Progress in Soil Science

Alex. B. McBratney
Budiman Minasny
Uta Stockmann *Editors*

Pedometrics

Springer

Progress in Soil Science

Aims and Scope

Progress in Soil Science series aims to publish books that contain novel approaches in soil science in its broadest sense – books should focus on true progress in a particular area of the soil science discipline. The scope of the series is to publish books that enhance the understanding of the functioning and diversity of soils in all parts of the globe. The series includes multidisciplinary approaches to soil studies and welcomes contributions of all soil science subdisciplines such as: soil genesis, geography and classification, soil chemistry, soil physics, soil biology, soil mineralogy, soil fertility and plant nutrition, soil and water conservation, pedometrics, digital soil mapping, proximal soil sensing, soils and land use change, global soil change, natural resources and the environment.

More information about this series at http://www.springer.com/series/8746

Alex. B. McBratney • Budiman Minasny
Uta Stockmann
Editors

Pedometrics

Springer

Editors
Alex. B. McBratney
Sydney Institute of Agriculture
& School of Life and Environmental Sciences
The University of Sydney
Sydney, NSW, Australia

Budiman Minasny
Sydney Institute of Agriculture
& School of Life and Environmental Sciences
The University of Sydney
Sydney, NSW, Australia

Uta Stockmann
Sydney Institute of Agriculture
& School of Life and Environmental Sciences
The University of Sydney
Sydney, NSW, Australia

ISSN 2352-4774 ISSN 2352-4782 (electronic)
Progress in Soil Science
ISBN 978-3-319-63437-1 ISBN 978-3-319-63439-5 (eBook)
DOI 10.1007/978-3-319-63439-5

Library of Congress Control Number: 2017955385

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Pedometrics, not in name, but in action, arguably began in the nineteenth century and grew into the twentieth century where we can see early attempts at quantitative expression of spatial and temporal soil variation. The information age which began with the advent of digital computers in the late 1950s saw a second phase of development largely in relation to numerical soil classification and multi-attribute description of soil objects. Geostatistics and information systems brought a large expansion of work in the 1980s and 1990s and also formal recognition and a label – *pedometrics*. The new millennium has seen pedometrics grow from strength to strength expanding the basic science and developing many new areas of application. This book attempts to cover (almost) all the topics that pedometrics comprises in a didactic way. We write in the hope that this text will lead to improved understanding of soil variation and its place in earth system functioning and society.

Sydney, Australia
March 31, 2017

Alex. B. McBratney
Budiman Minasny
Uta Stockmann

Contents

Part VII Applications of Pedometrics

Contributors

Dominique Arrouays INRA, US 1106, InfoSol Unit, Ardon, Orléans, France

Pat H. Bellamy Cranfield University, Cranfield, Bedfordshire, UK

Nigel R. A. Bird Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Thomas F. A. Bishop Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Florence Carré INERIS, Parc Technologique Alata, Verneuil en Halatte, France

Estelle J. Dominati Farm Systems and Environment, Land and Environment, AgResearch Grasslands, Palmerston North, New Zealand

Mario Fajardo Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Cécile Gomez LISAH, Univ Montpellier, INRA, IRD, Montpellier SupAgro, Montpellier, France

Juan B. Grau Group of Automation of Signals and Communications, Universidad Politécnica de Madrid (UPM), Spain

Jaap J. de Gruijter Alterra, Wageningen University and Research Centre, Wageningen, The Netherlands

Gerard B. M. Heuvelink Soil Geography and Landscape group, Wageningen University and ISRIC – World Soil Information, Wageningen, The Netherlands

Allan E. Hewitt Manaaki Whenua – Landcare Research, Lincoln, New Zealand

Edward J. Jones Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Mark Kibblewhite Cranfield University, Cranfield, Bedfordshire, UK

Philippe Lagacherie LISAH, Univ Montpellier, INRA, IRD, Montpellier SupAgro, Montpellier, France

Juan C. Losada Complex Systems Group, Universidad Politécnica de Madrid (UPM), Spain

R. Murray Lark British Geological Survey, Keyworth, Nottingham, UK

Brendan P. Malone Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Ben P. Marchant British Geological Survey, Keyworth, Nottingham, UK

Manuel P. Martin INRA, US 1106, InfoSol Unit, Ardon, Orléans, France

Juan J. Martín-Sotoca Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Complex Systems Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

Alex. B. McBratney Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Jeroen Meersmans Cranfield University, Cranfield, Bedfordshire, UK

Irina Mikheeva Institute of Soil Science and Agrochemistry of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation

Alice E. Milne Sustainable Agricultural Sciences, Rothamsted Research, Harpenden, UK

Budiman Minasny Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Jason Morris Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Melissa Moyce Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Inakwu O. A. Odeh Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Nathan P. Odgers Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Soils and Landscapes Team, Manaaki Whenua – Landcare Research, Lincoln, New Zealand

Thomas G. Orton University of Queensland, Brisbane, Australia

José Padarian Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Stacey Paterson Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Matthew J. Pringle Department of Science, Information Technology and Innovation, EcoSciences Precinct, Brisbane, QLD, Australia

David G. Rossiter School of Integrative Plant Science, Soil & Crop Sciences Section, Cornell University, Ithaca, NY, USA

ISRIC-World Soil Information, Wageningen, The Netherlands

Antonio Saa-Requejo Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Resource Assessment Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

Nicolas P. A. Saby INRA, US 1106, InfoSol Unit, Ardon, Orléans, France

Sebastien Salvador-Blanes Laboratoire Géo Hydrosystèmes Continentaux, Faculté des Sciences et Techniques, Parc de Grandmont, Tours, France

Uta Stockmann Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Ana M. Tarquis Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Complex Systems Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

Iván G. Torre Complex Systems Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

Department of Applied Mathematics and Statistics, Universidad Politécnica de Madrid (UPM), Madrid, Spain

Tom Vanwalleghem Departamento de Agronomia, Universidad de Córdoba, Córdoba, Spain

Ichsani Wheeler Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Brett Whelan Precision Agriculture Laboratory, Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW, Australia

Part I
Introduction: What Is Pedometrics?

"As a science grows, its underlying concepts change..."

Hans Jenny
– Factors of Soil Formation:
A System of Quantitative Pedology (1941), McGraw-Hill,
New York. p. 1

« Les méthodes sont ce qui caractérise l'état de la science à chaque époque et qui détermine le plus ses progrès »

Augustin Pyrame de Candolle (1778–1841)

The first words of Hans Jenny's classic book and those of de Candolle presage the need for this text. Soil science advances, and as a consequence we have to develop and grapple with new concepts and methods. This book is an attempt to present some of these concepts and to hang them in appropriate places on some kind of framework to construct a new body of knowledge. The concepts dealt with here have arisen largely because of a phenomenon that has been going on in the earth and biological sciences since the invention of digital computers just a few years after Jenny's book first appeared, namely that of quantification.

First a short geological excursion to illustrate this. An article entitled *Physicists Invading Geologists' Turf* by James Glanz appeared in the New York Times on November 23rd 1999. Here are some excerpts:

> Dr. William Dietrich has walked, driven and flown over more natural landscapes than he can remember. As a veteran geomorphologist, he has studied how everything from the plop of a raindrop to mighty landslides and geologic uplift have shaped the face of the planet.
>
> So when he admits to the growing influence on his field of an insurgent group of physicists, mathematicians and engineers with all-encompassing mathematical theories but hardly any field experience, the earth almost begins to rumble.
>
> Geology, a field that has always gloried in descriptive detail but has had less luck deriving mathematical generalizations, is changing. Invigorated by satellite maps, supercomputers and fresh ideas from physics, researchers are deriving sweeping theories without ever having put hammer to rock.
>
> The trend is unsettling to some old-school geologists, but even they concede that the work has prompted new research in traditional academic circles. While the new researchers have not yet proved that the shape of every hill and dale can be predicted by simple equations, they have at least raised the question of whether landscapes are sculptured by something other than a reductionistic accumulation of forces.

> "In some ways, they irritate us," said Dr. Dietrich, a professor at the University of California in Berkeley, speaking for university and government geologists.
>
> But Dr. Dietrich added that the new, physics-inspired theories of landscape formation, which have found success by ignoring the detail-oriented approach of traditional geology and focusing instead on the earth's overall patterns, are "a way of discussing some sense of regularity in what otherwise is a very messy world."
>
> "I've come to appreciate the perspective," he said.
>
> Inspired by ideas long familiar in physics, and fueled by the recent availability of high-quality satellite maps of much of the planet's surface, the new approach turns the earth sciences on their head, asserting that the most prominent structures on the surface of the planet are shaped not by just local factors, as had been thought, but by the most general properties of physics and mathematics.
>
> For all the insights of the new research, geologists are far from abandoning their worldview. In fact, they too are using satellite data — to perform detailed computer calculations of how rain and wind erode one particular landscape into another over time. "But even there", Dr. Dietrich said, "the physics-oriented researchers have inspired discoveries of order amid the reductionistic detail."
>
> "The places where they accomplish that," Dr. Dietrich said, "will, I think, have a lasting effect on how we think about the planet".

This example illustrates the disciplinary and human relationships[1] between traditional and mathematical geologists, which is completely akin to that between conventional pedologists and pedometricians. Pedologists study pedology, and pedometricans develop pedometrics. (Each group has its contribution to make.) Pedology and pedometrics are closely related. Flippantly, pedology is about augering; pedometrics is about auguring.

Pedology (from the Greek *pedon*, 'ground'), a term first coined in Saxony in 1862 (Simonson 1999), is the scientific study of the soil. More specifically it is the study of soil as part of the natural environment. It is concerned with soil description, spatial distribution, genesis and sustainable use (inter alia Joffe 1949; Buol et al. 1997). Traditionally, it has had a descriptive and field focus (Basher 1997).

Pedometrics generally addresses the same issues as pedology but focuses on specific kinds of problems, those that can be formulated quantitatively and can be solved with quantitative mathematical and statistical techniques. The coining of the term pedometrics and its first definition is by McBratney (1986):

> The use of quantitative methods for the study of soil distribution and genesis and as a sustainable resource.

Pedometrics is a neologism derived from the Greek roots:

πεδον PEDON, the ground, earth, soil
μετρον METRON, measurement

Webster (1994) reminds us that pedometrics is used analogously to other words such as biometrics, psychometrics, econometrics, chemometrics, and the oldest of all geometrics. Etymologically, the word covers two main ideas. First the 'pedo' part corresponds roughly to that branch of soil science we call pedology and the soil,

[1] Science is a human construct and as such shows all the glories and imperfections of human nature.

and the metric part has been restricted to quantitative mathematical and statistical methods. If we borrow from the wide definition of biometrics, we get:

> The development and application of statistical and mathematical methods applicable to data analysis problems in soil science.

So it is essentially the application of probability and statistics to soil. Webster (1994), in addition to McBratney's earlier definition, suggested an alternative problem-oriented meaning, which he paraphrased as

> Soil science under uncertainty.

In this sense, pedometrics deals with uncertainty in soil models that are due to deterministic or stochastic variation, vagueness and lack of knowledge of soil properties and processes. Thus, mathematical, statistical and numerical methods could be applied to resolve the uncertainty and complexity inherent in the soil system, including numerical approaches to classification, which deals with supposedly deterministic variation.

Pedometrics is not new, although it was first formally recognised as a different branch of soil science to traditional pedology at the end of the 1980s. Mathematical and statistical methods have been applied to soil studies generally since at least the 1960s with the availability of digital computers and software. The thread stretches back much further to precomputer days, however. It appears to have its origins in agronomy and soil survey, rather than strictly pedology, in the early part of the twentieth century. Harris' (1915) study of soil spatial variation in experimental fields and Robinson and Lloyd's (1915) concern over soil survey error are early examples. We suggest that Forbes' (1846) study of the temporal variation of soil temperature modelled by Fourier series is indeed a pedometric study, and no doubt there are even earlier ones.

For several decades of the twentieth century, pedometrics (although unrecognised and undefined) was a tool for designing experiments and surveys and in advisory work. In the 1960s, pedometricians were concerned with the difficult problem of soil classification and applied the methods of numerical taxonomy. In the late 1970s, pedometricians began to treat soil properties as spatially correlated random processes and to utilise geostatistics for analysis and prediction. Indeed, in a recent collection of the most important papers in soil science historically (Hartemink et al. 2009), this earlier work from pedometrics was recognised in four papers: Youden and Mehlich (1937) on efficient soil sampling, Rayner (1966) on numerical soil classification, Beckett and Webster on soil spatial variation and Burgess and Webster (1980) on soil spatial prediction. More recently, in addition to these earlier themes, pedometrics has begun to attempt to elucidate pedogenesis by quantifying relations between individual soil properties and controlling factors (e.g. Minasny et al. 2008).

Some might argue that Jenny (1941) was the seminal text, especially because of the title of his book *Factors of Soil Formation. A System of Quantitative Pedology*. The text of Webster (1977) and revised a decade or so later by Webster and Oliver (1990) were clearly milestones. Other important texts such as Burrough (1986), Goovaerts (1997), Webster and Oliver (2001), Nielsen and Wendroth (2004),

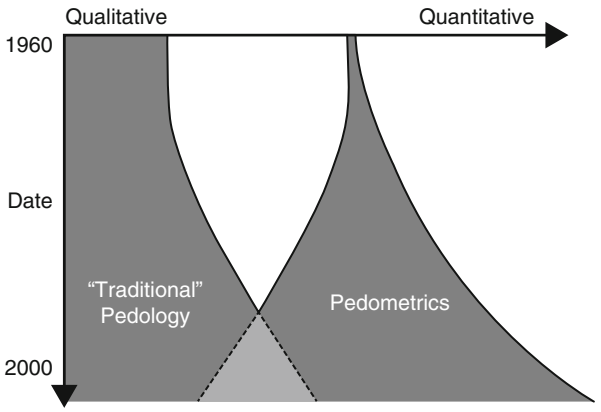


Fig. 1 A timeline of the growth of pedology and pedometrics (After McBratney et al. 2000, Fig. 1)

Grunwald (2005), de Gruijter et al. (2006) and Hengl and Reuter (2009) have a basis in pedometrics and contribute significantly to its canon. Pedometrics has contributed to, and gained from, applied statistics, geostatistics, GIS science, environmental sampling and geomorphometry.

Over time, the use of computers has increased in both pedometrics and in traditional pedology, and the difference between the two has decreased and in some cases overlapped (as shown in Fig. 1). Traditional pedology has, of necessity, become more quantitative through the increased use of computerised soil information systems and field-based measuring devices. Pedometrics has developed quantitative methods, which attempt to account for conceptual pedological models of soil variation. Now there is a strong and growing overlap and synthesis between traditional pedology and pedometrics.

In a bibliometric study of the composition of papers in a leading soil science journal from its inception in 1967 until 2001, Hartemink et al. (2001) showed that papers on pedometrics have risen from less than 3% in 1967 to around 18% of all papers in 2000. It seems, as shown in Fig. 2, that more qualitative soil genesis and morphological studies decreased to make way for the increase in more quantitative studies. By 2016, the proportion of papers in Geoderma on pedometrics had risen to ∼27%.

From a pedological point of view, Mermut and Eswaran (2001) saw pedometrics as a research tool with the potential to complement conventional soil surveys and a crucial technique in precision agriculture. By 2016, Brevik et al. (2016) recognised the role of pedometrics in soil mapping, classification and pedological modelling. A great potential has been demonstrated in applications such as digital soil mapping (Lagacherie et al. 2007; Hartemink et al. 2008; Boettinger et al. 2010) with a sixfold linear increase in published papers between 2001 and 2015 and proximal soil sensing (Viscarra Rossel et al. 2010).

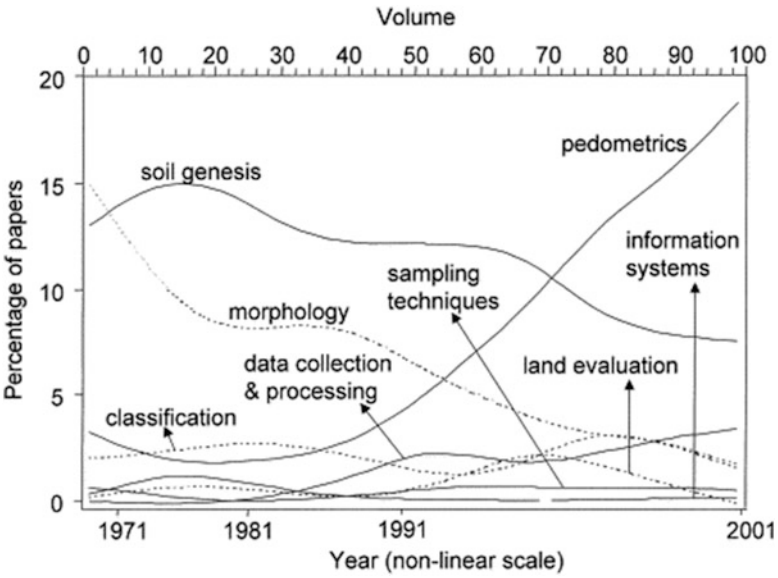


Fig. 2 Trends in pedological subdisciplines and subjects reported in Geoderma between 1967 and 2001. The ordinate is the percentage of papers in Geoderma (After Hartemink et al. 2001, Fig. 16)

References

Basher LR (1997) Is pedology dead and buried? Aust J Soil Res 35:979–994

Boettinger JL, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S (eds) (2010) Digital soil mapping: bridging research, environmental application, and operation. Springer, Dordrecht

Brevik EC, Calzolari C, Miller BA, Pereira P, Kabala C, Baumgarten A, Jordán A (2016) Soil mapping, classification, and pedologic modeling: history and future directions. Geoderma 264B:256–274

Buol SW, Hole FD, McCracken RJ, Southard RJ (1997) Soil genesis and classification. Iowa State University Press, Ames

Burgess TM, Webster R (1980) Optimal interpolation and isarithmic mapping of soil properties. 1. The semi-variogram and punctual kriging. J Soil Sci 31:315–331

Burrough PA (1986) Principles of geographical information systems. Oxford University Press, Oxford

De Gruijter JJ, Brus D, Bierkens M, Knotters M (2006) Sampling for natural resources monitoring. Springer, Heidelberg

Forbes JD (1846) Account of some experiments on the temperature of the earth at different depths and in different soils near Edinburgh. Trans R Soc Edinb 16:189–236

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Grunwald S (ed) (2005) Environmental soil-landscape modeling: geographic information technologies and pedometrics. CRC Press, Boca Raton

Harris JA (1915) On a criterion of substratum homogeneity (or heterogeneity) in field experiments. Am Nat 49:453

Hartemink AE, McBratney AB, Cattle JA (2001) Developments and trends in soil science: 100 volumes of Geoderma (1967–2001). Geoderma 100:217–268

Hartemink AE, McBratney AB, Mendonça Santos ML (eds) (2008) Digital soil mapping with limited data. Springer, Heidelberg
Hartemink AE, McBratney AB, White RE (eds) (2009) Soil science vol. I soil in space and time. Earthscan, London
Hengl T, Reuter HI (eds) (2008) Geomorphometry: concepts, software, applications. Elsevier, Amsterdam
Hengl T, Reuter HI (2009) Geomorphometry: concepts, software, applications. Elsevier, Amsterdam/Boston/Heidelberg/London/New York/Oxford/Paris/San Diego/San Francisco/ Singapore/Sydney/Tokyo
Jenny H (1941) Factors of soil formation. A system of quantitative pedology. McGraw Hill, New York
Joffe JS (1949) Pedology, 2nd edn. Pedology Publications, New Brunswick
Lagacherie P, McBratney AB, Voltz M (eds) (2007) Digital soil mapping: an introductory perspective, Developments in Soil Science 31. Elsevier, Amsterdam
McBratney AB (1986) Introduction to pedometrics: a course of lectures, CSIRO Division of Soils Technical Memorandum 53/1986. (mimeo). CSIRO Division of Soils, Glen Osmond. 150 pp
McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM (2000) An overview of pedometric techniques for use in soil survey. Geoderma 97:293–327
Mermut AR, Eswaran H (2001) Some major developments in soil science since the mid-1960s. Geoderma 100:403–426
Minasny B, McBratney AB, Salvador-Blanes S (2008) Quantitative models for pedogenesis – a review. Geoderma 144:140
Nielsen DR, Wendroth O (2004) Spatial and temporal statistics. Sampling field soils and their vegetation. Catena Verlag, Reiskirchen
Rayner JH (1966) Classification of soils by numerical methods. J Soil Sci 17:79–92. 20
Robinson GW, Lloyd WE (1915) On the probable error of sampling in soil survey. J Agric Sci (Camb) 7:144–153
Simonson RW (1999) Origin and acceptance of the term pedology. Soil Sci Soc Am J 63:4–10
Viscarra Rossel RA, McBratney AB, Minasny B (eds) (2010) Proximal soil sensing. Springer, Heidelberg
Webster R (1994) The development of pedometrics. Geoderma 62(1):1–15
Webster R (1977) Quantitative and numerical methods in soil classification and survey. Clarendon Press, The University of Michigan
Webster R, Oliver MA (1990) Statistical methods in land and resource survey. Oxford University Press, Oxford
Webster R, Oliver MA (2001) Geostatistics for environmental scientists. Wiley, New York
Youden WJ, Mehlich A (1937) Selection of efficient methods for soil sampling. Contrib Boyce Thompson Inst Plant Res 9:59–70

Chapter 1
Scope of Pedometrics

Alex. B. McBratney and R. Murray Lark

"Those are my principles, and if you dont't like them well, I have others."

Groucho Marx

1.1 The Agenda of Pedometrics

Why do we need pedometrics? What is its agenda? Pedometrics addresses certain key soil-related questions from a quantitative point of view. The need for the quantitative approach arises from a general demand for quantitative soil information for improved economic production and environmental management. Pedometrics addresses four main areas which are akin to the problems of conventional pedology:

1. Understanding the pattern of soil distribution in character space – soil classification
2. Understanding soil spatial and temporal variation
3. Evaluating the utility and quality of soil
4. Understanding the genesis of soil

A.B. McBratney (✉)
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: alex.mcbratney@sydney.edu.au

R.M. Lark
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
e-mail: mlark@bgs.ac.uk

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_1

1.1.1 Unravelling the Structure of the Soil Character Space: Soil Classification

We all search for organising generalisations in the face of complexity. Many soil properties need to be observed to characterise a soil. The organisation of this soil complexity through classification has been problematic, probably because classification is an innate human faculty. Leeper (1956, p. 59) noted "*When scientists discuss methods of analysing a solution for phosphate, they are practical, reasonable, and unemotional. When the same men discuss the classification of soils, these virtues are likely to evaporate*". Pedometrics attempts *to resolve some of the polemic of soil classification by a* search for insight into the structure of soil character space. Multivariate and numerical taxonomic methods (Webster 1977a) attempt to unravel this structure, with a view to better prediction and understanding. Chapters 4, 8 and 9 explore these issues in more detail.

1.1.2 Understanding Soil Spatial and Temporal Variation

The soil skins the land, and its attributes vary spatially and temporally in a sometimes continuous, sometimes discrete and sometimes haphazard fashion. Using current technologies, we can still only measure most attributes of the soil at a finite number of places and times on relatively small volumes, and therefore statements concerning the soil at other places or times involve estimation and prediction and an inevitable uncertainty. Such prediction and estimation are required for inventory, assessment and monitoring.

One of the tasks of pedometrics is to quantify this inexactitude in order that it can be known and managed accordingly (Heuvelink and Webster 2001). Pedometrics also seeks insight into such spatio-temporal patterns using geostatistical and other spatial and temporal description and prediction tools. Chapters 10, 11 and 12 explore these issues in more detail.

1.1.3 Evaluating the Utility and Quality of Soil

The importance of knowledge and awareness of soil resources ranging from individual fields to the global scale is axiomatic. Over and above the problems of spatial and temporal variation, we must put objective value judgments on the utility of soil for specified purposes. This may involve encapsulating practical experience and developing objective rules for management. Conventionally, this has been called land evaluation, and quantification has been underway since Storie (1933) or earlier and has been developed further by inter alia Rossiter (1990).

More recently, the focus has moved to the environmental performance of soil (Adhikari and Hartemink 2016). The somewhat problematic concept of soil quality includes assessment of soil properties and processes as they relate to the ability of

soil to function effectively as a component of a healthy ecosystem (Doran and Parkin 1994). The quality of soil as part of an ecosystem may depend on its pedodiversity (Ibáñez et al. 1995). Soil can provide a number of valuable ecosystem services (Dominati et al. 2010) that require to be quantified. This is also reflected in the newly emerging multidimensional concept of soil security where soil plays an integral part in the global environmental sustainability challenges (McBratney et al. 2014). In recent years, pedometrics has contributed more and more to these areas, with the use of pedometric products for land suitability analysis (e.g. Kidd et al. 2015) and assessment of the change in the soil resource in space and time (e.g. Stockmann et al. 2015), for example. These and related issues are explored in Chap. 17.

1.1.4 Understanding the Genesis of Soil

Ultimately, pedometrics would attempt to provide quantitative models of soil formation. This is achieved by encapsulating knowledge of pedological processes in mathematical forms as an alternative to purely statistical approaches. The success of such a modelling approach depends, however, on pedological knowledge and the non-linearity of processes (Phillips 1998). The advantages of a successful modelling of soil formation are substantial and manifold. The three previously discussed agenda items of pedometrics, namely, soil classification, spatial and temporal variation and soil utility and quality, would be predictable from such a model. Therefore, attempts at such an approach have emerged in recent years. Hoosbeek and Bryant (1992) perhaps first outlined the problem. Minasny and McBratney (1999, 2001), Cohen et al. (2010) and Vanwalleghem et al. (2013) have provided the first substantive, but still rudimentary models, followed by more sophisticated models on the soil profile scale and soil landscape scale. This intriguing approach is discussed further in Chaps. 18 and 19.

1.2 Types of Models and Their Evaluation

Models are abstractions of reality. Harvey (1969) gives an early general discussion of models and quantification. Dijkerman (1974) discusses the kinds of models used in soil science generally. Because the soil is a complex and variable system, we represent it with a simpler or more abstract model. Originally, these models were descriptive mental models, but over time they have become more quantitative (Dijkerman 1974). In the context of pedometrics, with its quantitative impulse, a model generally corresponds to one of Webster"s dictionary definitions, i.e. *a system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs*. Underlying this general statement, there is a variety of model types which we shall now consider and then go on to discuss how to judge when models are performing adequately.

1.2.1 Types of Models

Models can be described in various ways by dividing them into a number of binary categories.

Models may be *qualitative* or *quantitative*. In traditional pedology models tend to be qualitative, whereas in pedometrics the tendency is for quantification. The main advantage of quantitative models seems to be that of reproducibility. Some would argue that quantitative models are more objective, but this is a difficult area from a philosophical point of view. So qualitative models may be considered inferior; however, they do contain knowledge and that knowledge may be captured in some objective way. Qualitative models can be made quantitative. So a qualitative model may be a good starting point for subsequent quantitative investigation, e.g. qualitative models expressing the expert knowledge of a group of experienced scientists and practitioners may be formally written down and used as hypothesis for quantitative testing.

Models may be *static* or *dynamic*. Pedometrics tends to focus on static models, whereas the models of soil physics are dynamic. In the future, pedometric models probably will be increasingly dynamic.

Models may be *empirical* or *mechanistic*. By empirical models we simply describe a phenomenon with as few parameters as possible without necessarily seeking to describe the mechanism underlying the phenomenon, whereas mechanistic models attempt to describe the mechanism (at some scale). For example, we could model the changes in moisture content at some location as a function of time using a purely empirical time series or transfer function model. Alternatively, we could describe the same observations using Richards' equation (Pachepsky et al. 2003) and various other soil physical mechanisms. The latter requires knowledge of basic soil properties, whereas the former does not. Whether one uses a mechanistic or an empirical model will depend on our level of process knowledge and the relative predictability of the models. Another way of describing empirical versus mechanistic models is to term them functional and physical models (Addiscott and Wagenet 1985). Functional is largely synonymous with empirical. Sometimes the empirical and mechanistic categorisation is confused with static and dynamic. Mechanistic models are usually dynamic but as an example given above, empirical models can be dynamic as well as static.

A very important distinction is to consider *deterministic* or *stochastic* models. Underlying Laplace's (1749–1827) statement, "*All the effects of Nature are only the mathematical consequences of a small number of immutable laws*", is the origin of determinism described by Laplace and quoted by Addiscott and Mirza (1998), "*An intellect which at any given moment knew all the forces which animate nature and the mutual positions of all the beings that comprise it, if this intellect was vast enough to submit its data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom: for such an intellect nothing could be uncertain, and the future just like the past would be present before his eyes*". Stochastic models concern some

phenomenon that has randomness innate to its structure, whereas deterministic models have not. Many earth scientists do not readily accept that such randomness is inherent to nature. This unease recalls Einstein's famous saying "*The Lord does not play with dice*". No one has suggested that any soil process is constitutionally random unless we consider them at the quantum level. The ensemble averages are considered to be deterministic. So, for example, at the so-called Darcy scale, soil water phenomena can be described by a deterministic equation. However, given the inevitable lack of understanding and particularly lack of knowledge of soil processes, then randomness in models is a way of incorporating our ignorance. So in this sense, stochastic models may be seen as pragmatic. Up until the 1980s, stochastic models were rarely recognised or used in soil science. Dexter (1976) was one of the first examples. In hydrology on the other hand, stochastic models are abound (Yevjevich 1987). Pedometrics will probably make an ever-increasing use of stochastic models.

More recently, scientists have realised that deterministic models may have outcomes that look for all intents and purposes like stochastic ones – the outcomes appear random. These models describe processes which have a sensitive dependence on initial conditions. These are the so-called non-linear dynamic models, the basis of chaos theory (Gleick 1987). Such chaotic models blur the distinction between determinism and stochasticity (Addiscott and Mirza 1998). Phillips (1993, 1998) has suggested this kind of determinism as the basis for soil variability itself, i.e. soil formation is a chaotic process. Non-wetting phenomena in soil physics (Persson et al. 2001) are probably chaotic processes that cause soil variation. Minasny and McBratney (2001) give an example of a chaotic model and suggest how a chaotic deterministic model may be distinguished from a random one (Sugihara and May 1990).

As described above, models may be (1) qualitative or quantitative, (2) static or dynamic, (3) empirical or mechanistic (functional or physical) and (4) deterministic or stochastic. If we consider these four distinctions, then there are 16 combinations. Not all of them are sensible ones. In pedometrics, it is certainly worth considering the quantitative combinations and particularly those relating to stochastic and deterministic and empirical and mechanistic as shown in Table 1.1. So these four kinds of models are rather important in soil science. Another category has been

Table 1.1 Examples of quantitative models

	Deterministic		
	Certain	Uncertain	Stochastic
Empirical	Functional leaching model	Pedotransfer functions Fuzzy models	Regionalised variables geostatistics Markov chain models
Mechanistic	Richards' equation	Non-wetting phenomena Soil-landscape model	Molecular-scale diffusion Soil mechanics

added to the table, and we distinguish between certain and uncertain deterministic models. These models lie between purely deterministic and purely stochastic models. Examples of soil phenomena modelled in the various categories are given.

Pedometrics deals particularly with uncertain deterministic models and stochastic ones, i.e. models where randomness has to be introduced to deal with uncertainty (pedotransfer functions, Wösten et al. 2001) and/or randomness is seen as part of nature itself (soil as a realisation of random field – geostatistics). As such many pedometric models have a statistical basis. The uncertain deterministic, mechanistic and the stochastic, mechanistic models (e.g. soil mechanics, Manolis 2002) are not well developed in soil science or pedometrics.

One class of model that does not readily fit in to this scheme are the fuzzy models. They can be possibly thought of as uncertain deterministic, empirical models. They are probably best thought of as another categorisation, that of continuous or discrete states. Bárdossy et al. (1995) modelled soil water movement using the fuzzy approach as described in McBratney and Odeh (1997). Dou et al. (1999) modelled solute movement in a similar way. Pedotransfer functions (Wösten et al. 2001) which are largely linear or non-linear regression equations are empirical deterministic uncertain models, which will be discussed further in Chap. 7 of this book.

1.2.2 Critical Evaluation of Models and Their Parameters

Critical evaluatory procedures for models are needed to maintain the integrity of modelling and to ensure that the increasingly widespread use of models does not result in the propagation of misleading information. Generally speaking, models particularly quantitative pedometric ones, should be good predictors, while at the same time, they should not have too many parameters. The quality of prediction has to be tested with independent data sets by comparing predicted and observed values (Van Kuilenburg et al. 1982) or less satisfactorily with some kind of cross-validation procedure (Solow 1990).

The Franciscan friar, William of Ockham, was concerned about the number of parameters; he wrote in 1332, "*Pluralitas non est ponenda sine necessitate*", which roughly translates as "*Don't make things more complicated than you have to*". This is called Ockham's razor. The two ideas of quality prediction and minimising the number of parameters quantitatively are reflected in measures such as the Akaike information criterion (Webster and McBratney 1989).

Addiscott et al. (1995) pointed out that no model can be validated in the sense that it has been unequivocally justified. All that can be achieved is to show how small the probability is that the model has been refuted. Whether this probability is acceptable remains a subjective decision. In general, the further the data used for parameterisation are removed from the data to be simulated, the better. Problems can arise in both parameterisation and validation if the model is non-linear with

respect to its parameters, and the latter have appreciable variances. Parameterisation and validation become more difficult as the complexity of the model or the scale at which it is used increases.

1.3 Methods × Problems

The gamut of pedometrics can also be thought of as a two-way table. The rows are the problems that pedometrics can possibly address, and the columns are the potential mathematical and statistical methods that could be applied to such problems. Some cells of the table will be richly endowed, whereas others will require thought and invention to come up with new approaches. This book, which is essentially a contribution to soil science, and thereby earth, environmental and ecological sciences, is arranged with a problem focus rather than on the methods themselves which would be more appropriate to an applied statistics text.

1.3.1 The Problems of Pedometrics

"Sir Walter Blunt, new lighted from his horse.
Stain'd with the variation of each soil
Betwixt that Holmedon and this seat of ours;"

William Shakespeare, Henry IV Part 1 Act 1 Scene 1.

Pedometrics arises principally from this common observation that the soil is spatially variable. The soil varies laterally and with depth, and this variation has implications for its use and management. This is not a new insight. The rabbinical biblical commentators on the book of Genesis discussed the question of how much water is needed to sustain plant growth:

"How much rain must descend that it may suffice for fructification? As much as will fill a vessel of three handbreadths. This is Rabbi Meir's opinion. Rabbi Judah said: in hard soil, one handbreadth; in average soil, two; in humid soil three."

Midrash Genesis Rabbah XIII, 13.

Later in history, Thomas Tusser, in his *One Hundred Points of Good Husbandry* published in England in the sixteenth century, noted pithily that *each divers soil hath divers toile*. More recently, the development of agronomy, soil management and engineering has shown that the spatial variation of the soil means that its suitability for different purposes will vary in space. This is important for the planning of infrastructure, the government agencies making planning decisions about land use and, in recent years, the individual farmer trying to improve the efficiency of cropping systems by managing inputs in response to fluctuations in crop requirements at within-field scales.

We can think of other problems where the spatial variation of the soil will be of practical significance. Throughout the industrialised world, it is recognised that the soil is at risk of pollution from unregulated inputs and emissions. If we are to monitor this problem and focus protection and remediation where the problem is most severe, or the soil most vulnerable, then we must be able to detect changes in pollutants, which arise from complex processes of deposition and transport, against a background of complex intrinsic variation of the soil. A similar problem arises when we set out to monitor changes in the carbon content of the soil to evaluate its significance as a source and sink for greenhouse gases.

Soil variation is linked to some pressing problems, but it is also pertinent to other questions. Faced with variation in the soil cover of any terrain, the natural response is to ask how and why it arose. Many soil scientists would not have become interested in their subject if the soil was more or less uniform in space. To understand the variation of the soil at scales from the aggregate to the continent and the relationship of this variation to that of the vegetation which it supports is a basic scientific challenge. The better we understand these basic questions, the better equipped we will be to address the practical ones. Pedometrics is concerned both with meeting the practical requirements of soil management and also with generating insight into how and why the soil material varies as it does.

These, in summary, are the problems of pedometrics. In the remainder of this section, we want to set them in a general framework, with examples, as a prelude to introducing the methods which are used in their solution.

1.3.1.1 Prediction

The first type of problem is prediction. The basic question requiring a prediction has the form:

1. What soil conditions pertain at position **x**?

Vector **x** may contain two or three Cartesian coordinates which define a location in space and possibly a further number which defines a time. This very general question may be refined in different ways:

1.a What is the value of soil variable s at **x**, $s(\mathbf{x})$, *given a set of observations of the variable at other locations?*

The variable $s(\mathbf{x})$ might be the concentration of available potassium in the topsoil at a location in space $\{x,y\}$. Note that some finite volume of the soil is implicitly of interest. We call this volume the geometric support of $s(\mathbf{x})$.

This problem arises because very few if any soil properties of direct practical interest will have been measured exhaustively across a region. The soil is a continuum, and the constraints of costs and time mean that the soil may only be sampled and measured at a few sites. Information will inevitably be required about soil which has not been measured directly.

Reflection on this problem makes it clear that the best answer we can obtain to our question, short of actually sampling the soil at **x**, will be an estimate of $s(\mathbf{x})$,

$\widehat{s}(\mathbf{x})$, a number with attendant uncertainty. The way in which this uncertainty is quantified and controlled is a critical issue in pedometrics, to which we will return repeatedly.

A second refinement of the question is possible:

1.b. What is the value of, $y(\mathbf{x})$*, a mathematical function of soil property s at* $\mathbf{x}$*?*

This may seem at first sight like a pedantic variant of question *1.a.*, but it raises an important issue, and a practical one. We may be able, for example, to obtain a reasonable estimate of the clay content of the soil at $\mathbf{x}$, but are really interested in its available water capacity. A reasonable estimate of this latter quantity might be obtained as a function of the clay content. If y is a linear function of s, then the transformation of the estimate $\widehat{s}(\mathbf{x})$ to an estimate $\widehat{y}(\mathbf{x})$ is simple. If the relationship is not linear, then the uncertainty of the estimate $\widehat{s}(\mathbf{x})$ must be accounted for.

Other variants on the simple prediction problem are possible. For example:

1.c. What is the difference between $s\{x, y, t_i\}$ *and* $s\{x, y, t_j\}$*?*

where the third term in the vector denotes a time. This question will arise in environmental monitoring. If $s\{x, y, t_i\}$ is the concentration of a pollutant at x at time t_i, then the difference may be a measure of the success of a soil remediation campaign or the environmental impact of a change in regulations.

This problem will vary in form. A critical question is whether it is possible in principle to measure both variables – given the support of $s\{x, y, t_i\}$, does the disturbance of the initial sampling prevent a meaningful measurement from being made at $\{x, y, t_j\}$? This problem presents pedometricians with interesting challenges. Since politicians are increasingly interested in reliable estimates of the change of organic carbon in the soil, in response to the Kyoto protocol, the problem is also timely and topical.

So far, we have considered the soil over a small volume about a notional two- or three-dimensional location. In practice, we may be concerned more often with predictions about larger parcels or blocks of land, regularly or irregularly shaped. Such parcels may constitute management units, for example. The general problem, then, is

1.d. What is the value of $s(\mathbf{X}) = \int_{x \subset X} s(x)\mathrm{d}x$?

The integral implies that the new variable is effectively the average of all the notional point values, $s(\mathbf{x})$, in the region. Thus, we might be asked, 'what is the mean concentration of lead in the soil at this former factory site' or 'what is the mean concentration of available phosphorous in the soil of this field'? Since our prediction will be based on observations which are effectively point samples on volumes of soil, very small by comparison to the block $\mathbf{X}$, the problem involves generalisation from one spatial scale to a coarser scale. This is sometimes called 'upscaling' or 'aggregation'.

The change of scale can cause problems. Consider the following:

1.e. What is the value of $y(X)$ where y is a mathematical function of soil variables w and z at the scale of point observations $s(x)$ and $w(x)$?

Again, y might be the available water capacity of the soil and w and s bulk the density and clay content, respectively. If w and s are measured at all locations, then we might evaluate y at all these locations then aggregate these to the coarser unit. If the two variables have been measured independently, then we can combine two aggregated values, $s(X)$ and $w(X)$, but only if the functional relationship is linear. Otherwise, it is necessary to make some inference about the joint variation of variables s and w at the original point scale; this is sometimes called 'downscaling' or 'disaggregation'.

We have raised the issue of uncertainty already in the context of simple prediction, but uncertainty may sometimes be addressed directly in a pedometric problem. Consider the following:

1.f. What is the risk that $s(\mathbf{x})$ exceeds at threshold t?

The variable s might be the concentration of a pollutant and t a regulatory threshold. For example, the limit for concentration of lead in the soil set by the ANZECC guidelines is 300 mg/kg.

A type 1a question requires a simple prediction, $\widehat{s}(\mathbf{x})$, but in the context of the present problem, this is generally not adequate. If we act on the prediction, there remains the risk that contaminated soil is left untreated (because $\widehat{s}(\mathbf{x}) < t$ but $s(\mathbf{x}) > t$) or that expensive remediation is applied to land unnecessarily (because $\widehat{s}(\mathbf{x}) > t$ but $s(\mathbf{x}) < t$). If our predictions are unbiased, i.e. on average $\widehat{s}(\mathbf{x}) = s(\mathbf{x})$, and the costs of an error in one direction are more or less equal to the costs of an error in the other direction, then all we can do is try to reduce the uncertainty of our predictions as far as possible. Often, however, the costs of an error in one direction are much steeper than the other (e.g. the fines for leaving land unremediated may be large compared to the costs of remediation). In these circumstances, the best decision must account for the risk that $s(\mathbf{x}) > t$, given all available knowledge implicit in the prediction $\widehat{s}(\mathbf{x})$.

This latter problem was expressed in terms of a near-point support at $\mathbf{x}$. Answering the final prediction-type problem:

1.g. What is the risk that the mean of s over region $\mathbf{X}$ exceeds threshold t?

raises further problems for the pedometrician. Since management decisions will generally be made about a parcel of land, practical questions will often be framed this way.

1.3.1.2 Inventory and Allied Problems

A second category of pedometric problems can be recognised. If the problems in category 1 are variants of the question 'what conditions pertain at $\mathbf{x}$', then class 2 consists of variants on:

2. Over what subregion does conditions pertain?

At its simplest, the problem may be one of inventory – enumeration of the sites over which certain conditions are found. Thus, for example:

2.a. Over what region, X', does the value of soil variable s fall within the range $s_1 < s < s_2$?

As with *questions of* type 1.a., there is uncertainty attendant on any answer to this question since the soil properties are not measured exhaustively.

This problem might be posed by the land manager who wants to know where the depth of the soil over the underlying rock is large enough to permit the growth of a tree crop. It might also be asked by the farmer who wants to know where soil pH is likely to limit certain crops. In reality, a more complex problem might be posed:

2.b. Over what region, X', does some function of soil variables $y = f(s, w)$ fall within the range $s_1 < y < s_2$?

If we want to identify areas where the likely erosion losses exceed some threshold, then we may have to compute some non-linear function of soil properties like the universal soil loss equation (Wischmeir and Smith 1978). This poses similar problems to the analogous prediction problem 1.b.

Temporal monitoring of the soil may also generate problems with an inventory flavour such as:

2.c. Over what region, X', does the change in soil property s from time t_1 to time t_2 fall within the range $s_1 < s < s_2$?

This problem arises when the policy maker wants to know over what proportion of a landscape the concentrations of pollutants in the soil are diminishing or where the organic carbon content is increasing.

1.3.1.3 Decision Support

Prediction and inventory are tools. Managers require that the answers to these basic questions are integrated in a way which aids decision-making directly. This is the sphere of decision-support systems, and pedometric problems arise. There are two general types of problems:

3.a. What is the optimum management strategy over region X' to achieve goals A subject to constraints B?

So, for example, what is the optimum nitrogen rate to prescribe for a particular parcel of a field to maximise the economic return to the producer subject to the constraints that emissions to the environment through leaching and denitrification do not exceed some threshold? The problem requires process models of an appropriate level of sophistication to describe the whole system under different scenarios. Information on soil properties within the parcel of concern will also be required – subproblems of type 1.c., for example. The optimisation subject to a constraint

may be done numerically – i.e. by computing several runs of the system model in an ordered way to find the scenario which best meets the goals subject to the constraints.

In practice, decisions have to be made under uncertainty about many critical conditions, e.g. weather in the case of crop management. We then have a problem;

3.b. What is the optimum management strategy over region X′ to achieve goals A subject to constraints B and given that the conditions C have a particular statistical distribution?

1.3.1.4 In Pursuit of Insight

The problems so far have had a strongly practical flavour, but pedometric problems include basic scientific questions where a quantitative account of the spatial variation of the soil is required. We can identify such a problem:

4.a. Which factors appear to determine the lateral and horizontal spatial variation of soil property s?

This question implies a spatial scale – are we concerned with variations at the scale of microbial activity or geomorphic processes or some range between? This general problem may be addressed using analytical methods, but these always rest on assumptions which may not be realistic. An approach to the problem can never be driven purely by data analysis, however. Our investigation will be most fruitful if it is structured around a hypothesis. The approach may be statistical:

4.b. Is hypothesis H *about the causes of soil variation supported by the given observations of properties s...?*

or structured around a mechanistic model:

4.c. Given a process model linking input variables s, w and output y, can this particular set of observations of the variables be held to validate the model?

1.3.2 The Methods Pedometrics Uses

These questions are addressed by pedometrics. We now offer an overview of the pedometric methods that are treated in more detail in later chapters.

1.3.2.1 Statistical Prediction and Modelling

Random Variables

Ideally, soil scientists would like to base pedometric methods on quantitative understanding of soil processes. The ideal way of predicting the value of a soil property at location x would be to enumerate factors of soil formation which

pertain at x, i.e. the climate, the organic influences, the relief and parent material and the development of these factors over the time in which the soil material has developed. This information would then be combined with knowledge of the relevant soil processes in order to predict the soil property of interest. This recalls the determinism of Laplace, who held that if we know the momentum and position of all particles in the Universe at a given time, then all future changes are predictable. But modern physics has had to abandon Laplace's ideals, and neither is it an option for pedometrics. First, it is clear that our knowledge of the soil-forming factors operating over time at any location is very incomplete, and some of these factors will be unpredictable in principle. For example, a grazing *Megaloceros* in the Pleistocene disturbing a patch of metastable soil on a periglacial hillslope might influence the course of local solifluction events and play a large part in determining the clay content of the topsoil at positions downslope which a pedometrician later wants to predict. Second, many soil-forming factors have a complex and non-linear effect on the development of the soil and will interact. As a result small errors in our information about critical soil-forming factors may render our predictions quite inaccurate. This sensitivity to initial conditions recalls chaos theory. The solution to the problem is usually to replace our notional deterministic model of the soil with a statistical model. This latter model may have a structure which reflects the knowledge of soil-forming processes contained in the deterministic model, while representing the relationship between these factors and soil properties statistically.

At its simplest, we regard the value of a soil property s at location $\mathbf{x}$ as the outcome of a random process s. To do this is to treat the soil property as if it were the outcome of a process such as a toss of a coin or a roll of a die. At first glance, this is a deeply paradoxical approach for the scientist. By definition, the outcome of a random process is uncaused, and yet we know that soil properties are caused by processes which we can list and of which we have, more or less, a sound scientific understanding.

Treating a soil property as a random variable is an assumption. That is to say, we know that the soil property is not strictly the outcome of a random process, but that certain conditions permit us to treat it as such for certain purposes. More generally, we do not necessarily treat all the variation of the soil as random, but rather may limit this assumption to a component of the variation which a simplified mathematical description of the variation as a whole cannot account for.

When statistical models are used to describe the variation of the soil, there are broadly two conditions which justify the assumption that data on the variables of interest are outcomes of a random process. In the first instance, we may treat a set of measurements of a soil property as random variables if the selection of the sample has been done in a random way. This is the basis of design-based statistical analysis where data are obtained in accordance with a design which does not specify where the soil is sampled to collect a specimen for analysis, but which allows us to state in advance only the probability that a particular location will be sampled. The second situation is subtly different. Here, the assumption is that a set of measurements of the soil in space and/or in time may be regarded as a realisation of an assemblage of random variables which have a structure in space and/or time which has sufficient

complexity to encapsulate at least some of the features of the spatial variations of the real variables. In this case, we assume an underlying random process and that all observed and unobserved values of the soil within the region of interest constitute a single realisation of this process. This approach allows us to predict the values of the soil at unobserved locations from the values which have been observed in a way which is in some sense optimal. This approach is the basis of the so-called model-based analysis.

We must review some properties of random variables. A random variable is characterised by an underlying probability density function, (pdf) $p(S)$, such that the probability that a given value lies in the interval S_1 to S_2 is given by the integral

$$\int_{S_1}^{S_2} p(S)\mathrm{dS}.$$

It follows that the integral over the interval $-\infty$ to $+\infty$ is exactly 1. From the pdf, we may also define the distribution function for $S, f(S)$:

$$f(S) = p\,[S \leq s] = \int_{-\infty}^{s} p(S)\mathrm{dS}.$$

If all we know about a particular value of a soil property is that it is a realisation of a particular random variable, then our best estimate of the value is the statistical expectation $E[S]$, where

$$E\,[S] = \int_{-\infty}^{\infty} \mathrm{S}\ \ \mathrm{p}(S)\mathrm{dS}.$$

The expectation of a random variable is also known as its mean or first-order moment. Higher-order moments may also be defined. For example, the variance σ^2

$$\sigma^2 = E\left[\{S - E\,[S]\}^2\right] = \int_{-\infty}^{\infty} S\{S - E\,[S]\}^2 p(S)\mathrm{dS}.$$

If we assume, as is often done, that our soil property is a realisation of a normally distributed random variable, then these two moments alone are sufficient to characterise it and may be estimated from data.

The Linear Model

Simple random variables alone are of relatively little use for pedometric purposes. Ideally, we incorporate them into a statistical model to account for that variation which a simplified predictive equation cannot explain. Consider as an example the soil property organic carbon content. Many factors will determine this variable. One

such factor may be clay content, since the clay fraction of the soil may protect some of its organic matter. In addition, heavier soils which retain more moisture may have a larger input of fresh organic material in litter, roots and detritus from more vigorous vegetation. These and other considerations may lead us to expect that larger organic carbon contents of the soil may be associated with larger clay contents, although other factors will also be important and may mean that a good deal of observed variation in organic carbon content cannot be explained by a predictive relationship to clay content.

At its simplest, such a relationship might be a linear one of the form

$$\mathrm{OC} = a + b.\mathrm{Clay} + \varepsilon.$$

Here b is the slope of the linear relationship and a the intercept. The last term is the error or residual term, which sums up the other factors which determine organic carbon content of the soil. In statistical modelling, it is this term which we treat as a random variable. If the variability of this last term is small relative to the variability of organic carbon overall, then the model may be useful for predictive purposes. Under certain circumstances, we may estimate the parameters a and b from observations of the organic carbon and clay content of the soil by finding the values which minimise the deviations:

$$\mathrm{OC} - (a + b.\mathrm{Clay}),$$

in effect by minimising the variance of the error term.

The reader may have recognised that we have just described the ordinary least-squares regression model. This is a special case of the general linear model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is a vector of observations of a variable (which we may wish to predict) and matrix $\mathbf{X}$ contains a set of variables on which we assume the variable in y to be dependent. Vector $\mathbf{b}$ contains a set of coefficients, and $\boldsymbol{\varepsilon}$ contains realisations of a random variable, the error term.

Each column of matrix $\mathbf{X}$ corresponds to a predictor variable. These may be continuous variables (like clay content) or indicator variables which take the value 0 or 1 indicating whether or not an observation belongs to a particular class.

Consider a case where we wish to model the effect of parent material on soil properties. If clay content is the variable of interest, then different classes of parent material may weather to produce material of different textures. If g parent material classes have been defined, so that any one observation belongs to exactly one of the g classes, then the solution of the general linear model will be given by a vector of coefficients equal to the class means. Other variables, continuous or indicator, may be combined in the model. This general linear model (GLM) (McCullagh and

Nelder 1989) is equivalent to the familiar analysis of variance on the clay data testing the differences among the class means. In a similar vein, statisticians have recognised that nature is often not linear and have developed a class of non-linear models analogous to the GLM called the general additive models (GAMs) (Hastie and Tibshirani 1986, 1990).

Random Functions

A random function is an assemblage of random variables and is defined as a function of the location in space, $S(x)$. This function is assumed to have certain properties which characterise its behaviour over a set of all possible locations within some region of interest. A set of observations of a soil property at n locations in this region, $s(x_1)\,,\ \ldots\ s(x_n)$, is a sample of one realisation of this random function. We can only make progress by assuming that there are some restrictions on the joint distribution of $S(x_1)\,,\ \ldots\ S(x_n)$ which allows us to make inferences about it from the assemblage of observations.

We have introduced the concept of a pdf and a df of a random variable. We may similarly define the joint df of a random function over a set of locations:

$$F_{\{x_1,\ldots,x_n\}}\left(s_1,\ldots,s_n\right) = \text{Prob}\left[S\left(x_1\right) \leq s_1,\ldots,S\left(x_n\right) \leq s_n\right].$$

The simplest assumption which we might make about a random function is that the joint df for a configuration of sample points depends only on their relative distribution in space and not on their absolute position. Thus,

$$F_{\{x_1,\ldots,x_n\}}\left(s_1+\mathbf{h},\ldots,s_n+\mathbf{h}\right) = F_{\{x_1,\ldots,x_n\}}\left(s_1,\ldots,s_n\right),$$

where $\mathbf{h}$ is a vector defining a spatial separation or lag.

This is the assumption of strict stationarity. In practice, we work with less restrictive assumptions. One is to assume that only first- and second-order moments of the joint distribution are invariant with a shift in position. Thus, the mean is assumed to be constant:

$$E\left[S\left(x_i\right)\right] = \mu \text{ for all } x_i.$$

The covariance of any $S(x_i)$ and $S(x_j)$:

$$E\left[\left\{Z\left(xi\right) - E\left[Z\left(xi\right)\right]\right\}\left\{Z\left(xj\right) - E\left[Z\left(xj\right)\right]\right\}\right]$$

is also assumed to depend only on the interval $(x_i - x_j)$. If the joint df of the random function is multivariate normal, then strict stationarity is equivalent to this second-order stationarity since there are no higher-order moments.

Even this weaker version of stationarity may be too restrictive. In pedometrics, we often work with the assumption of intrinsic stationarity. This has a first-order assumption:

$$E\left[S\left(x_i\right) - S\left(x_j\right)\right] = 0 \text{ for all } x_i \text{ and } x_j,$$

and a second-order assumption that

$$E\left[\left\{S\left(x_i\right) - S\left(x_j\right)\right\}^2\right] = \gamma\left(x_i - x_j\right),$$

where γ is a function of the separation vector $x_i - x_j$. This function is the variogram and is widely used in pedometrics. It may be estimated from data and modelled with an appropriate parametric function. It is also possible to model the joint spatial variation of two or more variables in terms of random functions with variograms and cross-variograms. This is discussed in more detail in Chaps. 10 and 21. Having obtained a variogram model, this may be used in determining an optimal estimate of the value of a soil variable at an unsampled site as a weighted average of the values at neighbouring sites. This is the technique known as kriging.

State-Space Models

Consider a soil property which evolves over time. We may be able to express this change quantitatively by a model of the form

$$s_t = f\left(s_{\{t-i\}}\right)_{\{i=1,\ldots,n\}} + \varepsilon_t.$$

This states that the observed value is a function of past values, with an error term – a random variable which describes other factors which the model cannot account for. This is the so-called state equation. We recognise that our measurement of s, which we may make on several occasions, is made with error. The measurement process is described by a linear measurement relation:

$$u_t = g_t\left(s_t\right) + \varepsilon_t$$

where ε_t is an error term. The technique of Kalman filtering combines these two equations so that s_t may be estimated from a series of measurements. The state-space approach is of particular interest because the state equation can be a process model which incorporates our best scientific understanding of the processes whereby s is evolving over time. This state-space model may also be incorporated into a spatial model. In recent years, this approach is increasingly used by soil scientists in alliance with computational advances (e.g. Heuvelink et al. 2006; Huang et al. 2017).

Point Processes

There is another kind of process in the soil which we may usefully describe by or compare with probabilistic models of spatial variation. This is the distribution in space of discrete events. Consider, for example, the distribution in the soil of eggs of a nematode such as *Heterodera*. Soil scientists and agronomists might ask the pedometrician how quantitative analysis might illuminate the process by which the eggs are distributed. Does the distribution appear to be controlled by soil variation? Is it 'patchy'? We may make useful progress by comparing the observed distribution of eggs with what might be expected if they were scattered at random. In general terms, two departures from simple spatial randomness (SSR) are possible. The first is a so-called contagious distribution where occurrences of the event are more 'clumped' than SSR. This might occur if the nematodes grow and reproduce preferentially in areas of the field where certain soil conditions prevail and so tend to be aggregated here. The second departure from SSR is called overdispersion where the events are distributed more evenly than is expected of a random process. This might occur if competition between the adult nematodes is so intense that they tend to disperse in search of resources. Distinguishing SSR from contagious or overdispersed spatial patterns is not always simple. It is necessary to allow for the fact that we might only observe a fraction of the events in a field by sampling. Furthermore, the spatial distribution of a process might be complex with features such as patchiness being observed at particular spatial scales. Walter et al. (2005) used spatial point-process statistics to model lead contamination in urban soil. More conventional approaches are discussed in Chaps. 10 and 12. Glasbey et al. (1991) modelled the three-dimensional pore space within soil aggregates with randomly positioned overlapping spheres. Pore structure modelling is discussed further in Chap. 6.

Basis Functions and Decompositions

Any ordered set of n measurements on soil can be written down in a $n \times 1$ array or vector. Two instances are considered here. In the first, the vector contains measurements of one soil variable at n locations; here we consider regularly spaced locations on a linear transect. Location determines the order in this vector. In the second, the vector contains measurements of n soil properties from one location. A standard, though arbitrary, order of the properties is determined in advance.

It is easy to imagine a 3×1 vector as a point in the three-dimensional rectilinear space of our everyday experience, the three values being Cartesian coordinates relative to some arbitrary origin at {0,0,0}. In fact this is a special case of a general class of vector spaces. A vector space is a set of vectors with particular properties, so the set of all possible vectors of measurements of the volumetric water content, pH and clay content of the soil constitute one vector space with particular properties. All these soil variables are real numbers and vary continuously. This vector space is normed, i.e. we can define a distance between two vectors $\mathbf{x}$ and $\mathbf{y}$; in this case, a natural norm is the Euclidean distance $[\mathbf{x}\mathbf{y}^T]^{1/2}$.

Vectors of the first kind, where the soil properties vary continuously and with finite variance, and of the second kind, where the soil properties are continuously varying real numbers, occupy a subset of Hilbert space. The reader is referred to mathematical textbooks for a rigorous account of Hilbert space. For our purposes, the key fact is that Hilbert spaces can be described in terms of basis vectors. If $\mathbf{x}$ is any vector in such a space $\mathbf{X}$ and the basis vectors of $\mathbf{X}$ are $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$, then we may express $\mathbf{x}$ by

$$\mathbf{x} = \alpha\mathbf{a}_1 + \beta\mathbf{a}_2 \ldots$$

If the basis vectors are orthogonal, i.e. $\mathbf{a}_i\mathbf{a}_j^T = 0$ for any $i \neq j$, then n such vectors provide a complete basis for an n-dimensional space $\mathbf{X}$. As an intuitive example, note that we can characterise the position of any point in the 3D space of daily experience in terms of any three coordinates from a system where the axes are mutually perpendicular.

The reason for this mathematical digression is that pedometricians can often gain insight into data expressed as one or more vectors by expressing them as combinations of basis vectors. There are two general approaches. First, we may find the coefficients which express our data in terms of certain predetermined basis vectors. The analysis of the resulting coefficients may be informative. Second, we may determine both the basis vectors and the coefficients in our analysis where the former are chosen to illuminate how the data are distributed in vector space.

The first of these approaches is exemplified by Fourier analysis and wavelet analysis. The vectors here are of the first type (i.e. one vector represents one soil property measured at different locations). In Fourier analysis, we decompose a continuous variable (or its discrete sampling) into additive combinations of sine functions of different frequencies (where spatial frequency has units of distance^{-1}). The Fourier coefficients are complex numbers and so can convey two pieces of information, in this case, the amplitude of the particular component (the height of a peak over a trough) and the phase (how the peaks are aligned relative to our sample points). Analysis of these coefficients can show us how the variation of the soil property is partitioned between spatial frequencies, that is to say, between fluctuations over short distances in space and longer-range fluctuations. This may be informative. However, if our data show changes in the nature of their variability in space, non-stationarity in the sense of our discussion above, or show marked singular features or discontinuities, then the decomposition of the data on a basis of uniformly oscillating sine functions is not natural, and the coefficients will not yield a simple interpretation. For this reason, Fourier analysis has not been widely used in pedometrics. Where it has been found useful, the soil variation has been dominated by a more or less uniform periodic component such as the gilgai pattern analysed by Webster (1977a, b), McBratney (1998) and Milne et al. (2010).

Wavelet analysis is related to Fourier analysis in that data are decomposed on a basis of oscillating functions which respond to soil variation at different spatial frequencies. However, wavelets only oscillate locally and damp rapidly to zero so a single wavelet coefficient only describes the soil variation in a particular neighbourhood (the size of which depends on the spatial frequency).

A complete wavelet basis consists of dilations of a basic (mother) wavelet function which respond to different spatial frequencies and translations (shifts) of these which describe different neighbourhoods. The wavelet transform is a new tool in pedometrics, but a very promising one, and we discuss it further in Chap. 15.

When we have several data vectors of the second kind, observations on several soil variables at different locations, then we can define an orthogonal basis for the vector space in which these are distributed. An orthogonal basis is effectively a new set of coordinates for the vector space obtained by rotating the coordinate system of the original space. It is possible to find a basis such that the new variables which are defined are uncorrelated and so that one is the most variable such transformation of the original data which is possible, the second is the next most variable transformation orthogonal with the first and so on. These new variables are called principal components. If some or all of the original variables are substantially correlated, then it may be found that much of the original variability in n-dimensional space may be accounted for by fewer than n of the principal components. This may help the exploration of the multivariate structure of large and complex data sets and give insight into the joint variation of many soil variables. This is explored further in Chap. 4.

Classification and Allocation

Faced with many variable objects, the instinctive human response is to group them together into a manageable number of classes about which meaningful generalisations are possible. Soil classes may be formed by interpretation of profile characteristics in terms either of the evidence which they offer of the processes of soil genesis or the behaviour of the soil for practical purposes (e.g. drainage classes). This is principally a qualitative activity in which the expertise of the soil surveyor or pedologist is brought to bear on the problem. Pedometrics can offer two principal aids to the process:

1. Clustering. Consider again the vector space defined by several soil variables. If the space which is occupied by observations is more or less evenly populated, then it is clear that the recognition of classes defined on these properties will be difficult and somewhat arbitrary. We could, for examples, divide our observations into a number of more or less similar subvolumes of the vector space, but there are probably very many possible subdivisions which are of similar compactness as measured by the internal variability of the classes. A vector space occupied by data on soil which consists of a few distinct ‘types’ is likely to contain distinct ‘lumps’ or clusters of observations, since observations which are similar will be near to each other in the vector space. In the basic conceptual model which underlies classification, observations resemble typical or central members of the class to which they belong, although not exactly, and so should be clustered in the vector space defined on important soil properties.

Finding clusters in a vector space of many dimensions is a task for the computer. We will discuss the problem in more detail elsewhere. Suffice it to say that the key to the task is to define a model of within- and between-class variation in terms of a norm in vector space which is appropriate to the actual multivariate variability of the data.

2. Allocation. Having formed a classification, we may wish to allocate a new observation to the appropriate class. This is relatively straightforward where the classes have been defined (perhaps by clustering) in an n-variate vector space, and we have a new n-variate vector to allocate. This is often not the case. If we have measurements of one or more secondary variables which differ between the classes even though they are not definitive, then it may be possible to allocate an observation to a class using these. For example, we may have measurements of the reflectance of the soil surface in several frequency bands obtained by a remote-sensing device. Using several sites where the reflectances are known and the soil class, rules may be determined for predicting the soil class at sites where only the reflectances are known. Discriminant analysis due to Fisher is one technique for deriving these rules, which are formulated in terms of a norm in the vector space. In recent years, methods such as neural networks have been developed to solve the same problem.

   Classification and allocation are discussed further in Chaps. 8 and 9.

Modelling

A soil scientist may propose a model of some soil process which describes it quantitatively. The pedometrician may be interested in using this model to predict the behaviour of the soil in space and time. This entails four principal tasks:

1. Calibration. A model will often contain parameters – quantities which are constant in any one instance. Ideally, the values of parameters will be deduced from first principles, but usually it is necessary to estimate some or all of a set of parameters, some of which may bear a complex relationship to underlying soil properties (e.g. the 'tortuosity' of pathways for solute movement through the soil which depends on soil structure at different scales) or are somewhat artificial, not bearing a simple physical interpretation (e.g. the 'permeability' parameter of some functional models of solute leaching (Addiscott and Bailey 1991). Calibration requires that some data are available on the basic soil variables which are inputs to the model and those which are outputs. The temptation to estimate parameters by least squares must (generally) be resisted because this assumes that only the output variable is subject to random error, and this may often not be the case.
2. Validation. A model which generates quantitative predictions of a soil process must be measured against reality before it is relied on for practical purposes, and scientists will be interested in assessing model performance to identify

gaps in their understanding of the underlying processes. Model validation is often difficult. The outcome of a modelled process may often be difficult if not impossible to measure in the field, particularly when the model describes processes at coarse spatial scales. Where the modelled output can be measured directly, then ideally, we would wish to measure outputs from several locations or instances where the values of input variables are all the same. This allows us to estimate the variability of the output variable which is not explained by the model (since the model output will be the same in all instances) and so to estimate from the variability of the deviations between the measured and predicted outcomes how much of this can be attributed to 'lack of fit' of the model (Whitmore 1991).

3. Assimilation. Ideally, a model of a soil process, driven by some readily measured variables, will substitute for direct measurement of the process. In practice, we may wish to combine a model with limited measurements in order to predict outcomes or monitor a process. In this way, a purely statistical approach to estimating values is supplemented with understanding of how the system is likely to behave. The incorporation of measurements of a process into a model is known as data assimilation. It is widely used in oceanography and atmospheric science and has lately become fruitful in soil science. The state-space models described above are useful to this end; see, for example, Huang et al. (2017).
4. Error propagation. In practice, data with attendant errors are used as input to models with parameters estimated with error. Often the output of one model might function as input to another (e.g. modelled production of nitrate by mineralisation may be part of the input to a leaching model). It is clear that errors will propagate through such a system. If the models are linear, then the errors accumulate relatively slowly, and if all the errors have zero mean (i.e. are unbiased), then so is the final outcome. Neither is true of non-linear models, which may rapidly magnify the error and where the output may be biased even if the mean error of the inputs is zero. The error propagation of a set of models might be investigated by a Taylor series approximation where an analytical form of the model exists. For example, if an input variable, s, has a mean of $\widehat{s}$ and a variance of σ_s^2, and a model for an output variable y may be described by the non-linear function $y = f(s)$, then the approximate mean value of the output according to a second-order Taylor series approximation is

$$f\left(\widehat{s}\right) + \sigma_s^2 \frac{\delta^2 s}{\delta s^2} f\left(\widehat{s}\right).$$

A similar approximation to the variance of the output may be written.

A more generally applicable approach is to use Monte Carlo simulation. This requires that we have a reasonable knowledge of the variances and correlations of the input variables. We may then simulate many sets of inputs and obtain the mean and variance of the modelled outputs.

Heuvelink et al. (1989) used both approaches to study error propagation in the widely applied universal soil loss equation which combines several input variables

in a strongly non-linear way. They showed that the propagation of error was such that quite small uncertainty in the input variables could make the resulting predictions of little practical use. This is discussed further in Chap. 14.

1.4 Soil Attributes and Objects

It is worth considering the kind of objects that pedometrics studies. Pedometrics deals mainly with data on attributes observed on soil objects.

1.4.1 Soil Attributes

The attributes can either be direct or indirect (proximally or remotely sensed, including humanly sensed) measurements of physical, chemical or biological soil properties in the field or on specimens taken back to the laboratory or multivariate soil classes derived from them. The word attribute is used because the properties are *attributed* to the soil. Attributes are a function of the soil and the measurement process. The attributes can be discrete or continuous; see Chaps. 2, 3, 4 and 5.

1.4.2 Soil Objects

1.4.2.1 Geometric Objects

The objects on which attributes are observed can take a variety of forms, but most usually it is from some more-or-less fixed volume such as a soil core or pattern of soil cores or an auger boring or profile pit. Some workers have recognised the need for this volume to be of a minimum size in order that a good estimate of particular properties can be made. The measurement of water content of a saturated uniformly packed sand is a simple example of the dependence on scale of soil measurements. Sampling a very small volume gives a variation in the volumetric water content between zero and unity. Sampling a larger volume reduces the range of results, until, when a large enough volume is sampled, each measurement gives effectively the same result. The smallest volume at which this occurs is the representative elementary volume (REV). Buchter et al. (1994) effectively demonstrated the concept for stone content in a Swiss soil (Fig. 1.1). The range in percentage stones decreases exponentially from 100% for small sample lengths (volumes) to an asymptote of about 10%, which represents some larger-scale or macroscale variability. Figure 1.1 suggests that in this soil, the representative elementary volume is about $0.3 * 0.3 * 0.3 = 0.0027$ m^3 or about 30 l.

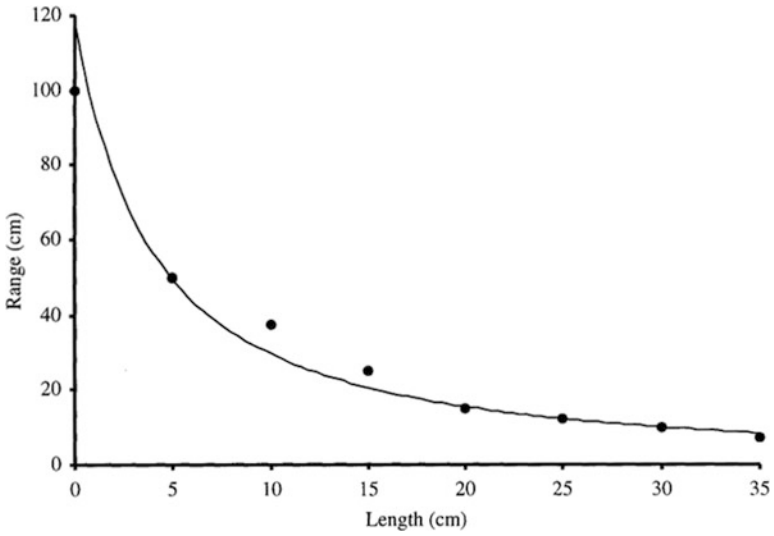


Fig. 1.1 The representative elementary 'length' of stone content in a Swiss soil. The range in percentage stones decreases exponentially from 100% for small sample lengths to an asymptote of about 10%. The graph suggests the representative elementary volume is about 0.03 m^3 (30 l) (Figure 6 from McBratney 1998)

The REV concept does not seem to have been applied to chemical and biological attributes or to soil classes for that matter. In a related way, geostatisticians (Isaaks and Srivastava 1989) recognise the need to define the so-called geometric support of observations. Geostatistics tells us that observations on larger supports or volumes within a fixed area will have smaller variances. The stabilisation of this variance is a way of defining the representative elementary volume. So, the REV concept and the geostatistical concept of geometric support are closely related.

1.4.2.2 'Natural' Objects?

These statistical concepts are not related to natural soil individuals, however. Soil science has always been troubled by a lack of clear definitions of individuals. When we study botany, we can almost always define and recognise an individual plant but in soil science, the definition of an individual soil is problematic. Pedologists recognise individuals such as the soil horizon (FitzPatrick 1993), the soil profile and the pedon (Johnson 1963). The problem with these is the lack of clear boundaries or of a fixed geometry. Russian pedologists recognise larger individuals such as the elementary soil area, the tessera and the tachet (Fridland 1976). Once again these are geometrically problematic.

1.4.2.3 Operational Objects

Holmgren (1988) sees the definition of soil individuals as a difficult problem and wrong-minded. He says 'that such distinct entities do not exist in soil is obvious'. He therefore suggests we should not seek such natural objects and argues for what he calls a point representation. He proposes a concept of a pedon that comprises a set of observational procedures which will lead to a set of attributes, including measurements of spatial and temporal variation, centred on a point. Holmgren's (1988) definition is:

> *A pedon is the possibility for soil observation in respect to a geographic point location. It can be realised by a set of observational propositions, each spatially and temporally specified in relation to that location.*

This concept seems to be more related to the practice in soil physics, chemistry and biology. This representation may be aligned with the geostatistical and REV concept above. The space-time geometric support of this operationally defined object is not clear, however. The two concepts could be combined, however. This is discussed further in Chaps. 8, 9, 10 and 21.

Ultimately, pedometrics is more a study of soil information (about given locations) than a study of natural or unnatural soil bodies and their attributes.

1.5 Scale Hierarchy

The word 'scale' is widely used in contemporary soil science, often confusingly. Conventionally, cartographers use the word 'scale' to refer to the ratio of a distance on a map to the corresponding distance on the ground. This simple descriptor denotes the detail which a particular map can convey. Thus, a 'large-scale' map (e.g. 1:5,000) is a relatively detailed representation of a relatively small area; a 'small-scale' map (e.g. 1:250,000) generalises over a large area. This notion of generalisation is important in soil science. It is obvious that we can describe the soil in terms of processes and phenomena across continents and landscapes or within horizons and aggregates. Confusingly, the former generalisation (which we would represent on a small scale map) is generally called large scale, and the generalisation over smaller intervals in space is referred to as small scale. Note that wavelet analysis also uses the term 'scale' in an opposite sense to the cartographers. We will use the term 'coarse scale' to refer to generalisations over large distances, represented at small cartographic scales, and 'fine scale' to refer to generalisations over short distances. This section aims to identify some of the key issues in current discussion of spatial and temporal scales in soil science which are relevant to pedometricians. It draws on an admirable review of the subject by Wagenet (1998).

In soil science, we may generalise over coarse or fine scales in both space and time. So, we may consider the soil system in terms of variations over a landscape

(coarse spatial scale) or variations within a profile (fine spatial scale). Similarly, we may consider coarse temporal scales (e.g. the decarbonation of a soil on an old land surface) or fine scales (e.g. hourly fluctuations in the water content of a cropped soil). Scale in space and time becomes of practical importance to soil scientists in general, and pedometricians in particular, when we find that some scales of generalisation are relatively rich in terms of our information about the soil and our ability to describe soil processes with confidence, but these do not necessarily coincide with the scales at which a manager might require predictions of soil behaviour under different scenarios.

The notion of a scale hierarchy is useful here and has been developed by ecologists who have recognised that ecosystems may be characterised as open systems with self-organising processes and that the flux of energy and matter through such a system tends to result in organisation over a hierarchy of spatial scales. This may be seen in soil systems too, where structure can be recognised in the organisation of clay particles by flocculation, the organisation of this material into aggregates, the structure of aggregates and macropores in horizons and so on up to the broad zonal patterns of soil variation at continental scale. Similarly, soil variation over time is organised in a hierarchy of scales from hourly fluctuations driven by rainfall events, the diurnal cycles of temperature, annual weather cycles and coarser scale variations still, associated with long-term cycles of climate and long-term processes of geomorphological change.

Bergkamp (1995) has identified concepts of scale hierarchy in the ecological literature which are relevant to soil scientists. Key to this development is the concept of a hierarchical structure of a complex system with explicit scales in time and space. The basic stable element of such a hierarchy is called the holon. A holon is essentially a distinct subsystem defined principally in terms of processes. It may therefore be defined in line with the concerns of a particular investigation. The principle interactions within a system take place within holons over particular scales of time and space. However, there are interactions within the hierarchical system of holons (holarchy) whereby processes at one scale of time and/or space constrain processes at other scales. Thus, for example, the process of aggregation at one scale both constrains and is constrained by the processes at coarser spatial and temporal scales which cause the differentiation of horizons. In general, however, holons at coarse temporal/spatial scales are more susceptible to the effects of events at comparable scale and may only respond to finer scale processes under conditions which are otherwise stable. Thus, processes of soil development in the profile – illuviation, development of clay skins and formation of pans – are likely to be overridden by coarse (temporal and spatial)-scale effects such as isostasy or climate change.

What are the implications of this analysis for soil research in general and pedometrics in particular? Wagenet (1998) identified three particular issues for research on a particular holon or subset of a holarchy at particular spatial and temporal scale. First, it is important to ensure that the basic unit of soil which is measured in such a study corresponds to the representative elementary volume

(REV) for the scale of interest. That is to say, the basic unit must be large enough to encompass the holons at finer scale which constitute the holon of interest, but not so large as to embrace the variability which occurs at this scale. The second point is that this variability must be adequately characterised. These are important issues for pedometricians and imply the careful choice of sample support and the selection of an appropriate sampling design. Since, as Wagenet (1998) recognises, the precise scale of a study may be constrained as much by technical limitations as the problem of interest, the pedometrician must be involved at all stages in the planning of the research activity. The third point made by Wagenet is that the process models which are used in a study must be appropriate to the scale. Thus, Darcy-type models must give way to two-domain models at some spatial scale.

We have discussed spatial and temporal scale as independent, but in fact they will generally be correlated, coarse spatial scale factors (topography) being correlated with coarse temporal scale factors (isostasy), while fine-scale processes (e.g. at molecular level) may tend to occur over fine temporal scales (Fig. 1.2). This assumption is implicit in Hoosbeek and Bryant's (1992) well-known figure linking scale (explicitly spatial) to appropriate modelling strategies (Fig. 1.3).

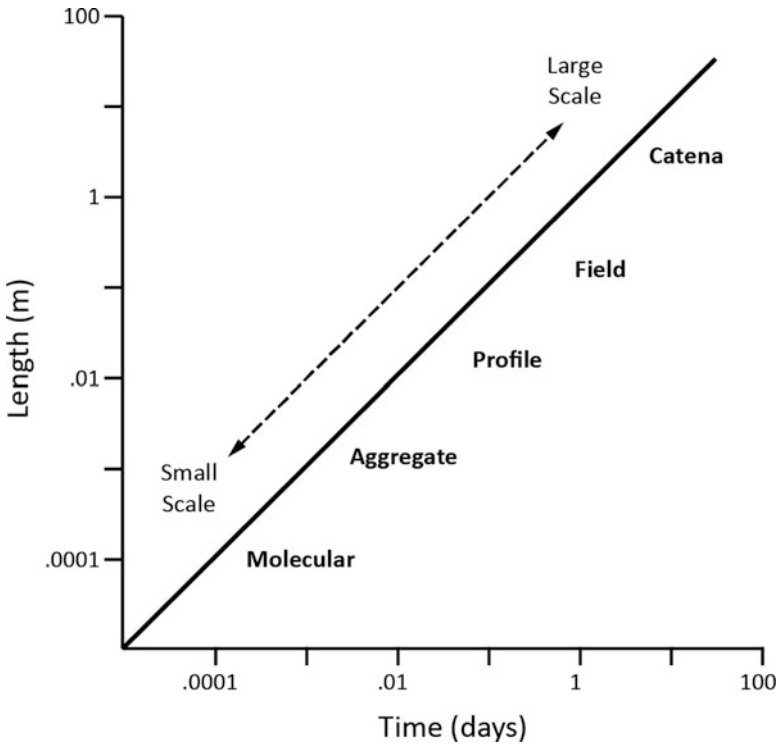


Fig. 1.2 Scales of soil processes (Figure 4 redrawn from Wagenet 1998)

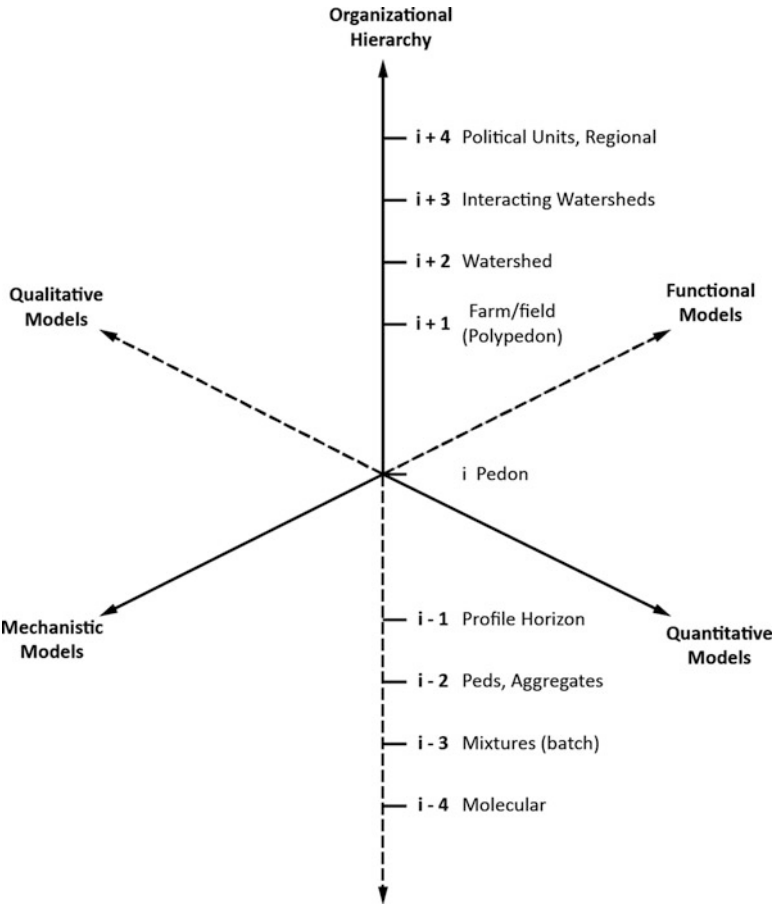


Fig. 1.3 A hierarchical representation of spatial scales of soil processes. At each scale, there is a plane of possibilities of research modelling approaches. The front half of each horizontal level corresponds to Table 1.1 (Redrawn from Hoosbeek and Bryant 1992)

An understanding of this relationship is important, particularly, to return to the key problem of scale, when we wish to translate knowledge and research findings between scales. McBratney (1998) discussed some of these issues and Bierkens et al. (2000) offer a detailed discussion, and we return to this issue in Chap. 15. For the present, however, it is worth noting that important processes may occur off the diagonal of Fig. 1.2. For example, non-equilibrium processes of pesticide-soil interaction occur over short distances (fine spatial scale) but long periods (coarse temporal scale), and this type of process is poorly modelled (Wagenet 1998). Similarly, the weathering of rock occurs at fine spatial scales but over long time periods (McBratney 1998). Rapid (fine temporal scale) processes may also occur

over relatively long distances, e.g. preferential flow through macropores (Wagenet's 1998 example) or global scale but rapid processes such as the Chernobyl incident (McBratney's 1998 example).

In short, pedometricians must be aware of the issues of spatial and temporal scale when collecting and analysing data on soil. A good summary of the key issues is due to Bergkamp (1995) who noted (1) that spatial patterns differ at different scales, (2) changes of spatial patterns differ at different temporal scales, (3) processes controlling changes in spatial pattern differ at different scales and (4) at different scales, different variables are often needed to describe similar processes. These considerations all pose mathematical and statistical challenges to the pedometrician. Some of these are met by the techniques which we describe in this book. Others are challenges for the future.

1.6 Pedometric Principles and Hypotheses

As a summary of this introductory chapter, we outline a few general and specific ideas that underlie the pedometric approach.

1.6.1 General Principles

All the 'metric' disciplines (e.g. biometrics, chemometrics, envirometrics), attempting a quantification of the natural environment and phenomena, will recognise at least three principles:

1. *A quantification principle*
   This suggests that *wherever possible*, *observations should be quantitative rather than qualitative*. The main rationale for this is that the difference between two observations can be calculated. For example, if we say the texture of a soil horizon is 'light' or 'heavy', we might not be able to easily understand how different 'light' is from 'heavy'. If however we describe one horizon as having a clay content of 10 dag/kg and another as 70 dag/kg, we know immediately that the difference is 60 dag/kg. This principle implies that vague descriptors should be quantified.
2. *An uncertainty principle*
   This principle is not the Heisenberg one, but simply further suggests that *wherever possible*, *a measure of uncertainty should accompany quantitative observations*. This may be a statistical uncertainty. For example, if we have a field estimate of horizon texture, then the uncertainty might have a standard deviation of ± 5 dag/kg, whereas if this was a laboratory measurement, the associated standard deviation might be ± 1 dag/kg. Further, if the soil attribute is predicted rather than measured, then the one which is least uncertain is to be preferred.

3. *A modelling parsimony principle (Ockham's razor)*
   This old principle described 700 years ago by William of Ockham states that *when comparing models for describing a phenomenon*, *choose the simplest one*. For example, if we can describe the particle-size distribution of a soil material equally well with a two- or a three-parameter distribution, then the model with two parameters is to be preferred.

We could suggest some principles from other disciplines, e.g. economic or sociological ones, but they do not seem to be in the domain of quantifying the natural environment. They may apply when various uses are made of soil information.

1.6.2 Specific Principles

In addition to these general principles, there are some principles or hypotheses specific to the domain of pedometrics. These may be summarised by a single all-embracing statement:

Soil is a more-or-less continuous, anisotropic, non-stationary, multivariate phenomenon in various states of equilibrium with its environment.

Attempts to quantify soil should recognise these features. They lead to various hypotheses that may be tested or to assumptions commonly underlying pedometric analysis. Therefore, various null hypotheses that could or should be tested by empirical observation at any location include:

Soil is a spatially discontinuous phenomenon.
Soil variation is constant with direction.
The mean of a soil property is independent of position in space or time.
The soil can be adequately described by a single property, e.g., *pH.*
Soil is a function of its current environment.

Finally, it is a difficult philosophical point to contend that these principles and hypotheses lead to a more objective treatment of reality. They are all human constructions. Commonsensically, it would seem that they probably lead to a legitimate empirical testing of humanity's hypotheses about the nature of soil. Nevertheless, as the famous biologist J.B.S. Haldane once wrote, 'If you are faced by a difficulty or a controversy in science, an ounce of algebra is worth a ton of verbal argument'.

References

Addiscott TM, Bailey NJ (1991) Relating the parameters of a leaching model to the percentages of clay and other components. In: Roth K, Flühler H, Jury WA, Parker JC (eds) Field-scale solute and water flux in soils. Birkhaüser Verlag, Basel, pp 209–221

Addiscott TM, Mirza NA (1998) New paradigms for modelling mass transfers in soils. Soil Tillage Res 47:105–109

Addiscott TM, Wagenet RJ (1985) A simple method for combining soil properties that show variability. Soil Sci Soc Am J 49:1365–1369
Addiscott T, Smith J, Bradbury N (1995) Critical evaluation of models and their parameters. J Environ Qual 24:803–807
Adhikari K, Hartemink AE (2016) Linking soils to ecosystem services – a global review. Geoderma 262:101–111
Bárdossy A, Bronstert A, Merz B (1995) 1-, 2- and 3-dimensional modeling of water movement in the unsaturated soil matrix using a fuzzy approach. Adv Water Resour 18:237–251
Bergkamp G (1995) A hierarchical approach for desertification assessment. Environ Monit Assess 37:59–78
Bierkens MFP, Finke PA, de Willigen P (2000) Upscaling and downscaling methods for environmental research. Kluwer, Dordrecht
Buchter B, Hinz C, Fluhler H (1994) Sample size for determination of coarse fragment content in a stony soil. Geoderma 63:265–275
Cohen S, Willgoose G, Hancock G (2010) The mARM3D spatially distributed soil evolution model: three dimensional model framework and analysis of hillslope and landform responses. J Geophys Res 115:F04013
Dexter AR (1976) Internal structure of tilled soil. J Soil Sci 27:267–278
Dijkerman JC (1974) Pedology as a science: the role of data, models and theories in the study of natural soil systems. Geoderma 11:73–93
Dominati EJ, Patterson M, Mackay AD (2010) A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol Econ 69:1858–1868
Doran JW, Parkin TB (1994) Defining and assessing soil quality. In: Doran JW, Coleman DC, Bezdicek DF, Stewart BA (eds) Defining soil quality for a sustainable environment, vol 35. Soil Sci Soc Am, Madison, pp 3–21
Dou C, Woldt W, Bogardi I (1999) Fuzzy rule-based approach to describe solute transport in the unsaturated zone. J Hydrol 220:74–85
FitzPatrick EA (1993) Principles of soil horizon definition and classification. Catena 20: 395–402
Fridland VM (ed) (1976) Soil combinations and their genesis. Amerind Publishing Co, New Delhi
Glasbey CA, Horgan GW, Darbyshire JF (1991) Image analysis and three-dimensional modelling of pores in soil aggregates. J Soil Sci 42:479–486
Gleick J (1987) Chaos. Making a new science. Sphere Books, London
Harvey D (1969) Explanation in geography. Edward Arnold, London
Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1:297–318
Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman and Hall, London
Heuvelink GBM, Burrough PA, Stein A (1989) Propagation of errors in spatial modelling with GIS. Int J GIS 3:303–322
Heuvelink GBM, Webster R (2001) Modelling soil variation: past, present and future. Geoderma 100:269–301
Heuvelink GBM, Schoorl JM, Veldkamp A, Pennock DJ (2006) Space–time Kalman filtering of soil redistribution. Geoderma 133:124–137
Holmgren GGS (1988) The point representation of soil. Soil Sci Soc Am J 52:712–715
Hoosbeek MR, Bryant RB (1992) Towards the quantitative modeling of pedogenesis – a review. Geoderma 55:183–210
Huang J, McBratney AB, Minansy B, Traintafilis J (2017) Monitoring and modelling soil water dynamics using electromagnetic conductivity imaging and the ensemble Kalman filter. Geoderma 285:76–93
Isaaks EH, Srivastava RM (1989) Applied geostatistics. Oxford University Press, New York
Ibáñez JJ, De-Alba S, Bermúdez FF, García-Álvarez A (1995) Pedodiversity: concepts and measures. Catena 24:215–232
Johnson WM (1963) The pedon and the polypedon. Soil Sci Soc Am Proc 27:212–215
Kidd D, Webb M, Malone B, Minasny B, McBratney A (2015) Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. Geoderma Reg 6:7–21

Leeper GW (1956) The classification of soils. J Soil Sci 7:59–64
Manolis GD (2002) Stochastic soil dynamics. Soil Dyn Earthq Eng 22:3–15
McBratney AB (1998) Some considerations on methods for spatially aggregating and disaggregating soil information. Nutr Cycl Agroecosyst 50:51–62
McBratney AB, Odeh IOA (1997) Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77:85–113
McBratney AB, Field DJ, Koch A (2014) The dimensions of soil security. Geoderma 213:203–213
McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Chapman and Hall, London
Milne AE, Webster R, Lark RM (2010) Spectral and wavelet analysis of gilgai patterns from air photography. Aust J Soil Res 48:309–325
Minasny B, McBratney AB (1999) A rudimentary mechanistic model for soil production and landscape development. Geoderma 90:3–21
Minasny B, McBratney AB (2001) A rudimentary mechanistic model for soil production and landscape development, II. A two-dimensional model incorporating chemical weathering. Geoderma 103:161–179
Pachepsky Y, Timlin D, Rawls W (2003) Generalized Richards' equation to simulate water transport in unsaturated soils. J Hydrol 272:3–13
Persson M, Yasuda H, Albergel J, Berndtsson R, Zante P, Nasri S, Öhrström P (2001) Modeling plot scale dye penetration by a diffusion limited aggregation (DLA) model. J Hydrol 250:98–105
Phillips JD (1993) Stability implications of the state factor model of soils as a nonlinear dynamical system. Geoderma 58:1–15
Phillips JD (1998) On the relations between complex systems and the factorial model of soil formation (with discussion). Geoderma 86:1–42
Rossiter DG (1990) ALES: a framework for land evaluation using a microcomputer. Soil Use Manag 6:7–20
Solow AR (1990) Geostatistical cross-validation: a cautionary note. Math Geol 22:637–639
Stockmann U, Padarian J, McBratney A, Minasny B, de Brogniez D, Montanarella L, Hong SY, Rawlins BG, Field DJ (2015) Global soil organic carbon assessment. Glob Food Sec 6:9–16
Storie RE (1933) An index for rating the agricultural value of soils. University of California Coop. Ext. Bull. 556
Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344:734–741
Van Kuilenburg J, de Gruijter JJ, Marsman BA, Bouma J (1982) Accuracy of spatial interpolation between point data on soil moisture capacity, compared with estimates from mapping units. Geoderma 27:311–325
Vanwalleghem T, Stockmann U, Minasny B, McBratney AB (2013) A quantitative model for integrating landscape evolution and soil formation. J Geophys Res Earth 118:1–17
Wagenet RJ (1998) Scale issues in agroecological research chains. Nutr Cycl Agroecosyst 50:23–34
Walter C, McBratney Alex B, Viscarra Rossel RA, Markus JA (2005) Spatial point-process statistics: concepts and application to the analysis of lead contamination in urban soil. Environmetrics 16:339–355
Webster R (1977a) Quantitative and numerical methods in soil classification and survey. Oxford University Press, Oxford
Webster R (1977b) Spectral analysis of gilgai soil. Aust J Soil Res 15:191–204
Webster R, McBratney AB (1989) On the Akaike information criterion for choosing models for variograms of soil properties. J Soil Sci 40:493–496
Whitmore AP (1991) A method for assessing the goodness of computer simulation of soil processes. J Soil Sci 42:289–299
Wischmeier WH, Smith DD (1978) Predicting rainfall erosion losses – a guide to conservation planning. US Department of Agriculture, Agriculture Handbook No 537, Washington, DC

Wischmeier WH, Smith DD 1978b Predicting rainfall erosion losses: guide to conservation planning. USDA, agriculture handbook 537. U.S. Government Printing Office, Washington, DC

Wösten JHM, Pachepsky YA, Rawls WJ (2001) Pedotransfer functions for estimating saturated hydraulic conductivity: implications for modeling storm flow generation. J Hydrol 251:202–220

Yevjevich V (1987) Stochastic models in hydrology. Stoch Hydrol Hydraul 1:17–36

Part II
Statistical Footings

"A certain elementary training in statistical method is becoming as necessary for everyone living in this world of today as reading and writing."

H.G. Wells

Statistical methods are needed to model the natural variation of soil across space and time and its multiple attributes. This section gently introduces some basic and exploratory statistical concepts in the context of soil observation to give an underpinning to any pedometric analysis.

Chapter 2 reminds us about the various kinds of scale on which soil attributes are measured, which in turn lead to various statistical analyses. Chapter 3 deals with the very interesting and under-researched topic of statistical distributions of nominal and continuously scaled soil attributes. A large variety of statistical distributions have been observed for soil properties. Chapter 4 gives an introduction to exploratory methods for looking at multiple soil attributes observed on the same soil object and at the special case when the sum of the variables is constant.

Chapter 2
Soil Statistical Description and Measurement Scales

Thomas F. A. Bishop and Alex. B. McBratney

"Without data, you are just another person with an opinion."

W. Edwards Deming American Engineer (1900–1993)

2.1 Why Statistical Description?

A soil scientist in most instances can only measure and describe soil at a few points in a landscape; at each location, he has ways to describe and measure soil features. These may be based on field observations, e.g. presence of mottling in the subsoil, or a sample may be collected for subsequent laboratory analysis, e.g. clay content. Many different measurements of soil properties can be made, and each of these has what we call a measurement scale which in simplistic terms tells us whether it is measured as numbers, e.g. clay content, or categories, e.g. texture class.

A dataset could be collected for many purposes; it could be for mapping or modelling or testing the impact of management practices. Whatever the ultimate purpose, once measurements have been made, our next step is to describe the dataset, and the approaches we take are largely determined by the measurement scale. We describe the dataset initially to summarise its key features, but also it is useful for looking at outliers which may be errors or unusual values. In many cases, the unusual values, for example, hotspots of soil contamination, are of most interest. From this summary of the data, and how we intend to use the data, we can then decide on which of the wealth of methods presented in this book to use.

Therefore, in this chapter, we focus on:

– The measurement scale
– Summaries of data used to describe a dataset, both numerical and graphical

T.F.A. Bishop (✉) • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: thomas.bishop@sydney.edu.au; alex.mcbratney@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_2

2.2 Scales of Measurements

Random variables may be *discrete* meaning that they have a countable number of possible values or *continuous* having an almost infinite number of possible values. Observation of a soil class, say the soil order in soil taxonomy (Staff 1999), would be a discrete random variable having 12 possible states, whereas the pH of the topsoil would be a continuous random variable, although in reality pH as measured in the field would lie somewhere in between.

Soil properties differ in "how well" they can be measured, i.e. in how much measurable information their measurement scale can provide. There is obviously some measurement error involved in every measurement, which determines the "amount of information" that we can obtain. Another factor that determines the amount of information that can be provided by a variable is its "type of measurement scale". Different soil attributes or soil properties are measured on different scales. These kinds of scales are widely recognised in the statistical literature, and different statistical methods pertain to the different types of scale. There are a number of different scales, and we now present them in order of quantitativeness. There are typically four levels of measurement that are defined in order of increasing quantitativeness: nominal, ordinal, interval and ratio. The first two, nominal and ordinal, correspond to discrete random variables, and interval and ratio correspond to continuous random variables. Different kinds of statistical procedures pertain to discrete and continuous random variables. These terms are also synonymous with categorical and numerical variables.

Let us elaborate the four measurement scales a little further.

2.2.1 Measurement Scales

2.2.1.1 Nominal

Nominal (or sometimes unranked multistate) variables allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories. For example, all we can say is that two soil samples are different in terms of variable A (e.g., they have a different pedality), but we cannot say which one "has more" of the quality represented by the variable. Typical examples of nominal variables are soil class or type of ped. This is the simplest type of scale and refers to different classes.

2.2.1.2 Ordinal

Ordinal variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still they do

not allow us to say "how much more." A typical example of an ordinal variable is small, medium or large peds. For example, we know that medium peds are larger than small ones, but we cannot say that it is, for example, 20% larger. Also this very distinction between nominal, ordinal and interval scales itself represents a good example of an ordinal variable. For example, we can say that nominal measurement provides less information than ordinal measurement, but we cannot say "how much less" or how this difference compares to the difference between ordinal and interval scales.

2.2.1.3 Interval

On interval measurement scales, one unit on the scale represents the same magnitude on the attribute being measured across the whole range of the scale. For example, temperature in Celsius is measured on an interval scale, and then a difference between 10 and 11°C would represent the same difference in temperature as would a difference between 40 and 41°C. Interval scales do not have a "true" zero point, however, and therefore it is not possible to make statements about how many times higher one score is than another; 40°C is not four times as hot as 10°C. Another example of an interval scale attribute is pH.

2.2.1.4 Ratio

Ratio-scaled variables have a fixed interval between values and also an absolute zero. For example, a soil temperature of 303 K is 303/273 times a soil temperature of 273 K (which is not the case if we describe them as 30°C and 0°C). Statistically, in most situations, interval- and ratio-scaled variables are treated much the same way. Most laboratory-derived attributes and some field ones such as soil depth are on the ratio scale.

It is important to recognise that there is a hierarchy implied in the level of measurement scale (Fig. 2.1). At lower levels of measurement, assumptions tend to be less restrictive, and data analyses tend to be less sensitive. At each level up the hierarchy, the current level includes all of the qualities of the one below it and adds something new. In general, it is desirable to have a higher level of measurement (e.g. interval or ratio) rather than a lower one (nominal or ordinal). A common problem with conventional soil description systems is that variables are recorded as ordinal-

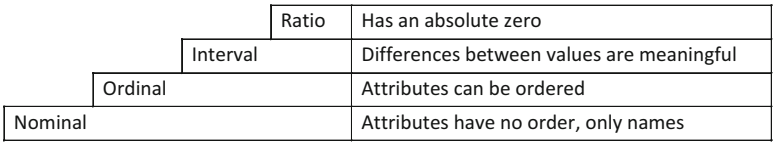


Fig. 2.1 Staircase of measurement scales

scaled variables when they could be interval or ratio. For example, the size of peds is measured and then allocated to a class or category removing some information from the description.

We should also realise that this scheme is a simplification. Because as we shall see later – tools such as fuzzy sets or multivariate methods allow us to put apparently unranked variables onto continuous scales by calculating taxonomic distance.

2.3 Numerical Summaries of Soil Data

Given a dataset of soil measurements, we need statistics to summarise its key features. The statistics we use depend on the type of the measurement scale of the data we have.

2.3.1 Summaries of Categorical Data

Categorical data have been measured on the ordinal and nominal measurement scales. Categorical data is best summarised in terms of the frequency, f, of observations found within each class. It is usually presented in the form of a table, a frequency table. The example dataset we will use in this chapter is from the catchment of Muttama Creek which is located in Southeastern Australia (Fig. 2.2). Further details about the catchment are given in Orton et al. (2016). One hundred seventy-four sites were sampled to a depth varying between 1 and 2 m, and various soil properties were measured at a subset of sites (Fig. 2.2).

In our first analysis, we wish to see how many sites are located within each geological unit which we have classified based on silica content. Gray et al. (2016) found that classifying geological classes on the silica content improved the usefulness of geological maps as predictors. Therefore we wish to make sure we have adequately sampled each of the silica classes. In Table 2.1 we present a frequency table for each of the classes; we also present the results in terms of the percentage of sites that fall within each class. The advantage of scaling the frequencies to a % is that it makes the results scale independent, or independent of n, the number of samples, enabling us to compare results to other datasets if needed. We also include another column, the % of area in the catchment covered by each class. This is based on the geological map (Warren et al. 1996) on which the silica classes were extracted. By using the % frequency we can easily see how well the sampling matches the areal proportions in the geological map. As can be seen by the results, the sample design does include any sites within the siliceous upper class, but it only covers 0.1% area so it is not a great concern. The other silica classes have at least four observations in each class so they have some representation for use in further modelling. In terms of being proportional to the area of each silica class in

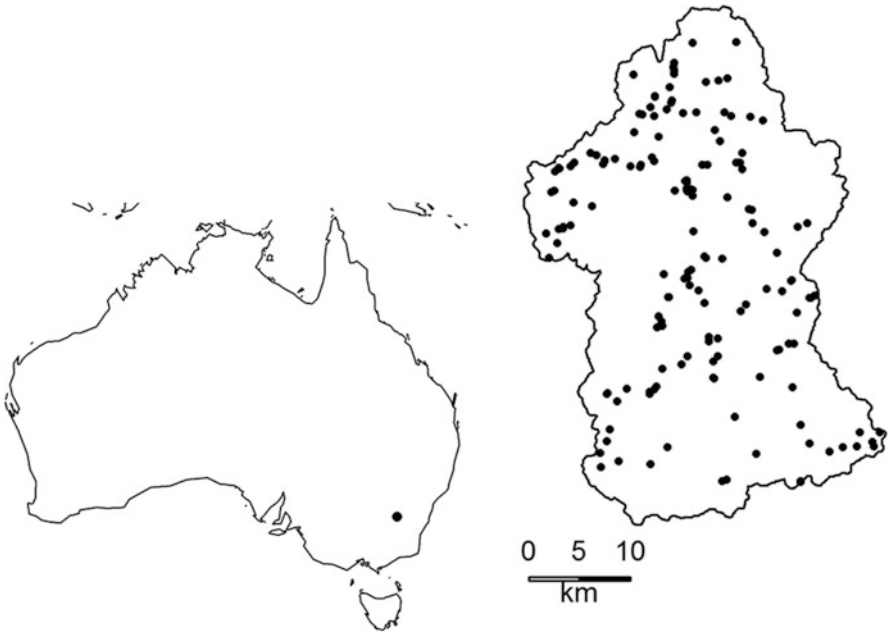


Fig. 2.2 Map of Australia with *solid circle* showing the location of the study area (*left*) and soil survey locations within Muttama Creek catchment (*right*)

Table 2.1 Frequency table for soil survey locations within each silica class

Silica class	Frequency	% frequency samples	% area of geology map
Mafic (Mafc)	36	20.7	24.4
Siliceous mid (SlcM)	42	24.1	13.7
Siliceous upper (SlcU	0	0	0.1
Siliceous lower (SlcL	80	46.0	55.6
Intermediate lower (IntL)	4	2.3	1.0
Intermediate upper (IntU)	12	6.9	5.2
Total	174	100.0	100.0

the geology map, we can see that the mafic, intermediate upper and intermediate lower classes are very close while the other classes are over- or under-sampled.

2.3.2 Summaries of Continuous Variables

Continuous data has been measured on the ratio and interval scales. In the case of continuous variables, we need to consider three types of statistics. This is best illustrated with an example. The figure below presents the histograms for electrical

conductivity (EC) measured using a 1:5 soil/water extract for the 0–30 cm and 60–90 cm layers for 55 soil sites in Muttama Creek catchment. While 174 sites have been sampled in Muttama Creek, only 55 had EC measured on them.

We can see that visually the distributions are quite different and we wish to use some metrics to summarise this. The first is one that describes a typical value for the dataset; or a measure of centre of the distribution, a simple estimate would be the value that corresponds to the highest bar in the histograms. This would be ~0.05 dS m^{-1} for the 0–30 cm layer and ~0.03 dS m^{-1} for the 60–90 cm layer. Of equal interest is the variation in the dataset or a measure of spread. A simple estimate would be the difference between largest and smallest values in the histogram. This would be ~0.18 dS m^{-1} for the 0–30 cm layer and ~0.4 dS m^{-1} for the 60–90 cm layer. Finally, the shapes of the histograms are quite different, and a statistic that describes this can be quite useful; key issues to consider are how symmetrical the distribution is or are there any unusual values. For example, there is one large observation at ~0.4 dS m^{-1} for the 60–90 cm layer.

In the next section we present summary statistics that relate to these three features. Another concept to consider when presenting data is to distinguish between statistics we calculate from samples and those from populations. Continuing with our example, we may wish to know a statistic that represents the typical value for EC across the catchment or perhaps for each silica class in the Muttama Creek catchment. The true value is the population statistic which we would obtain if we could measure all of the soil in the catchment. In reality, we only have a subset of samples from which to estimate the statistics; this is our sample statistic. A large part of statistics is about using our limited samples to make an inference about the true population statistics. For the purposes of this chapter, we present sample statistics but make mention of differences in notation and equations that exist for population statistics.

2.3.2.1 Measures of Centre

The most common measure of centre is the mean which is for the samples we denote as $\bar{y}$ and for populations we denote as μ. For the sample mean, the equation is:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \tag{2.1}$$

where y_i is the ith observation and n is the number of observations.

We now introduce another property in our case study, clay content which was measured at 77 sites in Muttama Creek catchment. The mean values for EC and clay in the 0–30 cm and 60–90 cm layers are presented in Table 2.2 below. There is not much difference in the mean EC between the two layers, but there is a large increase in the mean clay content with depth as would be expected as many of the soils in Muttama Creek have duplex profiles (i.e. an abrupt increase of clay content between the topsoil and subsoil). One issue with the mean is that it is susceptible

Table 2.2 Measures of centre

Property	EC (dS m^{-1})		Clay (%)	
Depth (cm)	0–30	60–90	0–30	60–90
Mean	0.071	0.061	22.66	47.26
Median	0.066	0.039	21.43	48.16

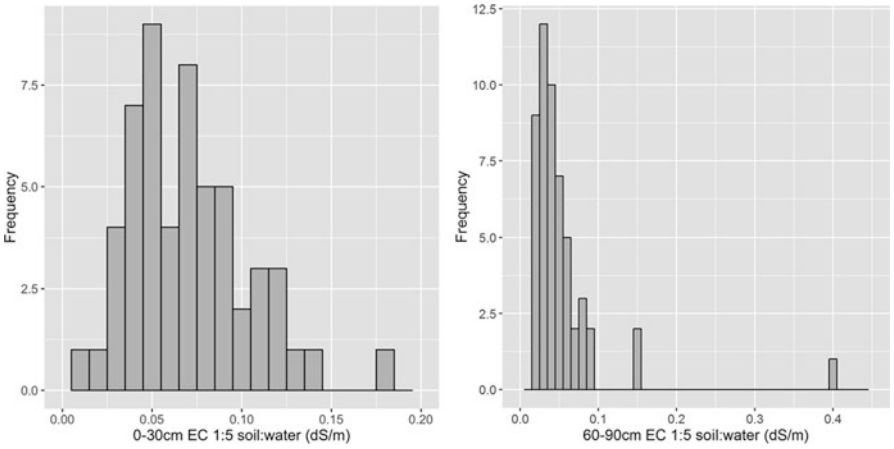


Fig. 2.3 Histogram of EC 1:5 soil/water extract for 0–30 cm layer (*left*) and 60–90 cm layer (*right*)

to a small number of outliers; these may be erroneous measurements or actually real observations that relate to unusual areas, and in case of EC, this could be a small hotspot where saline groundwater is close to the surface. In these situations the mean does not give a very good estimate of the typical value in a dataset as it can be inflated or deflated by a small number of outliers. This will be elaborated further in the coming sections.

Another option is the median which is the middle value in the dataset and is calculated by ordering the observations from smallest to largest. The sample median, $\tilde{y}$, and population median, M, is calculated by ordering the observations and using the equation below to work out the location of the middle value:

$$\text{median} = \frac{(n+1)\,\text{th}}{2}\text{ordered value}. \tag{2.2}$$

Unlike the mean, a few large or small values do not affect its value. Examining the data in Table 2.2 for the clay values, we can see that the mean and medians are quite similar, meaning that the distribution is symmetrical. The EC 0–30 cm layer also has a similar mean and median; however, they are quite different for the 30–60 cm layer. The median is much smaller than the mean which makes sense if we examine the histogram (Fig. 2.3) which shows that there are a few large values of EC which inflate the mean value, so much so that the mean value is located at a larger value past the peak in the histogram.

2.3.2.2 Measures of Variation

Equally important to the mean and median are measures that describe how much variation there is in a dataset. The most commonly used is the variance as shown by the formula below:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \tilde{y})^2}{n-1}. \tag{2.3}$$

One can see it is calculated simply by the average squared difference of each observation from the mean. In the formula above we calculate the sample variance, s^2, and for the population variance we use σ^2, and the denominator has N rather than n-1, where N is population size and n is the sample size. The difference ensures that the sample variance gives an unbiased estimate of population variance. One issue with the variance is the units are squared of the variable being measured; in our EC example, this would be (dS m^{-1})^2. This does make it difficult to interpret; so by taking the square root of the variance we get the standard deviation which is presented in Table 2.3. The advantage is that it is on the scale of the measurements. Based on the standard deviation, we can say that EC is more variable than clay content in our data, but as shown by the minimum and maximum values, the range of possible values is much greater for EC than clay (Table 2.3); the means are also quite different between depths and soil properties.

Due to this, when comparing variables with different magnitudes, a useful statistic is the coefficient of variation (*CV*) which is a measure of the variation as estimated by the standard deviation relative to the sample mean; it is calculated by

$$\text{CV} = \frac{s}{\bar{y}}. \tag{2.4}$$

Strictly, the coefficient of variation does apply to ratio-scaled variables but not to interval-scaled ones. For example, if we take five soil temperature observations measured in Celsius (5, 10, 11, 7, 13), the apparent *CV* is 0.35, whereas the same set of observations measured in Kelvin (278, 283, 280, 284, 286) has a *CV* of 0.011.

Table 2.3 Measures of variation

Property	EC (dS m^{-1})		Clay (%)	
Depth (cm)	0–30	60–90	0–30	60–90
s	0.033	0.077	10.76	17.19
Q_1	0.046	0.029	16.54	31.42
Q_3	0.089	0.060	27.86	59.61
IQR	0.043	0.031	11.32	28.19
CV	0.468	1.268	0.471	0.371
Minimum	0.013	0.016	4.52	10.38
Maximum	0.177	0.447	54.34	78.73

As shown in Table 2.3, using the coefficient of variation, both EC and clay in the 0–30 cm layer have the same amount of variability rather than EC being three times as variable as clay if we use standard deviation to assess variation. The EC 60–90 cm layer is still the most variable but one issue with using the standard deviation, and also the coefficient of variation is that for asymmetrical distributions these may not be useful measures. This is because the variation is calculated by the squared difference from the mean meaning that large values as shown in Fig. 2.3 will inflate the estimate of standard deviation.

This is where measures based on the order are useful; analogous to the median is the interquartile range (*IQR*) which describes the difference between the 1st (Q_1) and 3rd quartile (Q_3) or the middle 50% of the distribution. This means the values outside this range (the lower and upper quarter of the distribution) do not impact on the estimate of the variation. Examining the *IQR* alone we can see that for EC the 0–30 cm layer is more variable than the 60–90 cm layer. This is a very different result than what is obtained by standard deviation and CV.

2.3.3 Other Statistics

Until now we have qualitatively assessed the datasets for symmetry and discussed the effect of a small number of larger (though could be small values) on measures of centre and variation. This is an important characteristic as the normal distribution is symmetric and many of the statistical methods presented in this book are based on the assumption that the data follows a normal distribution. One estimate of the asymmetry of the distribution is the skewness. A positive value means the distribution is skewed or there are a small number of observations at the larger end of the distribution, creating a pointy or skewed distribution; a negative skew means the mirror image of this but the pointy end is towards small values. There is evidence of skewness in the EC 30–60 cm layer both by its histogram (Fig. 2.3) and the skewness value of 4.04 (Table 2.4). A heuristic rule proposed by Webster and Oliver (2001) for considering whether to transform a variable is when the skewness is outside the interval of [−1, 1]. In the case of the soil properties presented in Table 2.4, using this rule we would only consider transforming EC for the 30–60 cm layer.

One issue to consider is the effect of outliers; as noted previously, the mean and standard deviation are sensitive to outliers. The skewness is no different and analogous to the median, and IQR is the octile skew which was compared by Brys et al. (2003) with other measures of skew and as far as we are aware first applied to

Table 2.4 Measures of symmetry

Property	EC (dS cm^{-1})		Clay (%)	
Depth (cm)	0–30	30–60	0–30	30–60
Skewness	0.86	4.04	0.77	−0.03
Octile skew	0.20	0.43	0.11	0.07

Table 2.5 To transform or not to transform

		Octile skew	
Skewness	Interval	Inside [−0.2, 0.2]	Outside [−0.2, 0.2]
	Inside [−1, 1]	No transform	Possibly transform
	Outside [1, 1]	Possibly transform	Transform

soil data by Lark et al. (2006). It is less sensitive to outliers than the skewness, and Lark et al. (2006) proposed an equivalent heuristic of rule 0.20 for octile skew where we should consider transformation if the value is outside the range [−0.2, 0.2]. The value of the octile skew is evident in situations where a small number of observations result in a skewness greater than 1.0, but the issue here will be that the decision to transform the *entire* distribution is based on a few outliers. The same dataset may have an octile skew less than 0.2 which would mean the small number of larger values is insufficient to ruin the symmetry of the overall distribution so a transformation may not be recommended. In this situation the decision to transform depends on what part of the distribution is being modelled or of interest, the overall distribution or a few outliers. A decision matrix on whether to transform or not is presented in Table 2.5 based on the values of the skewness and octile skew.

2.4 Graphical Summaries of Soil Data

Like numerical summaries, how we graphically summarise data depends on the measurement scale. The advantage of graphical summaries is that we can view the shape of the entire distribution unlike summary statistics which describe one feature only, i.e. the typical value or the variation in the data.

2.4.1 Categorical Data

The figure below presents the frequency table from Table 2.1 as a bar plot where the y-axis is the frequency and the x-axis is the different silica classes. We clearly see the range of classes and their associated frequencies, and identify the most sampled class, Siliceous Lower (SicL) (Fig. 2.4).

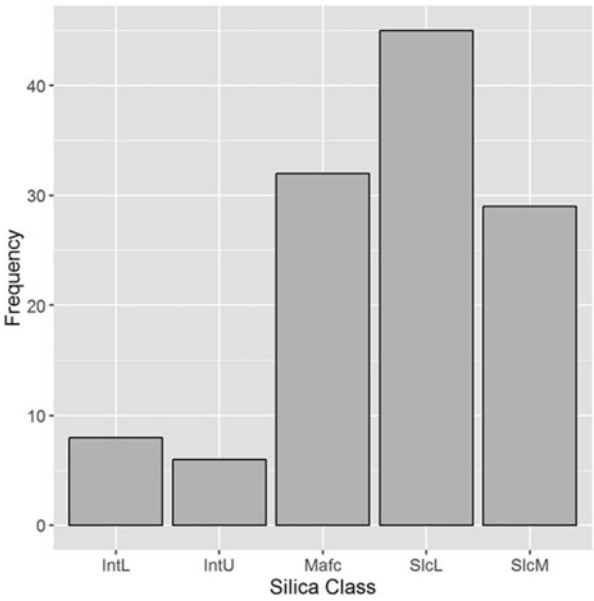


Fig. 2.4 Bar plot of frequencies within each silica class

2.4.2 Numerical Data

The most common type of graphical summaries for numerical data is boxplots and histograms which have advantages and disadvantages depending on the situation which will be elaborated on further in the following sections.

2.4.2.1 Histogram

A histogram is essentially a bar plot of frequencies where frequencies are the number of observations within a range of values, called bins. In the figures below (Fig. 2.5), we present histograms of clay content where the bin size was 6%. Considering the histograms in terms of the summary statistics, we can see that the 0–30 cm layer is positively skewed which is supported by the skewness value of 0.77 (Table 2.4), whereas the distribution for the 60–90 cm layer is quite symmetrical (skewness $= -0.03$ from Table 2.4). When comparing the distribution for EC (Fig. 2.3) and clay for the 0–30 cm layer, it can be seen that the positive skew for clay is weak compared to the EC which has a skewness of 4.04 (Table 2.4).

An issue with using histograms is the size of the dataset which is best illustrated by examining histograms for each silica class (Fig. 2.6). The number of observations within each silica class where clay content was measured ranged between 3 and 29. As shown by the histograms, when the number of observations is small, they

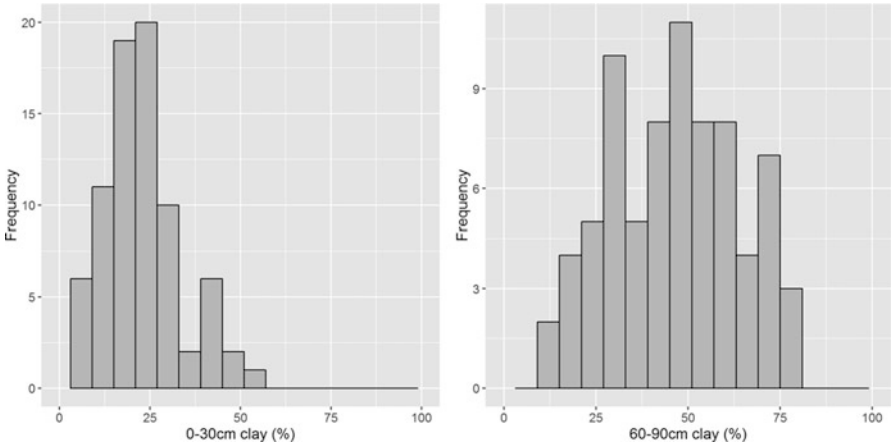


Fig. 2.5 Histograms of clay content (%) for 0–30 cm layer (*left*) and 60–90 cm layer (*right*)

are not particularly useful; this is especially clear for the intermediate lower and intermediate upper silica classes. Clearly, alternative approaches are needed when the dataset size is small.

2.4.2.2 Boxplots

Below boxplots are presented for the clay content (Fig. 2.7). They have a number of key features:

- The width of the box is determined by the 1st (Q_1) and 3rd quartiles (Q_3).
- The line down the middle is the median.
- The length of the whiskers is determined by the largest or smallest observation that falls within the bounds determined by [Q_1–1.5IQR] and [Q_3 + 1.5IQR].
- Observations outside the bounds of [Q_1–1.5IQR] and [Q_3 + 1.5IQR] are identified as outliers and plotted individually.

Boxplots do provide less information about the distribution of a variable than histograms. For example, the boxplot for the 60–90 cm layer does show approximately equal whisker lengths which indicates a symmetrical distribution but not the actual shape as shown by the histogram (Fig. 2.5). One useful feature is that boxplots automatically identify outliers which can be examined more closely. However, since a boxplot provides less information than a histogram, it is recommended that when sufficient data is available, a histogram is used. For smaller datasets, a boxplot should be used which is illustrated by comparing the histograms of clay for each class (Fig. 2.6) with boxplots of clay for each silica class (Fig. 2.8). Key features of the distribution such as symmetry are easier to determine for the boxplots as compared to the histograms.

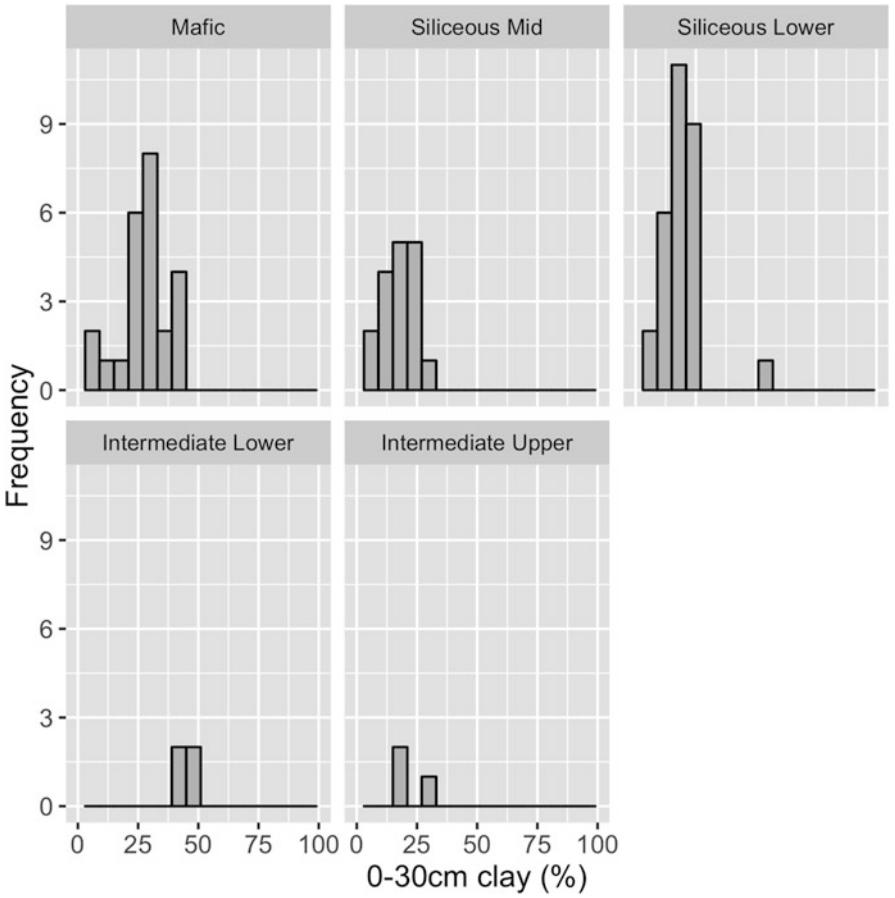


Fig. 2.6 Histograms of clay content (%) for 0–30 cm layer for each silica class

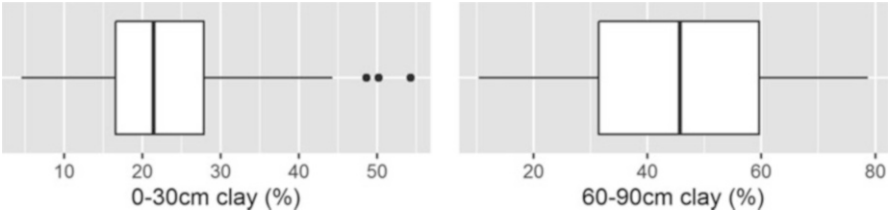


Fig. 2.7 Boxplots of clay content (%) for 0–30 cm layer (*left*) and 60–90 cm layer (*right*)

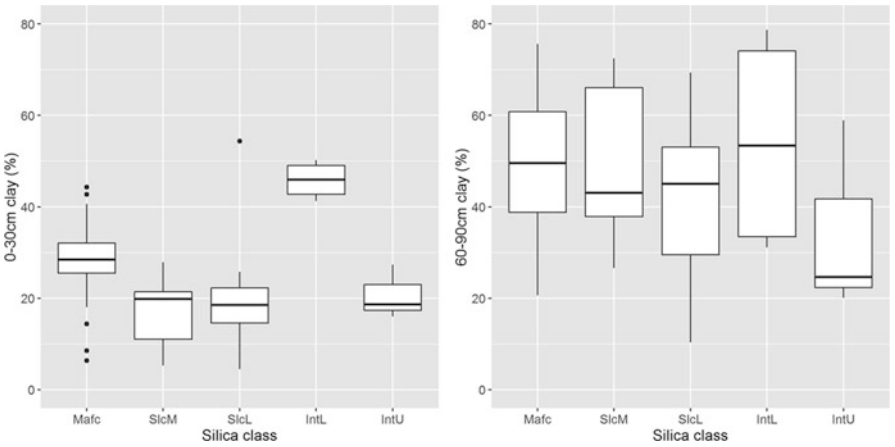


Fig. 2.8 Boxplots of clay content (%) for 0–30 cm layer and (*left*) 60–90 cm layer (*right*) for each silica class

It should be noted that for the intermediate lower and intermediate upper silica classes the number of observations is three or less, and in these cases, it would probably be wisest to plot the raw data if a graphical summary was required.

As we have discussed, the wisest choice of graphical summary is largely determined by the dataset size. While these are only guidelines, some heuristic rules could be:

– 1–5 observations: Plot raw data.
– 6–20 observations: Use a boxplot.
– 20+ observations: Use a histogram.

2.5 Concluding Remarks

We have presented different measurement scales of observations and focused on statistics and graphical methods to describe such data with a focus on categorical and continuous data. In the case of categorical data the choices are limited but this is not the case for continuous data. In this case, the type of numerical summary of the data depends on how skewed the data is; for symmetrical distributions, the classical measures, mean and variance, are best, but when the data is skewed, other measures are more useful such as the median and interquartile range. A decision matrix was presented based on the skewness and octile skew to determine whether the data should be transformed to make the distribution symmetrical (Table 2.5). Finally, we presented two common graphical summaries, the boxplot and histogram, the choice of which to use is largely determined by the dataset size.

References

Brys G, Hubert M, Struyf A (2003) A comparison of some new measures of skewness. In: Dutter R, Filzmoser P, Gather U, Rousseeuw PJ (eds) Developments in robust statistics: international conference on robust statistics 2001. Physica-Verlag HD, Heidelberg, pp 98–113

Gray JM, Bishop TFA, Wilford JR (2016) Lithology and soil relationships for soil modelling and mapping. Catena 147:429–440

Lark RM, Bellamy PH, Rawlins BG (2006) Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. Geoderma 133(3–4):363–379

Orton TG, Pringle MJ, Bishop TFA (2016) A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262:174–186

Staff SS (1999) Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys, 2nd edn. Natural Resources Conservation Service, USDA, Washington, DC

Warren AYF, Gilligan LB, Raphael NM (1996 Cootamundra 1:250 000 Geological Sheet SI/55–11,. Geological survey of New South Wales, Sydney, Australia

Webster R, Oliver MA (2001) Geostatistics for environmental scientists. Wiley, Chicester

Chapter 3
Statistical Distributions of Soil Properties

Alex. B. McBratney, Budiman Minasny, Irina Mikheeva, Melissa Moyce, and Thomas F. A. Bishop

"The true logic of this world is in the calculus of probabilities."

James Clerk Maxwell

3.1 Statistical Distributions

A basic idea concerning collections of soil observations is to obtain statistical parameters from the data distribution. In soil, we recognise two kinds of statistical distributions relating to discrete or continuous random variables.

3.1.1 Discrete Distributions

To understand distributions of discrete random variables, we consider the Bernoulli trials process, named after the Swiss mathematician James Bernoulli. It is one of the simplest yet most important random processes in probability. Essentially, the process is the mathematical abstraction of coin tossing, but because of its wide applicability, it is usually stated in terms of a sequence of generic trials that satisfy the following assumptions:

A.B. McBratney (✉) • B. Minasny • M. Moyce • T.F.A. Bishop
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: alex.mcbratney@sydney.edu.au; budiman.minasny@sydney.edu.au;
thomas.bishop@sydney.edu.au

I. Mikheeva
Institute of Soil Science and Agrochemistry of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation
e-mail: pulya80@ngs.ru

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_3

1. Each trial has two possible outcomes, generically called success and failure.
2. The trials are independent. Intuitively, the outcome of one trial has no influence over the outcome of another trial.
3. On each trial, the probability of success is p and the probability of failure is $1 - p$.

Mathematically, we can describe the Bernoulli trials process with a sequence of indicator random variables:

$$I_1, I_2, I_3, \ldots$$

An indicator variable is a random variable that takes only the values 1 and 0, which in this setting denote success and failure, respectively. The jth indicator variable simply records the outcome of trial j. Thus, the indicator variables are independent and have the same density function:

$$P\left(I_j = 1\right) = p, P\left(I_j = 0\right) = (1 - p)$$

Thus, the Bernoulli trials process is characterised by a single parameter p.

The mean and variance of a generic indicator variable I with $P(I = 1) = p$ are

$$E(I) = p$$
$$\text{var}(I) = p\,(1 - p)$$

Putting this now in a soil science context, we are interested in surveying an area at random to discover the proportion of a certain important soil class, for say wine production, the Terra Rossa, or the sites of soil contamination, say where total Pb >300 mg/kg, or some other particular soil feature. We do this by taking points at random within the region of interest by generating uniformly distributed random co-ordinates (x,y) in northings and eastings. (Alternatively we could make observations systematically, if we really believe the soil is distributed randomly.) If we observe the soil of interest, we can call this a *successful soil observation* and we record 1; if not we record 0. In this case the total number of soil observations is equal to the number of trials.

Recalling that the number of subsets of size k from a set of size n is given by the binomial coefficient:

$$C\,(n, k) = n!/\,[k!\,(n - k)!]$$

The discrete random variable X_n, the number of successful soil observations in the first n observations, has the *binomial distribution*:

$$P\,(X_n = k) = C\,(n, k)\,p^k(1 - p)^{n-k} \text{for } k = 0, 1, \ldots, n.$$

with parameters n and p and mean and variance

$$E(I) = np$$
$$\text{var}(I) = np\,(1-p)$$

The discrete random variable Y_k, the number of observations needed for k successful soil observations, has the *negative binomial distribution*:

$$P\,(Y_k = n) = C\,(n-1, k-1)\,p^k(1-p)^{n-k} \text{ for } n = k, k+1, k+2, \ldots$$

with two parameters, the number of successful soil observations k and the success probability p. Its mean and variance are

$$E\,(Y_k) = k/p$$
$$\text{var}\,(Y_k) = k\,(1-p)\,/p^2$$

The binomial and negative binomial sequences are inverse to each other in the sense that

$$X_n \geq k \text{ if and only if } Y_k \leq n.$$

The discrete random variable Y_1 is the number of observations needed to get the first success, which has the *geometric distribution*:

$$P\,(Y = n) = p(1-p)^{n-1} \text{for } n = 1, 2, \ldots,$$

with parameter p and with mean and variance

$$E(Y) = 1/p.$$
$$\text{var}(Y) = k\,(1-p)\,/p^2.$$

So we can see that this is a special case of the negative binomial distribution with $k = 1$.

So, for example, if the conditions of a Bernoulli trial are met and if, in reality, the proportion p of an area being Terra Rossa is 0.1 (10 %) and we make 100 soil observations, then we can expect that the number of successful of soil observations will be 10 with a variance of 11.1. The number of observations expected to find the first successful soil observation is $1/0.1 = 10$ with variance of 90. If we need to find 20 successful soil observations, then we can expect to make 200 observations with a variance of 1,800.

In soil survey we would generally be interested in all the classes observed. Let us say for any soil observation there are k possible classes, then we have a *multinomial trials process*, a sequence of independent, identically distributed random variables:

$$U_1, U_2, \ldots,$$

each taking k possible values. Thus, the multinomial trials process is a simple generalisation of the Bernoulli trials process (which corresponds to $k = 2$). For simplicity, we will denote the outcomes by the integers 1, 2, ..., k, (N.B. this is a form of coding only and no metric is associated with these integers). Thus, we will denote the common density function of the observation variables by

$$p_i = P\left(U_j = i\right) \text{ for } i = 1, 2, \ldots, k \text{ (and for any } j).$$

Of course $p_i > 0$ for each i and $p_1 + p_2 + \cdots + p_k = 1$.

As with our discussion of the binomial distribution, we are interested in the variables that count the number of times each outcome occurred. Thus, let Z_1 as the number of times in which outcome 1 occurs, Z_2 as the number of times in which outcome 2 occurs, etc.

Note that

$$Z_1 + Z_2 + \cdots + Z_k = n,$$

so if we know the values of $k - 1$ of the counting variables, we can find the value of the remaining counting variable. As with any counting variable, we can express Z_i as a sum of indicator variables:

$$Z_i = I_{i1} + I_{i2} + \cdots + I_{in} \text{ where } I_{ij} = 1 \text{ if } U_j = i \text{ and } I_{ij} = 0 \text{ otherwise.}$$

Recalling the definition of the multinomial coefficient

$$C\left(\mathrm{n}; j_1, j_2, \ldots, j_k\right) = n!/\left(j_1! j_2! \cdots j_k!\right) \text{ for positive integers } j_1, j_2, \ldots, j_k \text{with}$$
$$j_1 + j_2 + \cdots + j_k = n.$$

For positive integers j_1, j_2, ..., j_k with $j_1 + j_2 + \cdots + j_k = n$

$$P\left(Z_1 = j_1, Z_2 = j_2, \ldots, Z_k = j_k\right) = C\left(n; j_1, j_2, \ldots, j_k\right) p_1^{j2} p_2^{j2} \ldots p_k^{\mathrm{jk}}.$$

The distribution of (Z_1, Z_2, ..., Z_k) is called the *multinomial distribution* with parameters n and p_1, p_2, ..., p_k and mean and variance

$$E\left(Z_i\right) = np_i$$
$$\operatorname{var}\left(Z_i\right) = np_i\left(1 - p_i\right)$$

which we can see is a generalisation of the binomial result. Even though the events are independent, there is a correlation between the number of observed classes because the sum is constrained to be n, thus the covariance and correlation for distinct i and j

$$\operatorname{cov}\left(Z_i, Z_j\right) = -n\, p_i p_j.$$
$$\operatorname{cor}\left(Z_i, Z_j\right) = -\left\{p_i p_j / \left[\left(1 - p_i\right)\left(1 - p_j\right)\right]\right\}^{1/2}.$$

These kinds of results are not extensively used in the soil science literature, perhaps because we are more interested in how good our estimates of the various p_i are.

Discrete distributions such as the negative binomial (also known as the Pascal or Polya distributions) and the Poisson distribution have been used to describe the spatial pattern of viable weed seeds in soil (Mulugeta and Stoltenberg 1997), root contact points on horizontal plane sections (Wulfsohn et al. 1996) and the degree of clumping of calcium phosphate particles (Kirk and Nye 1986). These are really examples of *spatial point processes*.

3.1.2 Continuous Distributions

Collections of interval and ratio-scaled variables can be described by various continuous probability distributions; the most commonly known is the normal or Gaussian distribution. Whether this is the most common for soil variables is questionable. Many soil variables show distinctly long-tailed or positively skewed distributions. These are often modelled by the lognormal distribution, but there is a range of other possibilities that are rarely investigated. So it is worthwhile here to elaborate some of the more common continuous distributions, realising that we cannot be exhaustive because there are hundreds of them. The main ones are given in Evans et al. (2000), and we only present a small subset of those here.

3.1.2.1 Normal or Gaussian Distribution

The normal distribution was originally developed as an approximation to the binomial distribution when the number of trials is large and the Bernoulli probability p is not close to 0 or 1 (Evans et al. 2000). Parameters are estimated either using the maximum likelihood method or using an unbiased estimator. Just as in the case of discrete distributions, a nonnegative function can often be scaled to produce a probability density function (PDF).

The normal PDF is (Fig. 3.1)

$$\sigma\left(2\pi\right)^{1/2} / \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

The normal distribution has been used in soil science to describe many soil properties, including penetrometer resistance and spatial bulk density variability (Utset

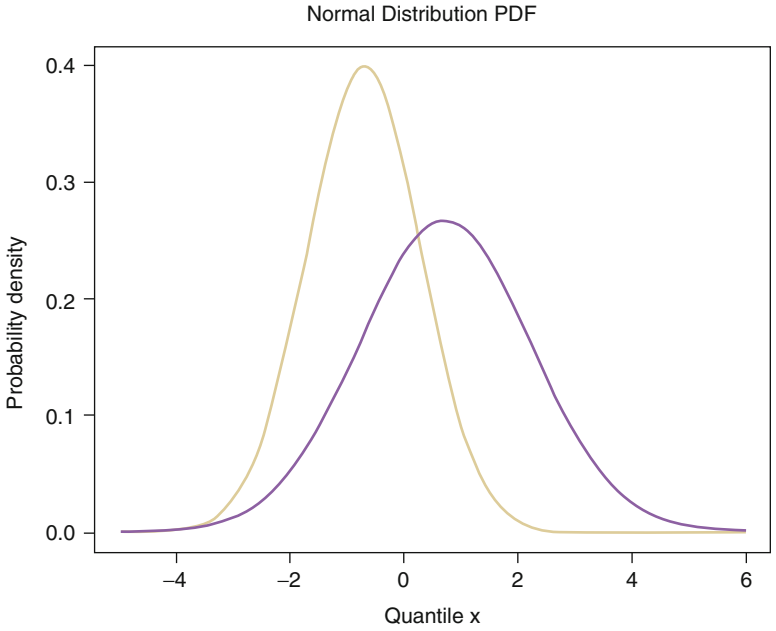


Fig. 3.1 The probability density function of the normal distribution

and Cid 2001). These examples represent the tendency of the spatial variability of static soil physical properties to follow normal probability distributions. However, non-normality is also encountered in soil properties. One strategy to make non-normal data resemble normal data is by using a transformation. There is no dearth of transformations in statistics; the issue is which one to select for the situation at hand. Unfortunately, the choice of the 'best' transformation is generally not obvious.

This was recognised in 1964 by G.E.P. Box and D.R. Cox, who proposed a useful family of power transformations. These transformations are defined only for positive data values. This should not pose any problem because a constant can always be added if the set of observations contains one or more negative values.

The Box-Cox power transformations are given by

$$x(\lambda) = \frac{(x^{\lambda}-1)}{\lambda} \quad \lambda \neq 0$$
$$x(\lambda) = \ln(x) \quad \lambda = 0$$

Given the vector of data observations $\mathbf{x} = x_1, x_2, \ldots x_n$, one way to select the power λ is to find the λ that maximises the logarithm of the likelihood function:

$$f(x,\lambda) = -\frac{n}{2}\ln\left[\sum_{i-1}^{n}\frac{(x_i(\lambda)-\bar{x}(\lambda))^2}{n}\right] + (\lambda-1)\sum_{i=1}^{x}\ln x_i$$

where

$$\bar{x}(\lambda) = \frac{1}{n}\sum_{i=1}^{x} x_i(\lambda)$$

is the arithmetic mean of the transformed data.

Another type of transformation includes the Johnson system, which involves three main transformations: LN, lognormal; SB, log ratio; and SU, hyperbolic arcsine (Carsel and Parrish 1988). Examples of properties where a log transform produced normally distributed data include studies of soil hydraulic properties (Russo and Bouton 1992), soil nutrient content (Wade et al. 1996) and trace metal soil concentrations (Markus and McBratney 1996).

3.1.2.2 Logistic Distribution

The logistic distribution is used as a model for growth and can also be used in a certain type of regression known as logistic regression. This distribution has a similar shape to the normal distribution but has heavier tails.

The PDF of the standard logistic distribution (Fig. 3.2) is given by

$$\frac{\exp\left[\frac{(x-a)}{b}\right]}{b\left(1+\exp\left\{\frac{x-a}{b}\right\}\right)^2}, \quad -\infty < x < \infty$$

Parameters of this distribution can be estimated using the maximum likelihood method.

The logistic distribution can sometimes be used as a substitute for a normal distribution. The logistic distribution has been used to predict the soil water retention based on the particle-size distribution of Swedish soil (Rajkai et al. 1996).

3.1.2.3 Uniform (Rectangular) Continuous Distribution

Although the continuous uniform distribution is one of the simplest of all probability distributions, it is very important because of its role in simulating other probability distributions. This is the distribution taken by uniform random numbers. Every value in the range of the distribution is equally likely to occur (Evans et al. 2000). It is appropriate for representing the distribution of round-off errors in values tabulated to a particular number of decimal places (Siergist 2001).

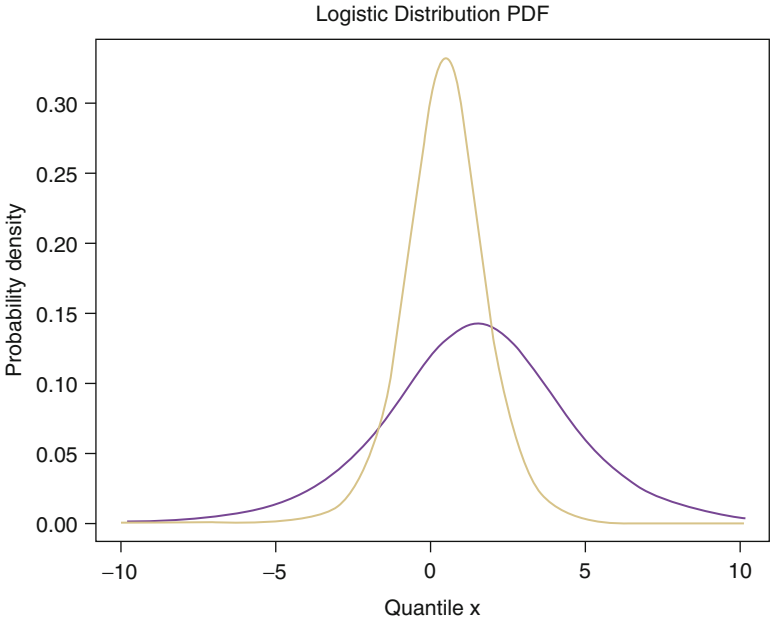


Fig. 3.2 Probability density function of the standard logistic distribution

The PDF for a continuous uniform distribution on the interval [a, b] is

$$\begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases}$$

Since the density function is constant, the mode is not meaningful.

The uniform distribution has been used in soil science as a basic tool for simulating other probability distributions because the uniform distribution corresponds to picking a point at random from an interval, for example, selection of random locations when designing sampling schemes.

3.1.2.4 Lognormal Distribution

The lognormal distribution is applicable to random variables that are constrained by zero but have a few very large values. The resulting distribution is asymmetrical and positively skewed.

Parameters are estimated by transformation to the maximum likelihood estimators of the normal distribution. A random variable x has the lognormal distribution

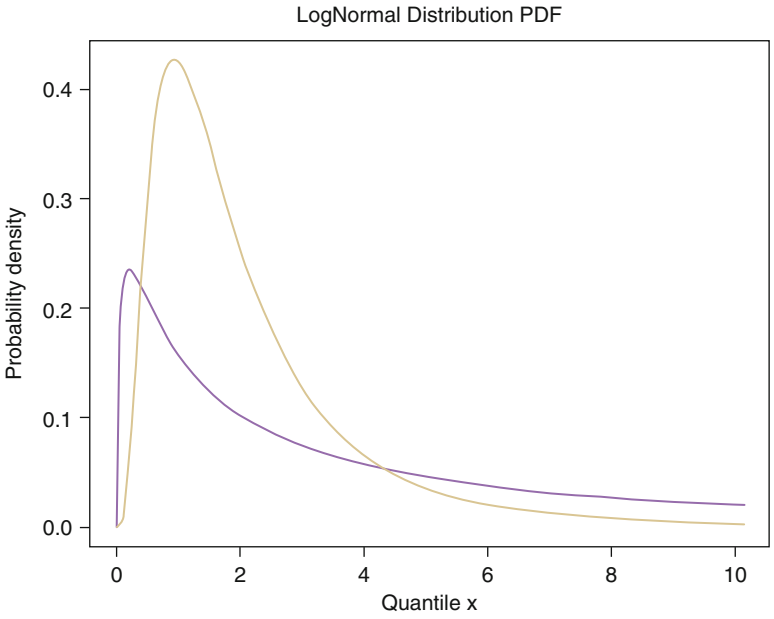


Fig. 3.3 The probability density function of the lognormal distribution

with the location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$. The PDF of the lognormal distribution (Fig. 3.3) is given by

$$\frac{1}{x\sigma(2\pi)^{1/2}} \exp\left(\frac{-(\log(x/\mu))^2}{2\sigma^2}\right)$$

Soils frequently show a lognormal distribution of particle sizes so that a graph of a function versus the logarithm of particle diameters appears to be normally distributed. The lognormal distribution has been used to develop soil water retention models, which assume a lognormal pore-size distribution (Kosugi 1996).

The lognormal distribution has also been used to describe soil organic carbon concentrations in topsoil (McGrath and Zhang 2003; Van Meirvenne et al. 1996).

Mikheeva (2001, 2005b) considered the PDFs of several soil properties of chestnut soils (Russian classification) of Kulunda steppe in Western Siberia. The lognormal distribution was used to model the humus content of nondeflated (unaffected by wind erosion), cohesive sand in the A_p horizon of chestnut soils.

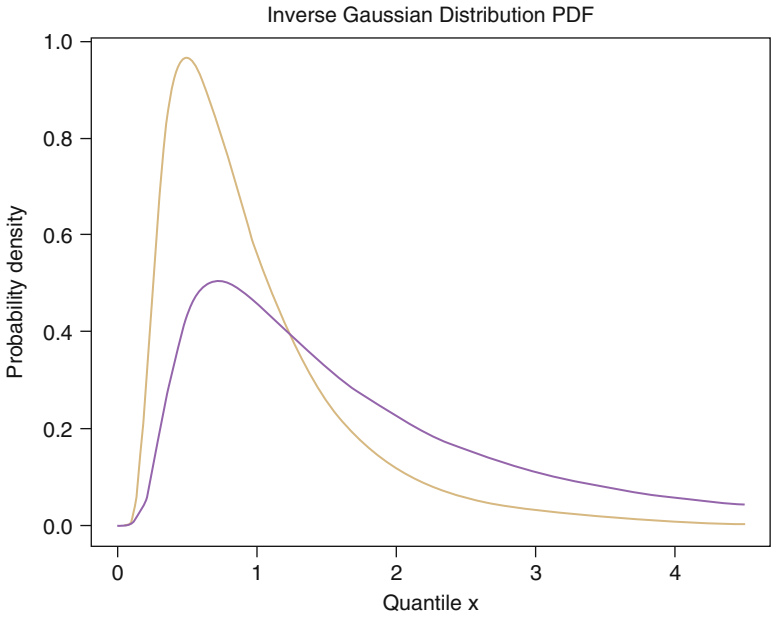


Fig. 3.4 The proability density function of the inverse Gaussian distribution

3.1.2.5 Inverse Gaussian (Wald) Distribution

The inverse Gaussian distribution has applications in the study of diffusion processes and as a lifetime distribution model (Evans et al. 2000). This distribution may be appropriate when one wishes to admit some extreme values to the data.

The inverse Gaussian distribution has a PDF (Fig. 3.4) given by

$$\left(\lambda / \left(2\pi x^3\right)\right)^{1/2} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right)$$

The inverse Gaussian distribution is used in transfer function modelling of solute transport (Jury 1982; Jury and Sposito 1985; Jury et al. 1986). Jury (1982) measured the distribution of solute travel times from the soil surface to a reference depth. Jury demonstrated that this distribution function can be used to simulate the average solute concentrations at any depth and time for arbitrary solute inputs or water application variability and that the model can be used to predict the probability of extremely long or short travel times for a mobile chemical.

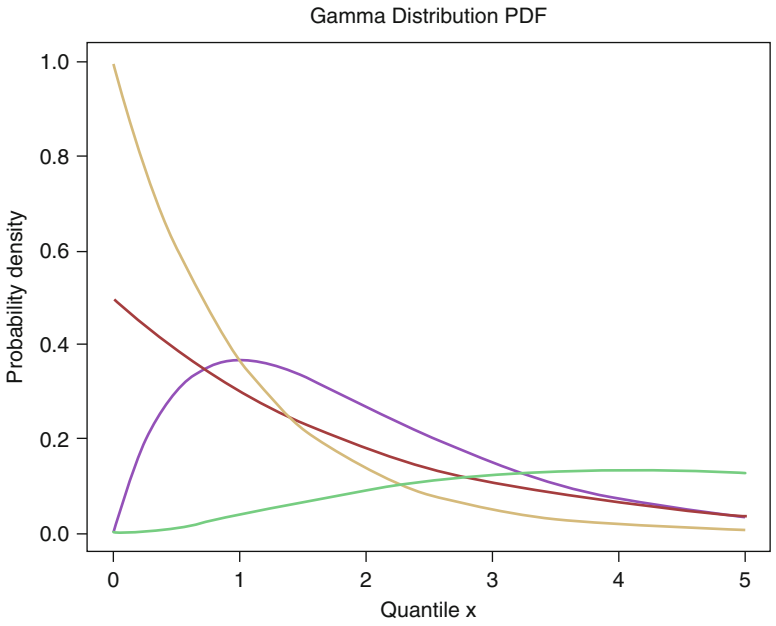


Fig. 3.5 The probability density function of the gamma distribution

3.1.2.6 Gamma Distribution

Before we can study the gamma distribution, we need to introduce the gamma *function*, a special function whose values play the role of the normalising constants (Siergist 2001). The gamma function is defined as follows:

$$\Gamma(c) = \int_0^\infty \exp(-u)\, u^{c-1} du$$

The function is well defined, that is, the integral converges for any $c > 0$. On the other hand, the integral diverges to ∞ for $c \leq 0$.

The PDF (Fig. 3.5) is given by

$$(x/b)^{c-1} \left[\exp(-x/b)\right] / b\Gamma(c)$$

where $\Gamma(c)$ is the gamma function with argument c.

The gamma distribution includes the chi-squared, Erlang and exponential distributions as special cases, but the shape parameter of the gamma is not confined to integer values. The gamma distribution starts at the origin and has a flexible shape. The gamma distribution is commonly used to model continuous variables that are always positive and have skewed distributions. The parameters are easy to estimate by matching moments (Evans et al. 2000).

Russell (1976a, b, c) studied particle-size distribution characterisation and tested the incomplete gamma function as a descriptor for granite disintegration. He found that the function was able to describe the size distribution of the disintegration products from a hot semiarid environment reasonably well. Some consider the gamma distribution to be the most general mathematically convenient model for size distributions of particle types (Petty and Huang 2011).

3.1.2.7 Exponential Distribution

The exponential distribution is a distribution of the time to an event when the probability of the event occurring in the next small time interval does not vary through time. It is also the distribution of the time between events when the number of events in any time interval has a Poisson distribution (Forbes et al. 2010). It is a particular case of the gamma distribution.

The PDF of an exponential distribution (Fig. 3.6) is

$$(1/b)\exp(-x/b) = \lambda \exp(-\lambda x)$$

Mikheeva (2005b) identified the PDFs of several soil properties of chestnut soils (Russian classification) of Kulunda steppe in Western Siberia to be exponential. A

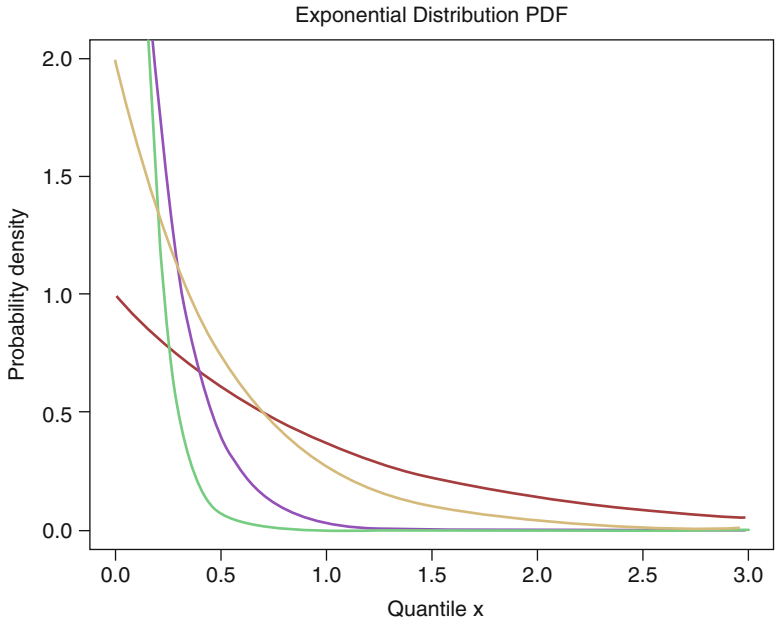


Fig. 3.6 The probability density function of the exponential distribution

variation of the exponential distribution (double exponential distribution) reflected the contents of soil organic matter in fine-loamy sandy chestnut soils.

3.1.2.8 Weibull Distribution

The Weibull distribution is commonly used as a lifetime distribution in reliability applications. It is a versatile distribution and can take on the characteristics of other statistical distributions. The two-parameter Weibull distribution can represent decreasing, constant or increasing failure rates. These correspond to the three sections of the 'bathtub curve' of reliability, referred to also as 'burn-in', 'random' and 'wearout' phases of life (Forbes et al. 2010). The bi-Weibull distribution can represent combinations of two such phases of life (Evans et al. 2000). Parameters are estimated by the method of maximum likelihood.

For strictly positive values of the shape parameter β and the scale parameter η, the PDF (Fig. 3.7) is given by

$$\left(\beta x^{\beta-1}/\eta^{\beta}\right) \exp\left[-(x/\eta)^{\beta}\right]$$

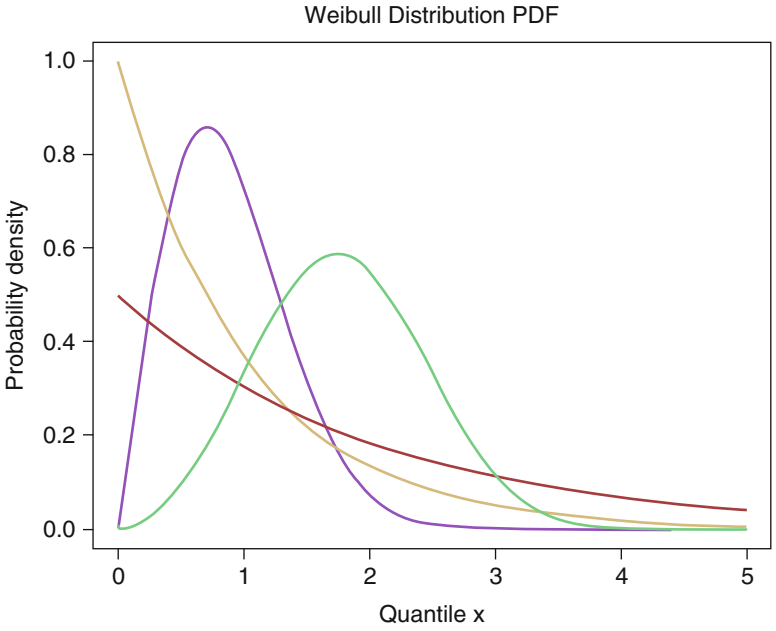


Fig. 3.7 The probability density function of the Weibull distribution

The Weibull distribution can be used to model the distribution of soil moisture deficit and critical soil moisture deficit (Mukherjee and Kottegoda 1992). The Weibull distribution has also been used to study the distribution of radioactive pollutants in soil (Dahm et al. 2002). This is due to its ability to describe a 'snapshot' of the distribution to describe and explain the depth profiles in relatively simple empirical terms. Dahm et al. (2002) used the Weibull distribution to extrapolate the total amount of a radioactive pollutant and then used this data to describe its vertical distribution.

The Weibull cumulative distribution function has also been used to describe soil texture and airborne dust particle-size distributions (Zobeck et al. 1999). The function has a physical basis and is able to describe fragmentation distributions. Weibull distribution is appropriate to describe the airborne soil grains because 'the breakup of a single particle into finer particles', is the physical basis of the Weibull, as described by Brown and Wohletz (1995).

3.1.2.9 Extreme-Value (Gumbel) Distribution

The Gumbel distribution is used to model extreme events. It is also known as the log-Weibull distribution because its variates correspond to the logarithmic transformations of Weibull extreme-value variates.

The Gumbel distribution is appropriate for modelling strength. It may also be appropriate for modelling the life of products that experience very quick wear-out after reaching a certain age. The distribution of logarithms of times can often be modelled with the Gumbel distribution (in addition to the more common lognormal distribution) (Meeker and Escobar 1998).

It has a PDF

$$(1/b)\exp\left[-(x-a)/b\right]x\exp\left\{-\exp\left[-(x-a)/b\right]\right\}$$

The extreme-value distribution can be used when seeking the extreme values of a variable that may occur within a given probability or return period. The extreme-value distribution has been used to model the drainage arising from excess winter precipitation (Rose 1991). In this study, the distribution was used to describe the variability of winter leaching, likely because it can be used successfully as an empirical climatological summary (Rose 1991). It has also been used to estimate rainfall event recurrence, because of its ability to capture the non-normality of recurrence frequencies of hydrologic and associated surface erosion phenomena encountered in semiarid areas (Mannaerts and Gabriels 2000).

In the Mikheeva (2001, 2005a, b) studies, the maximum value distribution was used to model organic matter content in protected coarse-loamy sand and protected sandy loam in the A_p horizon of chestnut soils.

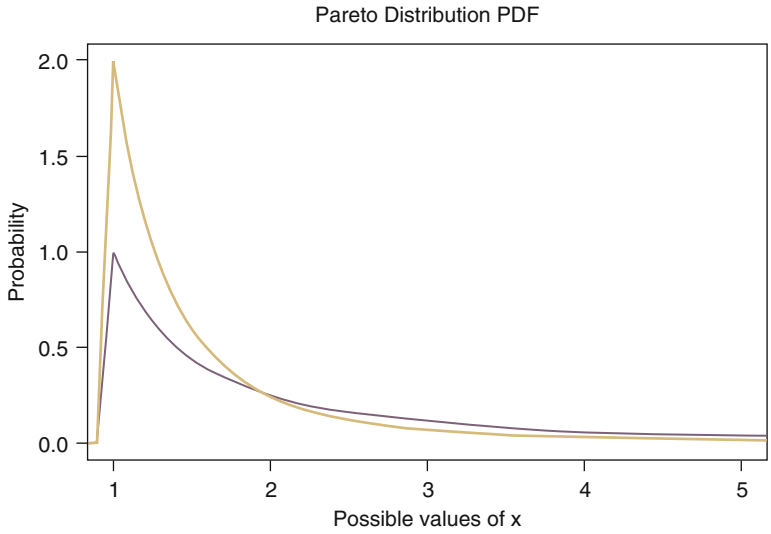


Fig. 3.8 The probability density function of the Pareto distribution

3.1.2.10 Pareto Distribution

The Pareto distribution is often used to model the tails of another distribution because it provides a good fit to extremes of complicated data. It is a power-law probability distribution, meaning that it represents a functional relationship between two quantities where a relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities (Clauset et al. 2009). It can occur in an extraordinarily diverse range of phenomena.

Maximum likelihood estimates of the parameters are

$$\overline{A} = \min x_i$$
$$1/\overline{C} = (1/n) \sum_{i=1}^{n} \log \left(x_i / \overline{A} \right)$$

The PDF (Fig. 3.8) is given by

$$ca^c / x^{c+1}$$

Because it is a power-law relationship, it is related to much of the work on fractal soil geometry (Pachepsky et al. 1999). Perfect et al. (1993) fitted the tensile strength of soil aggregates by a Pareto distribution, the parameters of which determined the probability of failure of the largest aggregate and the rate of change in scale dependency.

The high-value tails of the sizes and size-grade distributions for sets of mineral deposits often can be modelled as Pareto distributions plotting as straight lines on log-log paper suggesting a multifractal model for size-grade distributions (Agterberg 1995). Because of the relationship between soil and underlying rocks, we might speculate that the distribution of areas of polygons of a particular soil class on soil maps might also show Pareto distributions.

3.1.2.11 Su-Johnson Distribution

The Su-Johnson distribution represents one of the four types of Johnson distributions. 'Su' refers to the unbounded Johnson distribution. It is a four-parameter family of probability distributions, proposed as a transformation of the normal distribution.

Overall, the PDF of a Johnson distribution is unimodal with a single 'peak' (i.e. a global maximum), though its overall shape (its height, its spread and its concentration near the y-axis) is completely determined by the values of its arguments. Additionally, the tails of the PDF are 'thin' in the sense that the PDF decreases exponentially rather than decreasing algebraically for large values of x (Wolfram 2010).

The PDF (Fig. 3.9) is given by

$$\frac{\delta}{\lambda\sqrt{2\pi}}\frac{1}{\sqrt{1+\left(\frac{x-\xi}{\lambda}\right)^2}}e^{-\frac{1}{2}\left(\gamma+\delta\sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right)^2}$$

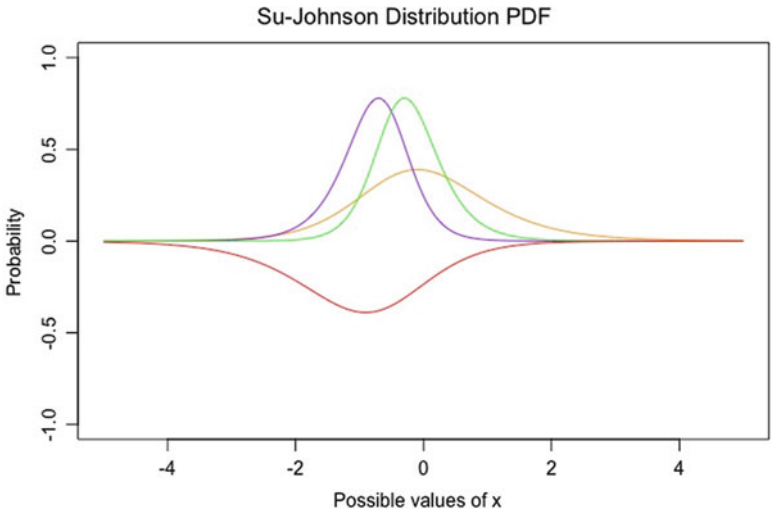


Fig. 3.9 The probability density function of the Su-Johnson distribution

Johnson distributions can be applied to a wide range of possibly non-normal probability distributions via a series of simple transformations that are all easily computable. Because of its flexibility, the family of Johnson distributions has been used to analyse real-world datasets across a number of fields including atmospheric chemistry, biomedical engineering and material science.

In the Mikheeva (2001, 2005a, b) study, the Su-Johnson distribution was used to accurately describe the contents of humus (soil organic matter) in a number of chestnut soil varieties including coarse-loamy sandy, sandy loamy, loamy, coherent sandy and coarse-loamy sandy.

3.1.2.12 Cauchy Distribution

The Cauchy distribution, also called the Lorentzian distribution or Lorentz distribution, is a continuous distribution describing resonance behaviour. It also describes the distribution of horizontal distances at which a line segment tilted at a random angle cuts the x-axis (Weisstein 1995).

The Cauchy distribution is unimodal and symmetric, with much heavier tails than the normal. The PDF is symmetric about a, with upper and lower quartiles $a \pm b$ (Forbes et al. 2010). Additionally, the tails of the PDF are 'fat' in the sense that the PDF decreases algebraically rather than decreasing exponentially for large values of x.

The probability density function (Fig. 3.10) is given by

$$\left\{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1}$$

The Cauchy distribution has a number of applications across a variety of fields of study. It has been used to model a number of phenomena in areas such as risk analysis, mechanical and electrical theory and physical anthropology.

3.1.2.13 Nakagami Distribution

The Nakagami distribution represents a continuous statistical distribution supported on the interval (0, ∞) and parameterised by positive real numbers, a 'shape parameter' and a 'spread parameter', which together determine the overall behavior of its PDF. Depending on the values of its two parameters, the PDF of a Nakagami distribution may have any number of shapes, including unimodal with a single 'peak' (i.e. a global maximum) or monotone decreasing with potential singularities approaching the lower boundary of its domain. In addition, the tails of the PDF

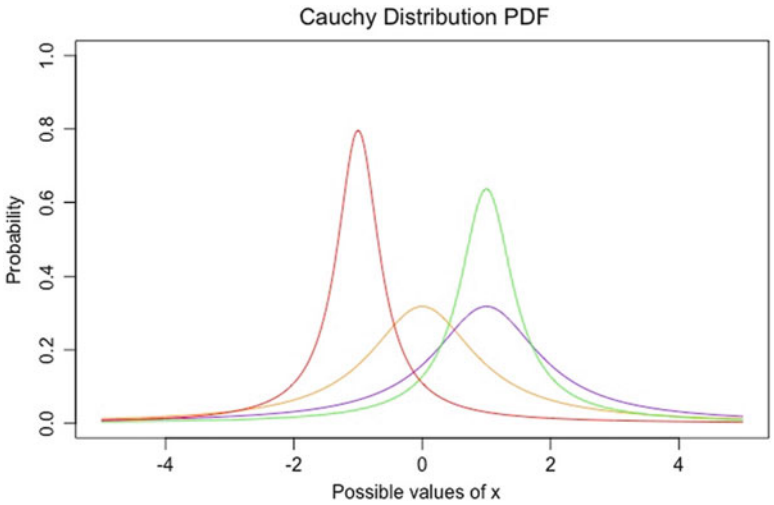


Fig. 3.10 The probability density function of the Cauchy distribution

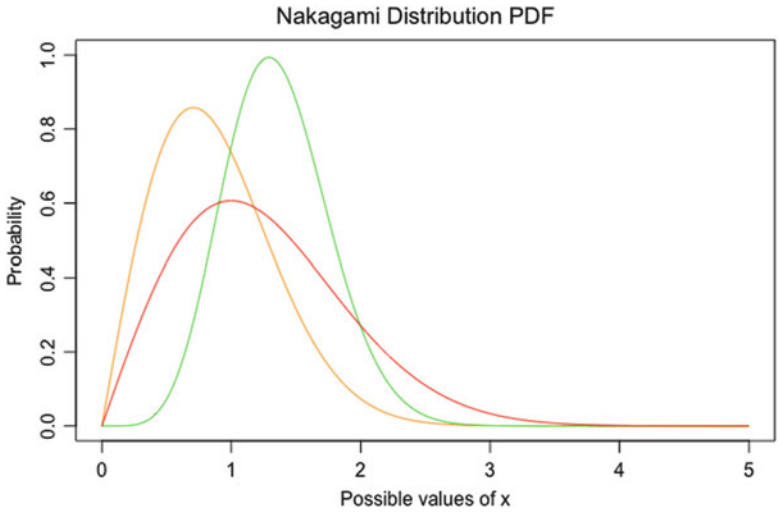


Fig. 3.11 The probability density function of the Nakagami distribution

are 'thin' in the sense that the PDF decreases exponentially rather than decreasing algebraically for large values of x (Wolfram 2010).

The PDF (Fig. 3.11) is given by

$$\frac{2\mu^{\mu}}{\Gamma(\mu)\,\omega^{\mu}} x^{2\mu-1} e^{\frac{-\mu}{\omega}x^2}$$

The Nakagami distribution was first proposed as a mathematical model for small-scale fading in long-distance high-frequency radio wave propagation. Many applications of the distribution have been wave related, and it has been used to model phenomena related to meteorology, hydrology and seismology (Wolfram 2010).

In the Mikheeva (2001, 2005a, b) studies, the Nakagami distribution was used to model organic matter content in protected fine-loamy sand in the A_p horizon of chestnut soils.

3.1.3 Directional Variables

When considering data distributions, one may encounter observations based on angular displacements and orientations. Their analyses, however, require special attention since the methods advocated for usual linear data are not only often misleading but also not applicable to such directional data. Data of this kind are called *directional data* (Mardia 1972).

The directions are regarded as points on the circumference of a circle in two dimensions or on the surface of a sphere in three dimensions. In general, directions may be visualised as points on the surface of a hypersphere, but observed directions are obviously angular measurements (Mardia 1972).

Describing and analysing such data statistically present interesting problems. The problem is that there is no natural 'starting point' for measuring directions; any direction can be chosen, arbitrarily, as the 'baseline'. It is also completely arbitrary as to whether to work clockwise or anticlockwise, up or down. However, this can be overcome by representing the directions as vectors (Fisher 1993).

3.1.3.1 Circular Distributions

The von Mises distribution is a 'natural' distribution for circular attributes such as angles, time of day, day of year, phase of the moon, etc. It has certain characterisations analogous to those of the linear normal distribution (Evans et al. 2000). The von Mises distribution M(a, b) has a mean direction a and concentration parameter b. For small b it tends to a uniform distribution, and for large b it tends to a normal distribution with variance 1/b.

A continuous distribution defined on the range $x \in [0, 2\pi)$ with PDF

$$P(x) = \frac{e^{b\ \cos(x-a)}}{2\pi I_0(b)}, \tag{3.1}$$

where $I_0(x)$ is a modified Bessel function of the first kind of order 0 and distribution function

$$D(x) = \frac{1}{2\pi I_0(b)} \left\{ xI_0(b) + 2\sum_{j=1}^{\infty} \frac{I_j(b)\sin\left[j\left(x-a\right)\right]}{j} \right\}, \tag{3.2}$$

which cannot be done in closed form. Here, $a \in [0, 2\pi)$ is the mean direction and $b > 0$ is a concentration parameter. The von Mises distribution is the circular analogue of the normal distribution on a line.

The mean is

$$\mu = a \tag{3.3}$$

and the circular variance is

$$\sigma^2 = 1 - \frac{I_a(b)}{I_0(b)}. \tag{3.4}$$

The von Mises distribution (Fig. 3.12) has been described as a natural choice of model in theory and, at the same time, a difficult model to handle in practice.

The von Mises distribution is most commonly seen in soil science to describe soil fabric patterns. When analysing any complex, multicomponent fabric pattern, one must be able to recognise the multicomponent individuals and their geometrical relationships towards adjacent individuals. Those relationships can then be considered as directional data. Murphy et al. (1977) studied voids in soil thin sections and modelled the orientation of components and their discrimination according to shape. Crack segments, analysed by their length and direction, are an example of a naturally directional distribution. The von Mises distribution can then be used to test for isotoropy or uniformity of direction.

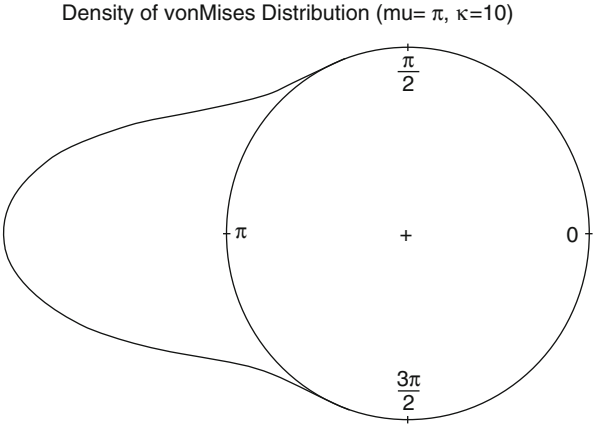


Fig. 3.12 Density of the von Mises distribution

Directional data which follows a circular distribution can be presented in a rose diagram, which can be thought of as a 'circular histogram'. This technique was used by FitzPatrick (1975) to describe stone orientation patterns in soils of north-east Scotland. It was concluded that the pattern of the orientation suggested frost heaving. In this particular study, FitzPatrick considered the data in two dimensions – the orientation of the stones in relation to slope. However, if the author had wanted to incorporate the dipping angle, this data could also be considered in three dimensions. To incorporate the additional dimension, one would have to consider a spherical distribution.

3.1.3.2 Spherical Distributions

The two most commonly described vector distributions are exact analogues of the scalar uniform and normal distributions. On the unit sphere, a uniform distribution can be viewed as a completely random shotgun pattern over the entire sphere surface. Any single vector direction has an equal likelihood of occurrence. The spherical analogue of the normal distribution is called Fisher's distribution (Bingham 1974).

Any vector can be represented by a point P on the surface of a unit sphere of centre O, $\overrightarrow{OP}$ being the directed line; any axis can be represented by a pair of points, Q, P at opposite ends of a diameter QOP of the sphere, or equivalently by a point P on a unit hemisphere. Two quantities are required to define a point P on the sphere or hemisphere and hence a spherical measurement, whether vectorial or axial.

If we assume that the distribution is symmetrically distributed about its mean direction, the direction of the resultant of the sample unit vectors is a natural estimate of the mean direction, and, provided that the sample size is not too small, we can get an adequate estimate of the error of the sample mean direction by a simple procedure (Fisher et al. 1987). On the other hand, if a probability distribution such as the Fisher distribution can be fitted satisfactorily to the data, we are in effect assuming additional information and so can get an improved estimate of the error in the sample mean direction. The principal exploratory tool is to look at the eigenvalues of the orientation matrix.

Three-dimensional directional data relate particularly to soil mechanical processes. For example, Lafeber (1965) described the spatial orientation of planar pore patterns in soils, explaining that these patterns were an expression of the depositional and/or stress-strain history of the particular soil. Another example related to soil mechanical processes was described by Knight (1979) who studied gilgai mircrorelief in the Australian state of Victoria. Knight studied both two-dimensional crack patterns on vertical sections and three-dimensional measurements on monoliths. The author used these measurements to propose that differences between lateral and vertical stresses due to swelling pressures and overburden loads are sufficient to cause small, inclined shear displacements, which cause the gilgai microrelief.

There is not extensive formal analysis of such data, but recent work by van Dusschoten et al. (2016) has analysed the three-dimensional directional data of plant root growth in soil using magnetic resonance imaging. In order to do this, the authors developed a data analysis tool that computed root traits from MRI-acquired images. The study found that the angular projection maps of root mass densities, root length densities and root tip densities showed similar distribution patterns for those traits and that the depth distribution of root mass, root length and number of root tips also matched rather well.

3.2 Displaying Distributions

It needs to be recognised that pedometrics studies two kinds of distribution:

- Distribution of a soil property over an area, e.g. distribution of organic matter content of a farm
- Distribution of a property within an individual soil sample, e.g. the particle-size distribution of a soil sample taken from a field

There are various ways of displaying distributions; the most basic is a histogram. A kernel density plot is a non-parametric estimate of the PDF of a variable. Figure 3.13 shows the soil production rate (Chap. 19) from a global dataset as measured by two different methods.

Figure 3.14 shows some alternative of displaying the distribution, including the box plot and violin plot. A violin plot is a box plot with a rotated kernel density plot on each side.

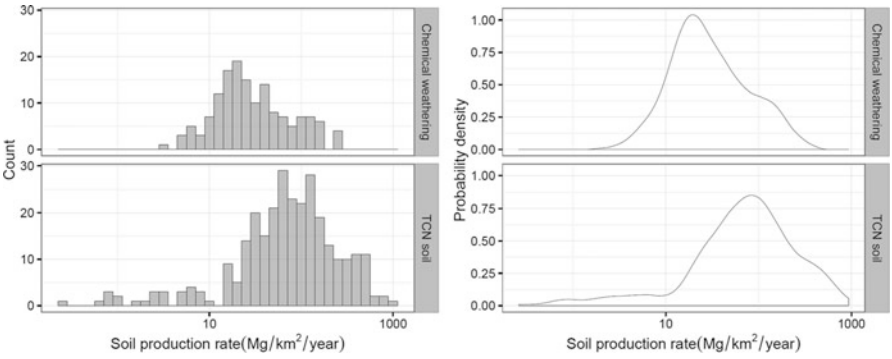


Fig. 3.13 Global soil production rates determined from chemical weathering studies and from terrestrial cosmogenic nuclides (TCN), displayed using a histogram and probability density plot. Note that the values are in logarithmic scale

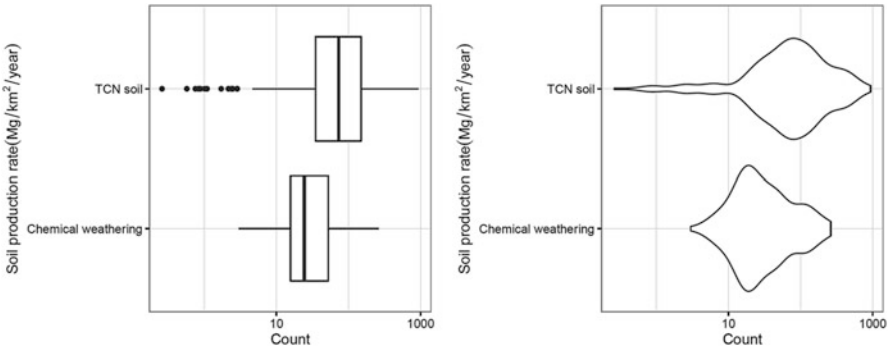


Fig. 3.14 Global soil production rates determined from chemical weathering studies and from terrestrial cosmogenic nuclides (TCN), displayed using the box plot and violin plot. Note that the values are in logarithmic scale

3.3 Statistical Distributions of Continuous Soil Properties for Soil Genesis

In Russian studies of soil geography, statistical distributions have been used as a tool for modelling temporal change and soil evolution. In this section, we illustrate the use of statistical distribution as a tool to characterise evolutionary change in soil properties.

3.3.1 Changes in Spatial Variability of Soil Properties Under Anthropogenic Impact

Statistical distributions of soil properties under anthropogenic and natural processes have long been discussed in Russian literature. Vauclin reasoned that the normal distribution is inherent to static soil properties, while the lognormal distribution is typical for dynamic properties. Kozlovskii noted that the spatial differentiation of the soil evolution processes gives rise to the asymmetry of empirical statistical distributions for dynamic parameters of soil, in particular, causing the normal distribution to be lognormal. The authors note that the change in the dispersion parameter with time can serve as a reliable indicator of soil change.

The statistical analysis of a great body of soil data allowed the development of a model for the variations of the shape of statistical distribution (Fig. 3.15). Consider soil process Y which is caused by environmental conditions which are either natural or anthropogenic and results in the quantitative increase of soil property X, whose initial distribution had a small dispersion (Fig. 3.15a). Over time, the notable increase in dispersion and right asymmetry for variable $X(X_1,X_2,...,X_n)$ in space (Fig. 3.15b) indicate that several values which substantially exceed the mode correspond to the areas that are susceptible to the given environmental change. At the same time, most of variable X retain moderate values; hence, the

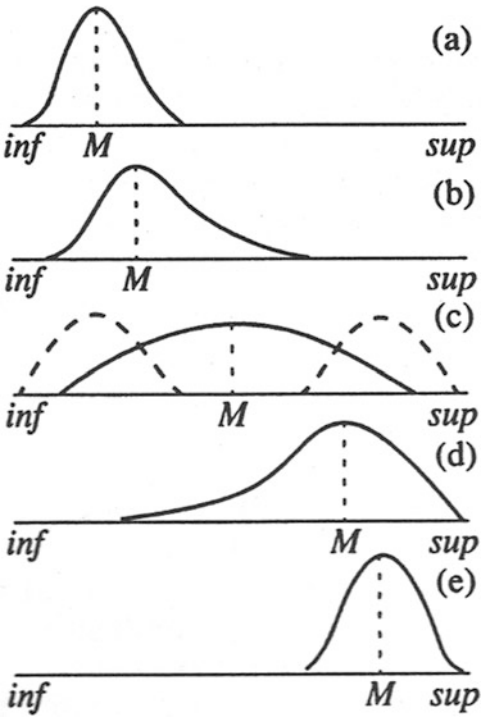


Fig. 3.15 Stages of changes in the statistical distribution of soil properties; *(inf)* is the least possible limit of the variability of a property in a specific soil; *(sup)* is the highest possible limit of the variability of a property in a specific soil; *(M)* is the mean value of a property; for (*a–e*) explanations in text

corresponding areas are tolerant to the environmental influence. Thus, the right asymmetry of statistical distribution suggests a weak dispersion or the beginning stages of process *Y* on the area. Under constant environmental conditions responsible for the development of the process *Y*, the asymmetry of statistical distribution of *X* decreases, and the distribution becomes more symmetrical with higher dispersion (Fig. 3.15c).

A large decrease in the distribution tail can result in the split of the distribution curve (dotted line, Fig. 3.15c), which indicates heterogenisation. In other cases, further development of the process *Y* can result in the left asymmetry of statistical distribution of parameter *X* (Fig. 3.15d), which suggests the approaching of the limiting spatial distribution of the process. Here, the upper boundary of variability reaches its limit, which is determined by the nature of soil properties, and most of values are grouped around the mode in the upper part of the variation interval. Further development of the process results in decreasing scattering of values, and the shape of distribution becomes more symmetrical (Fig. 3.15e), which suggests the completion of the process on the area and its homogenisation in relation to a given property. Considering the process resulting in the quantitative decrease of a property, one can observe the reverse sequence of changes in statistical distribution (Fig. 3.15e–3.15a). Certain ratios of conditions, factors and internal properties of soil systems, the existence of straight lines and feedback create quasi-equilibrium

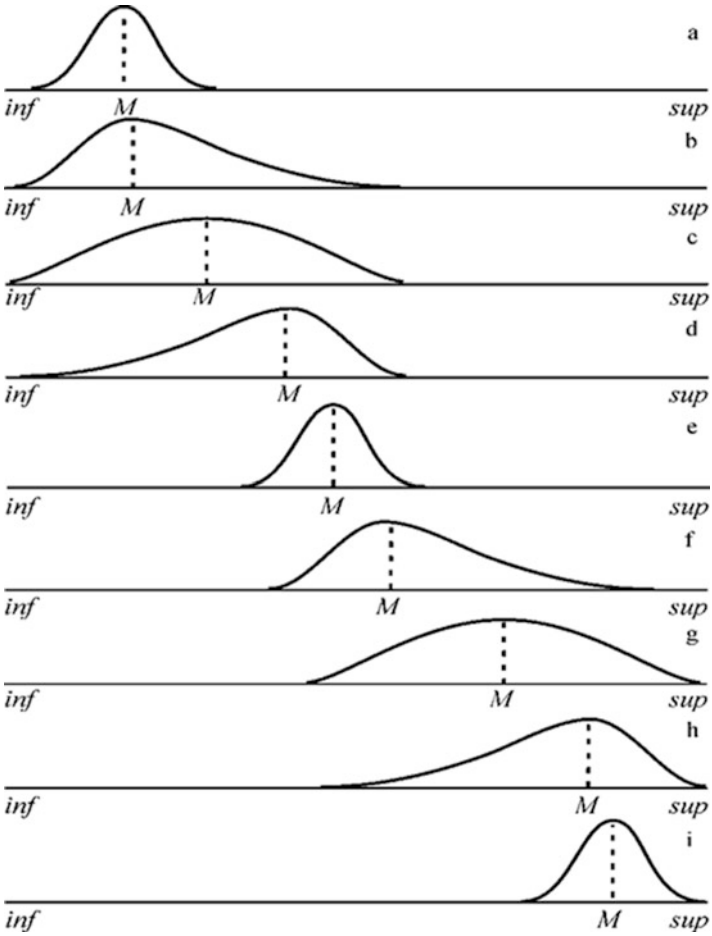


Fig. 3.16 Theoretical stages of changes in the statistical distribution of soil properties; *inf* is the least possible limit of the variability of a property in a specific soil; *sup* is the highest possible limit of the variability of a property in a specific soil; *M* is the mean value of the property

states which define the real limits of variation of soil properties (min – max) which do not coincide with their potential limits (inf – sup).

Stages of transformations of soil probability density functions during soil evolution are shown in Fig. 3.15 in the case where the potential limits of variation of a soil property are close to their real limits. In the case where the potential limits of variation are far from the real limits, the stages of change of the shape of statistical distribution can go on several paths (Fig. 3.16a–i) according to the recurrence of environmental change.

The actual statistical distributions of soil properties can correspond to any of the patterns presented in Fig. 3.15 or an intermediate one, depending on the devel-

opment of the processes responsible for these properties and on the soil-forming factors governing these processes. The multiplicity of statistical distributions of soil properties possesses some regular trends, which can be used for the diagnostics of the development of natural and anthropogenic processes.

The analysis of the shape of statistical distribution of properties is promising for characterisation and evaluation of soil transformation, because the shape of distribution reflects the degree of manifestation of processes responsible for these properties and can serve as an indicator of the stages of their development and spatial distribution of soil properties in an area.

3.4 Conclusions

Probability distributions should be studied carefully to gain insight:

1. To the correct stastical treatment
2. To sampling and measurement methods
3. To suggest or explain processes
4. To environmental and evolutionary sequences

A suite of possible distributions should be considered. It is an area of pedometrics that requires more detailed attention.

References

Agterberg FP (1995) Multifractal modeling of the sizes and grades of giant and supergiant deposits. Int Geol Rev 37:1–8

All-Union Guidelines for Soil Survey and Compilation of Large-Scale Maps of Land Tenure (1975) Kolos, Moscow, 95 [in Russian]

Bingham C (1974) An antipodally symmetric distribution on the sphere. Ann Stat 2:1201–1225

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc Ser B Methodol 26(2): 211–252

Brown WK, Wohletz KH (1995) Derivation of the Weibull distribution based on physical principles and its connection to the Rosin-Rammler and lognormal distributions. J Appl Phys 78: 2758–2763

Carsel RF, Parrish RS (1988) Developing joint probability distributions of soil water retention characteristics. Water Resour Res 24(5):755–769

Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. Soc Ind Appl Math Rev 51(4):661–703

Dahm H, Niemeyer J, Schroder D (2002) Application of the Weibull distribution to describe the vertical distribution of cesium-137 on a slope under permanent pasture in Luxembourg. J Environ Radioact 63(3):207–219

Evans M, Hastings N, Peacock B (2000) Statistical distributions, 3rd edn. Wiley, New York

Fisher NI (1993) Statistical analysis of circular data. Cambridge University Press, Cambridge

Fisher NI, Lewis T, Embleton B (1987) Statistical analysis of spherical data. Cambridge University Press, Cambridge

FitzPatrick EA (1975) Particle size distribution and stone orientation patterns in some soils of north-east Scotland. In: Gemmell AMD (ed) Quaternary studies in North-East Scotland. Department of Geography, University of Aberdeen, Aberdeen, pp 49–60

Forbes C, Evans M, Hastings N, Peacock B (2010) Statistical distributions, 4th edn. Wiley, New York

Jury WA (1982) Simulation of solute transport using a transfer function model. Water Resour Res 18(2):363–368

Jury WA, Sposito G (1985) Field calibration and validation of solute transport models for the unsaturated zone. Soil Sci Soc Am J 49:1331–1341

Jury WA, Sposito G, White RE (1986) A transfer function model of solute transport through soil. 1. Fundamental concepts. Water Resour Res 22:243–247

Kirk GJD, Nye PH (1986) A simple model for predicting the rates of dissolution of sparingly soluble calcium phosphates in soil. II Applications of the model. J Soil Sci 37:541–554

Knight MJ (1979) Structural analysis and mechanical origins of gilgai at Boorook, Victoria, Australia. Geoderma 23:245–283

Kosugi K (1996) Lognormal distribution model for unsaturated soil hydraulic properties. Water Resour Res 32(9):2697–2703

Lafeber D (1965) The graphical representation of planar pore patterns in soils. Aus J Soil Res 3:143–164

Mannaerts CM, Gabriels D (2000) A probabilistic approach for predicting rainfall soil erosion losses in semiarid areas. Can Underwrit 48(4):403–420

Mardia KV (1972) Statistics of directional data. Academic, London

Markus JA, McBratney AB (1996) An urban soil study: heavy metals in Glebe, Australia. Aust J Soil Res 34:453–465

McGrath D, Zhang C (2003) Spatial distribution of soil organic carbon concentrations in grassland of Ireland. Appl Geochem 18(10):1629–1639

Meeker WQ, Escobar LA (1998) Statistical methods for reliability data. Wiley, New York

Mikheeva IV (2001) Statistical probability models of soil properties (on example of Kashtanozems in Kulunda steppe). Publishing of Siberian Branch of Russian Academy of Sciences, Novosibirsk. 103. (in Russian)

Mikheeva IV (2005a) Monitoring and probabilistic-statistical assessment of stability and changeability of natural objects under contemporary processes (on example of Kashtanozems in Kulunda steppe). Publishing of Siberian Branch of Russian Academy of Sciences, Novosibirsk. 103. (in Russian)

Mikheeva IV (2005b) Spatial fluctuations and statistical probability distributions of chestnut soil properties in the Kulunda steppe. Eurasian Soil Sci 38(3):278–288

Mukherjee D, Kottegoda NT (1992) Stochastic model for soil moisture deficit in irrigated lands. J Irrig Drain Eng 118(4):527–542

Mulugeta D, Stoltenberg DE (1997) Seed bank characterization and emergence of a weed community in a moldboard plow system. Weed Sci 45:54–60

Murphy CP, Bullock P, Turner RH (1977) Measurement and characterization of voids in soil thin-sections by image analysis. 1. Principles and techniques. J Soil Sci 28(3):498–508

Pachepsky Y, Crawford JW, Rawls WJ (eds) (1999) Fractals in soil science (Special Issue). Geoderma 88(3-4):103–364

Perfect E, Kay BD, Rasiah V (1993) Multifractal model for soil aggregate fragmentation. Soil Sci Soc Am J 57:896–900

Petty GW, Huang W (2011) The modified gamma size distribution applied to inhomogeneous and nonspherical particles: key relationships and conversions. J Atmos Sci 68(7):1460–1473

Rajkai K, Kabos S, Van Genuchten MT, Jansson PE (1996) Estimation of water-retention characteristics from the bulk density and particle size distribution of Swedish soils. Soil Sci 161(12):832–845

Rose DA (1991) The variability of winter drainage in England and Wales. Soil Use Manag 7: 115–121

Russell DA (1976a) Particle-size distribution characterization for the coarse fraction of a granite soil: I. A model. Soil Sci Soc Am J 40:409–413
Russell DA (1976b) Particle-size distribution characterization for the coarse fraction of a granite soil: II. Disintegrating granite. Soil Sci Soc Am J 40:414–418
Russell DA (1976c) Particle-size distribution characterization for the coarse fraction of a granite soil: III. The coarse fraction of a granite sand. Soil Sci Soc Am J 40:418–422
Russo D, Bouton M (1992) Statistical analysis of spatial variability in unsaturated flow parameters. Water Resour Res 28(7):1911–1925
Siergist K (2001) Virtual laboratories in probability and statistics. Department of Mathematical Statistics. University of Alabama in Huntsville. http://www.math.uah.edu/stat/index.html
Utset A, Cid G (2001) Soil penetrometer resistance spatial variability in a Ferralsol at several soil moisture conditions. Soil Tillage Res 61(3–4):193–202
van Dusschoten D, Metzner R, Kochs J, Postma JA, Pflugfelder D, Bühler J, Schurr U, Jahnke S (2016) Quantitative 3D analysis of plant roots growing in soil using magnetic resonance imaging. Plant Physiol 170(3):1176–1188
Van Meirvenne M, Pannier J, Hofman G, Louwagie G (1996) Regional characterization of the long-term change in soil organic carbon under intensive agriculture. Soil Use Manag 12:86–94
Wade SD, Foster IDL, Baban SMJ (1996) The spatial variability of soil nitrates in arable and pasture landscapes: implications for the development of geographical information system models of nitrate leaching. Soil Use Manag 12(1):95–101
Weisstein EW (1995) “Cauchy Distribution” from MathWorld web resource. Wolfram Research, Inc
Wolfram Research, Inc. (2010) Mathematica, Version 8.0, Champaign, IL
Wulfsohn D, Gu Y, Wulfsohn A, Mojlaj EG (1996) Statistical analysis of wheat root growth patterns under conventional and no-till systems. Soil Tillage Res 38:1–16
Zobeck TM, Gill TE, Popham TW (1999) A two-parameter Weibull function to describe airborne dust particle size distributions. Earth Surf Process Landf 24:943–955

Chapter 4
Effective Multivariate Description of Soil and Its Environment

Alex. B. McBratney, Mario Fajardo, Brendan P. Malone, Thomas F. A. Bishop, Uta Stockmann, and Inakwu O. A. Odeh

> *" . . . nature's laws are causal; they reveal themselves by comparison and difference, and they operate at every multivariate space/time point".*
>
> Edward Tufte

4.1 Introduction

When we wish to characterize soil, it soon becomes very clear that one or two properties of soil materials, horizons, profiles or pedons will not suffice to give an adequate description. Soil classification, land capability, soil quality, condition and health assessments often involve the observation of tens or scores of soil properties on a single soil entity; e.g. the new soil microbial DNA descriptions involve hundreds or thousands of attributes. For analysis of such high-dimensional data, multivariate statistical techniques are most appropriate, particularly ordination techniques which help to reduce the dimensionality down to a few (typically 2 or 3) which can be graphed simply and the relationships between soil entities, and between observed soil attributes on those entities, displayed.

Additionally, we are interested in understanding the relationship between the attributes of soil entities and the environment in which they exist. Univariate statistical techniques, such as ANOVA, have commonly been used to study the impact of environmental effects on soil properties (Islam and Weil 2000; Braimoh 2004; Crave and Gascuel-Odoux 1997; McGrath et al. 2001; Latty et al. 2004); they may not be suitable when a large number of variables is involved. This is because a separate analysis will be required for different factors. It may therefore be more

A.B. McBratney (✉) • M. Fajardo • B.P. Malone • T.F.A. Bishop • U. Stockmann • I.O.A. Odeh
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: alex.mcbratney@sydney.edu.au; mario.fajardo@sydney.edu.au;
brendan.malone@sydney.edu.au; thomas.bishop@sydney.edu.au; uta.stockmann@sydney.edu.au;
inakwu.odeh@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_4

informative and economical to handle the different sets of data all simultaneously using multivariate ordination techniques (Shaw 2003) to explore the complex interrelationships among different sets of variables. These multivariate techniques are powerful tools for investigating and summarizing the underlying trends in complex data structures (Kenkel et al. 2002) and have been widely, and successfully, applied in species-environment relationships (Ter Braak 1986, 1987, 1995; Palmer 1993; Graffelman and Tuft 2004; Zhang and Oxley 1994; Dray et al. 2008). Moreover, these techniques in many cases do not require Gaussian distribution of data.

There are different ordination techniques, among them are principal component analysis (PCA), correspondence analysis (CA), canonical correlation analysis (COR), canonical correspondence analysis (CCA) and redundancy analysis (RDA). These techniques are widely used in environmental sciences. While PCA is used for continuous data, CA is preferred for categorical data. The canonical counterparts of them (from Greek κανων pronounced 'kanon': rule) are COR and CCA commonly used when the *direct interrelations* (the ordination of one set of variables intervenes in the calculation of the final ordination of the second set) between two sets of variables are meant to be explored, e.g. environment variables vs soil variables. RDA can be compared with COR as it differs in the way of relating both sets; RDA uses linear regressions and COR uses linear correlations (Legendre and Legendre 2012).

In this chapter we will briefly describe each of these methods in terms of the data they can be used on (e.g. categorical versus continuous) and the information they can provide. We will illustrate the methods with a soil and environmental dataset from Australia. Finally, we discuss approaches to analyse compositional data, in particular particle-size fractions of soil, as when analysing multivariate soil data, we need to account for its compositional nature.

4.2 Transformations into Multivariate Spaces

4.2.1 Principal Component Analysis

Principal component analysis (PCA) is a procedure for identifying a smaller number of uncorrelated variables, called 'principal components', from a large set of data. The goal of PCA is to explain the maximum amount of variance with the fewest number of principal components. Thus, PCA is commonly used in the earth, ecological and environmental sciences that use large datasets, including soil science.

PCA has been used as a step to a series of complex analysis that involves continuous data only. PCA is used to reduce the number of continuous variables and avoid multicollinearity or when there are too many predictors relative to the number of observations. Given the wealth of easily accessible data, it is becoming

more commonly used as preprocessing step for reducing the number of predictors before performing variable selection in statistical modeling of soil. For example, Karunaratne et al. (2014) used PCA to reduce the number of potential predictor variables from 31 to the first 6 principal components in a soil carbon mapping study. In the case of more complex datasets that could involve a mixture of continuous, categorical and nominal data, principal coordinate analysis could be used (Rayner 1966). A more detailed description of this type of analysis and similar other ordination techniques can be found in Sect. 9.3.1.1.

PCA involves some form of linear combinations such that the first component accounts for as much of the variance as possible and the next component accounts for as much as of the variance as possible, etc., while the components are uncorrelated with each other. PCA is a linear method that is appropriate to use when the data is continuous (Ter Braak and Prentice 1988). As an ordination technique, it constructs the theoretical variable that minimizes the total residual sum of squares after fitting straight lines to the sample data (Ter Braak 1995).

PCA relates site data to environmental variables in two steps: (i) a few ordination axes that summarize the overall variation are extracted, and then (ii) the weighted sums of the environmental variables that most closely fit each of these ordination axes are calculated (Ter Braak and Prentice 1988). This two-step approach (ordination followed by environmental gradient identification) is termed as an indirect gradient analysis (Ter Braak 1986; Ter Braak and Prentice 1988). A joint analysis of all environmental variables can be carried out by multiple regression of each ordination axis on the environmental variables (Eq. 4.1):

$$x_i = c_0 + \sum_{j=1}^{q} c_j z_{\mathrm{ij}} \tag{4.1}$$

in which x_i is the score of site i on that one ordination axis, z_{ij} denotes the value at site i of the jth out of q actual environmental variables and c_j is the corresponding regression coefficient.

4.2.2 Correspondence Analysis

Correspondence analysis (CA) is a non-linear counterpart to PCA. As unimodal response models are more general than monotonic, it is usually appropriate to start analysing the data using CA and its canonical version (CCA) (Ter Braak and Prentice 1988; Odeh et al. 1991). As PCA, CA is also an indirect gradient analysis. CA extracts continuous axes of variation from contingency tables (categorical data) or continuous data (using discretization), typically interpreted with the help of external knowledge and data on environmental variables (Ter Braak 1986; Ter Braak and Prentice 1988). CA is a popular ordination technique in environmental sciences and especially in ecology (Ter Braak 1986; Kenkel 2006); it uses an efficient reciprocal weighted averaging algorithm as shown in Eq. 4.2:

$$u_k = \sum_{i=1}^{n} y_{\mathrm{ki}} x_i / \sum_{i=1}^{n} y_{\mathrm{ki}} \tag{4.2}$$

where y_{ki} is abundance of category k (e.g. species in ecology terms, soil or soil type in pedology) at site i, x_i is the score of site i and μ_k is the score of category k.

If the gradient length is reduced to less than about 3 SD (standard deviation), the approximations involved in weighted averaging become worse, and ultimately the methods yield poor results because most species or soil types are behaving monotonically over the observed range. In these situations (variation within a narrow range), the linear ordination methods – PCA or RDA – may be more appropriate. However, if the community variation is over a wider range, the non-linear methods such as CA and CCA are better (Ter Braak and Prentice 1988).

4.2.3 Canonical Correlation Analysis

COR is probably the first statistical methodology (Hotelling 1936) that offered a way to measure the relations between two sets of multiple variables. In pedology, Webster (1977) used COR to illustrate the well-known interactions between soil and environment. Intuitively, the method relies on two steps: (1) a principal component analysis performed to each of the sets of variables (e.g. soil and environment variables) and (2) a rotation of the principal axes making sure that their correlation is maximized. Briefly, let p and q be variables that have been measured on every observation site; thus X are the $n \times p$ observed variables (e.g. soil data) and Y are the $n \times q$ observed co-variables (e.g. environmental data), with both being centred so that their values are deviations from their means. Then S is the variance covariance matrix partitioned into S_{11} ($p \times p$), S_{22} ($q \times q$), S_{12} ($p \times q$) and S_{21} ($q \times p$) submatrices (with $S_{11} = \text{var-cov}(X)$ and $S_{22} = \text{var-cov}(Y)$ and so on) as follows:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{4.3}$$

As explained in Legendre and Legendre (2012), canonical correlations are then obtained by solving the characteristic equation:

$$\left| S_{12} S_{22}^{-1} S_{12} - \lambda_k S_{11} \right| = 0 \tag{4.4}$$

and

$$\left| S_{21} S_{11}^{-1} S_{12} - \lambda_k S_{22} \right| = 0 \tag{4.5}$$

Then, canonical correlations r_k are the square roots of the eigenvalues ($\lambda_K = r^2$). Once the eigenvalues are obtained, the eigenvectors can be calculated as follows:

$$\left(S_{12}S_{22}^{-1}S_{12} - \lambda_k S_{11}\right) u_k = 0 \tag{4.6}$$

and

$$\left(S_{21}S_{11}^{-1}S_{12} - \lambda_k S_{22}\right) v_k = 0 \tag{4.7}$$

Canonical variates (or the new coordinates) are then obtained by multiplying the respective eigenvectors with the original coordinates:

$$c_i = \mathrm{X}\mathrm{u}_k \tag{4.8}$$

and

$$d_i = \mathrm{Y}\mathrm{v}_k \tag{4.9}$$

Finally, we obtain two matrices ($n \times p$ *for soil variables* and $n \times q$ *for environmental variables*) of canonical ordination scores. A more detailed description of this method can be found in Legendre and Legendre (2012), Webster (1977) and Kendall and Stuart (1977).

4.2.4 Canonical Correspondence Analysis

Canonical correspondence analysis (CCA) is designed to detect the patterns of variation in the categorical (or continuous) data that can be explained 'best' by the observed environmental variables. The resulting ordination diagram expresses not only a pattern of variation in categories composition (species in ecology or soil types in soil science) but also the main relations between the species and each of the environmental variables. CCA thus combines aspects of regular ordination with aspect of regression (Ter Braak 1986, 1995). The difference between CCA and CA is analogous to the difference between RDA and PCA. CCA is therefore a 'restricted correspondence analyses' in the sense that the site scores are restricted to be linear combinations of measured environmental variables and thereby maximizes the dispersion of the categories scores. By incorporating this restriction in the two-way weighted averaging algorithm of CA, we obtain an algorithm for CCA. The fitted values of this regression are by definition a linear combination of environmental variables (Eq. 4.3) (Ter Braak 1995). The eigenvalues in CCA are usually smaller than those in CA because of the restrictions imposed on the site scores in CCA:

$$x_i = c_0 + c_1 z_{1i} + c_2 z_{2i} + \ldots\ldots\cdots + c_j z_{\mathrm{ji}} \tag{4.10}$$

where z_{ji} is the value of an environmental variable j at site i, c_j is the weight (not necessarily positive) belonging to that variable and x_i is the value of the resulting compound environmental variable at site i.

The parameters of the final regression in the iteration process are the best weights, also called canonical coefficients. The multiple correlations of this regression are called category-environment correlations (originally species-environment correlations in ecological terms) (Ter Braak 1995).

Each of these correlations is between the site scores that are weighted averages of the categories scores and the site scores that are a linear combination of the environmental variables. Thus, the species-environment correlation is a measure of association between categories and the environment. However, this correlation may not be the ideal one; an axis with small eigenvalue may have misleadingly high category-environment correlations. The importance of the association is expressed well by the eigenvalue, because the eigenvalue measures how much variation in the categories data is explained by the axis and, hence, by the environmental variables.

The restriction of CCA becomes less with increasing number of environmental variables. Ter Braak and Prentice (1988) found CCA to be extremely robust even when the assumption of unimodal response does not hold. CCA can also accommodate nominal explanatory variables by defining dummy variables.

In general, the choice between a linear and a non-linear ordination method is not a matter of personal preferences. Where the gradient is short, there are sound statistical reasons to use linear methods like PCA and RDA. As the gradient length increases, linear methods become ineffective; Gaussian methods such as CA and CCA become more suitable (Ter Braak and Prentice 1988).

4.2.5 Redundancy Analysis

As mentioned previously, RDA is comparable with COR and also can be seen as a constrained canonical form of PCA because the site scores are linear combinations of environmental attributes (Ter Braak and Prentice 1988; Odeh et al. 1991; Ter Braak 1995). In other words, RDA is the technique selecting the linear combinations of environmental variables that gives the smallest total residual sum of squares. RDA is thus an intermediate between PCA and separate multiple regression models for each of the soils, i.e. it is a constrained ordination, but it is also a constrained form of multiple regression (Ter Braak and Prentice 1988). One of the attractive features of RDA is that it leads to an ordination diagram that simultaneously displays (1) the main pattern of soil variation as far as this variation can be explained by the environmental variables and (2) the main pattern in the correlation coefficients between the soils and each of the environmental variables.

4.3 Biplots

4.3.1 Linear Biplots

A useful product of multivariate ordination analysis is the biplot, sometimes referred to as an ordination diagram. In simple terms, a biplot is characterized by a data matrix X, a method of ordination or multidimensional scaling or approximation (e.g. previously mentioned PCA, COR or RDA) and a choice of inter-sample distance representation (e.g. Euclidean distance and its special case of Pythagorean distance, Mahalanobis distance, χ^2 distance) (Gower and Hand 1996).

The term biplot was first introduced by Gabriel (1971). He originated the PCA biplots, canonical biplots, and (together with Bradu) the biadditive model diagnostic biplots (Gower and Hand 1996).

Let us assume that X is a $n \times p$ matrix containing numerical information on p variables for each of n samples. Then, in a classical biplot, the samples are shown as n points and the variables are shown as p vectors relative to the same axes and origin. Underlying the classical biplot is the singular value decomposition (SVD) which is expressed as follows (Gower and Harding 1988):

$$X = \mathrm{URV}' \tag{4.11}$$

where U of size $n \times n$ and V of size $p \times p$ are orthogonal matrices and R is a matrix with non-negative entries with $R_{i,j} = \gamma_i$ and if $i = j$, otherwise $R_{i,j} = 0$; further $\gamma_1 \geq \cdots \geq \gamma_r$ and $\gamma_i = 0$ when $> r$. Here, r is the rank of X, so usually $r = \min(n, p)$. As discussed in Gower and Harding (1988), the algebraic result of this expression has been known from about 1880, but its statistical importance resulted from the work of Eckart and Young (1936). Eckart and Young (1936) showed that the best least-squares fit of rank s to X can be obtained by replacing R in (Eq. 4.11) by R_s in which γ_i is set to zero for all $i > s$. Readers are referred to Gower and Harding (1988) for further details on biplots.

4.3.1.1 Case Study

Let us use the Edgeroi dataset described in McGarry et al. (1989) to illustrate some of the most common linear biplots. The Edgeroi dataset contains 210 sampled sites on a systematic, equilateral triangular grid with a spacing of 2.8 km between sites, plus 131 samples obtained through a different sampling design as shown in Fig. 4.1. A comprehensive suite of soil properties was obtained for each of the samples, including exchangeable cations (Ca^{2+}, Mg^{2+}, Na^{+} and K^{+} in mmol kg^{-1}); pH; soil colour (CIElab); sand, slit and clay content (%); carbon

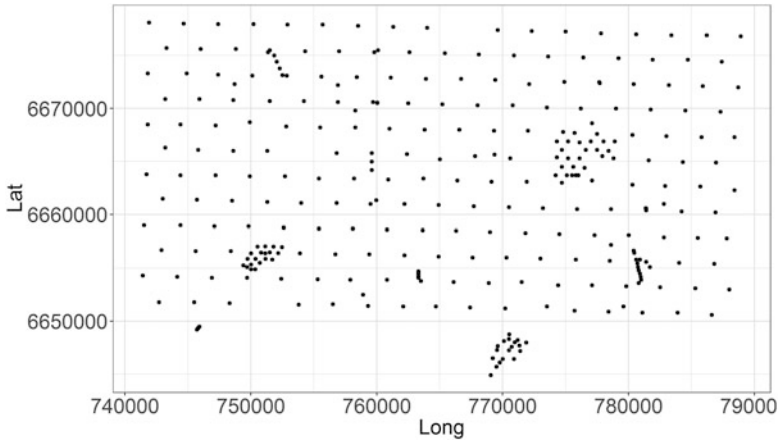


Fig. 4.1 210 Sampling points obtained for the Edgeroi dataset, employing two different sampling designs

content (%); and calcium carbonates (%). In addition, a comprehensive suite of environmental variables is available, including gamma radiometrics (^{232}Th, ^{238}U and ^{40}K), a digital elevation model and its derivatives such as MRVBF (multiresolution valley bottom flatness), slope, curvature and TWI (terrain wetness index) and products from satellite imagery (e.g. NDVI (normalized difference vegetation index)).

The most common dimensional reduction approach is the PCA which is a form of metric scaling that approximates the distance d_{ij} between the ith and jth samples defined by

$$d^2{}_{\mathrm{ij}} = \sum_{k=1}^{p} \left(x_{\mathrm{ik}} - x_{\mathrm{jk}}\right)^2 \tag{4.12}$$

This is known as Pythagorean distance or often, somewhat misleadingly, as Euclidean distance.

Figure 4.2 shows a common biplot picturing known linear relations between soil attributes (e.g. carbon content vs colour lightness and clay vs sand content in bold arrows).

On the other hand, it is possible to 'constrain' these relations between one set of attributes to another dataset as described in Sect. 4.2.2 (canonical correlation analysis). Figures 4.3 and 4.4 show examples of how these two sets of variables can be related.

The immediate effect of the 'conditioning' of this new multidimensional space is evident when comparing Fig. 4.3 with Fig. 4.2, as the linear relations presented previously even though still present are now somehow modified by the constraining relations of the landscape dataset.

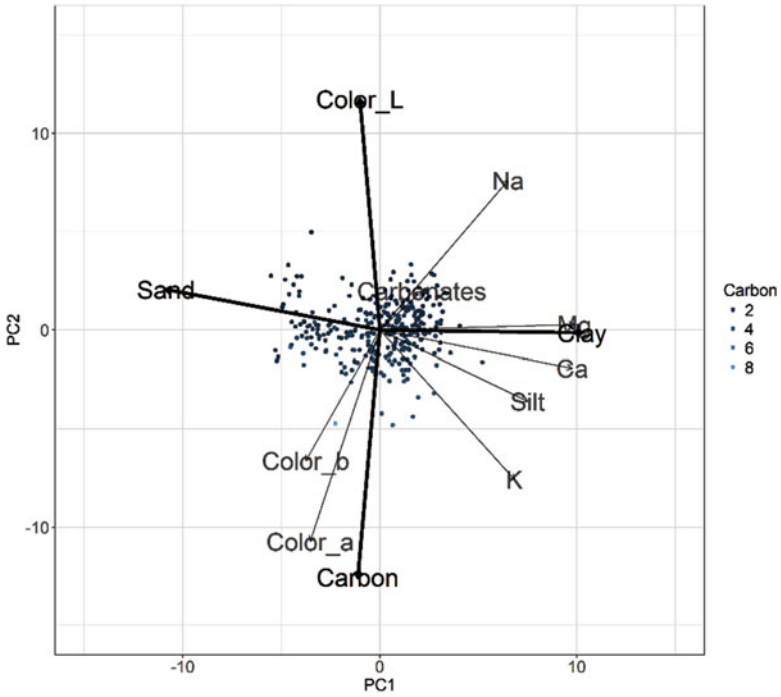


Fig. 4.2 PCA biplot. Note that the colour scale represents carbon content, where carbon content is increasing in the direction of the vector carbon. Note: The reader is referred to Sect. 5.2 of this book for explanations of the colour space used in this example (CIELab)

If a direct comparison is desired, a word of caution needs to be issued first, since the different canonical axes (from the soil attributes and landscape dataset) represent a different ordination; hence, the observed samples will have a different projection in both spaces (Fig. 4.4).

An intuitive way of visualizing this phenomenon can be observed in Fig. 4.5 where both first canonical axes of the soil and landscape ordinations are plotted together.

It is evident that both axes are conditioned to best fit the linear relations between the two datasets, by doing so, revealing the known connections between landscape and soil attributes; however, a compound plot may be misleading if these relations are meant to be represented in a single figure and their ordinations are not sufficiently significant (which is not the case for our dataset) (Fig. 4.6).

4.3.1.2 Categorical Data

A different approach is needed when the input information comes in categorical form; however, the practical meaning is similar to the previous section. A soil

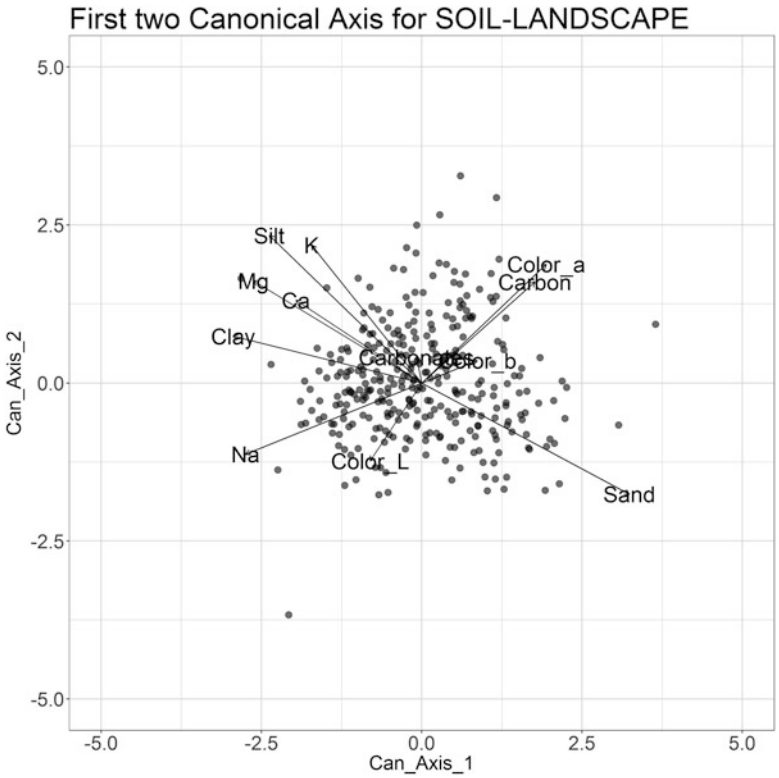


Fig. 4.3 COR biplot where samples are projected in a matrix ordinated by soil attributes. *See* Sect. 4.2.2 of this chapter. Note: The reader is referred to Sect. 5.2 of this book for explanations of the colour space used in this example (CIELab)

science example of constrained forms of categorical ordination (i.e. CCA and RDA) can be found in the work of Odeh et al. (1991). Odeh et al. (1991) attempted to elucidate the relations between a set of soil attributes usually presented as categories or frequencies (e.g. soil structure grade or mottle abundance) and landscape attributes (e.g. curvature, upslope distance, etc.) that may or may not explain the soil variation present.

Figure 4.7 shows the CCA analysis performed for their dataset. As observed in previous sections, similar relations are found by using this analysis, e.g. the known relations between clay and sand content (as seen in Fig. 4.7, with *acla*, topsoil clay content, and *asan*, topsoil sand content).

The usefulness of this kind of ordination is that it permits, for example, to relate structure grade (presented as a value ranking between 0 and 5) with continuous

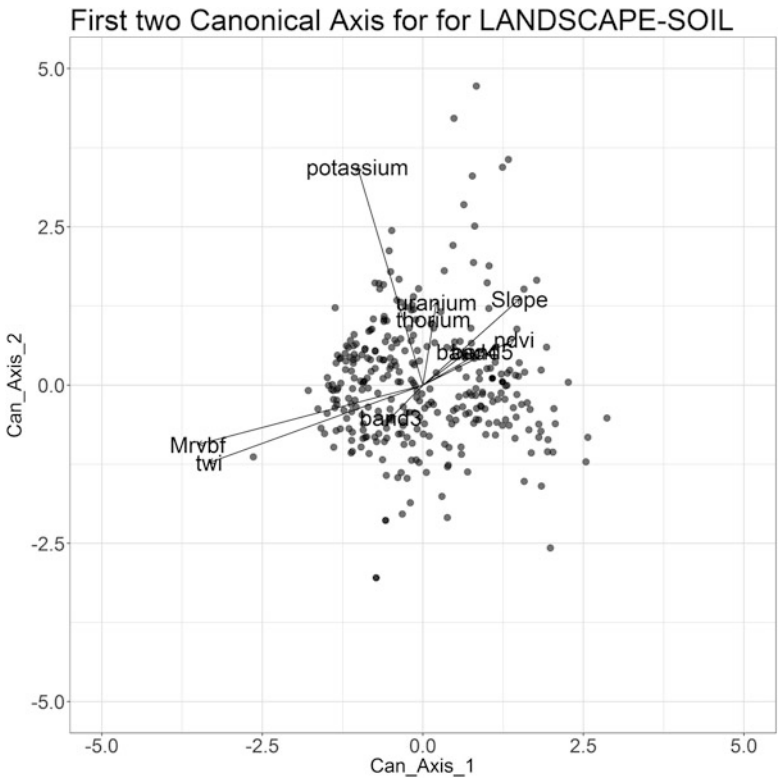


Fig. 4.4 COR biplot where samples are projected in a matrix ordinated by landscape attributes. Note the different projection of the sample points

data such as clay content (compositional data; *see* Sect. 4.4 regarding compositional data) allowing at the same time to constrain the linear relations to those found in external environmental information such as rock depth or aspect and visualizing all of these in only two dimensions.

A different approach (also constrained to environment attributes) is presented in Fig. 4.8, where an RDA analysis was performed on the same dataset.

Considering that only a few of the measured soil variables had unimodal relationships among themselves or with the environmental attributes used, the authors considered it as advantageous to use linear relations (smallest total residual sum of squares in RDA) when constraining the soil matrix with the environment matrix instead of using their correlations (CCA), thus recommending RDA over CCA under the studied conditions.

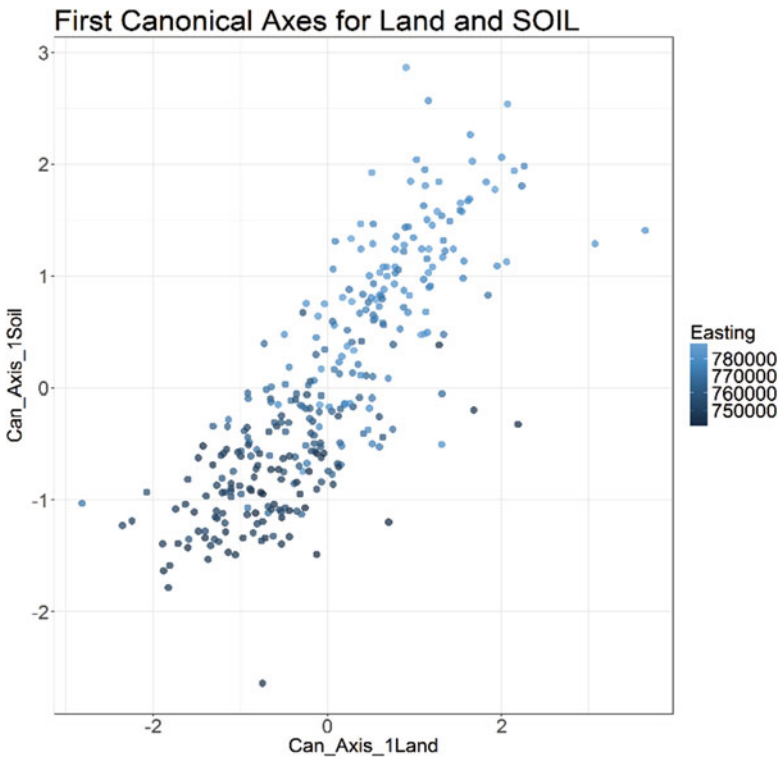


Fig. 4.5 COR plot where both the first canonical axes of the soil and landscape ordinations are plotted together. Colour scale represents easting, showing the same E-W pattern in both multidimensional vectors, i.e. soil and landscape

4.3.2 Non-linear Biplots

The previous section has interpreted the classical biplot as a component analysis augmented by information on the variables obtained by projecting notional sample units with values concentrated in a single variable. In general, there are many other ways of defining d_{ij} in terms of the sample values. Gower and Legendre (1986) discuss some of the more commonly used definitions.

Gower and Harding (1988) give an example for a classical, non-linear biplot with d_{ij} defined as in Eq. 4.12 (Fig. 4.9). The data used to populate this example were sourced from a soil survey conducted in Pembrokeshire, Wales (Glamorganshire soils). The concentration of 12 trace elements in parts per million (variables shown as integers in Fig. 4.9) was determined at 15 sampling sites (shown as lower-case letters in Fig. 4.9), and their logs taken to calculate the values x_{ik} which are plotted as deviations from the mean. In the example, we can see that the trajectories which represent the range of the different variables clearly differ in length which

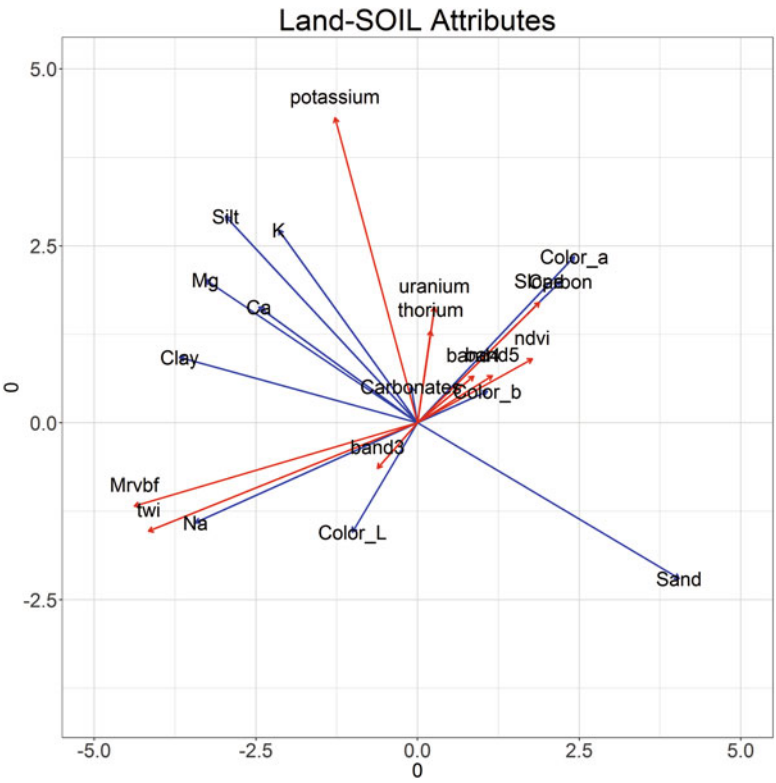


Fig. 4.6 COR plot where two sets of variables (soil and landscape attributes) are presented in the same two dimensional space

reflects their contribution to the ordination *Y*. Here, the variables 5, 7 and 9 are best represented, and consequently the other variables have little weight. Sample g (as well as e and k), for example, is shown as an outlier and lies well beyond the range of variable 9; other variables therefore contribute to g. Here, g could only be interpolated by using extreme values of 5 (negative), 7 (negative) and 9 (positive). On the other hand, even if a sample lies well within the range of a variable, this does not necessarily mean that the variable is a major contributor. This is the case for sample d which is close to 5 but also includes large contributions from variables 7 and 9.

A subset of the above described dataset was used to create the non-linear biplot shown in Fig. 4.10. Here, only 4 of the originally 12 trace elements (variables 5, 7, 9 and 12) were used, i.e. only effective trajectories were chosen, and all trajectories were scaled from the common point O by a factor of 4, which provides a less diffuse and more balanced interpretation of the biplot. General trends of the dataset remain, with sampling site g still being identified as an obvious outlier.

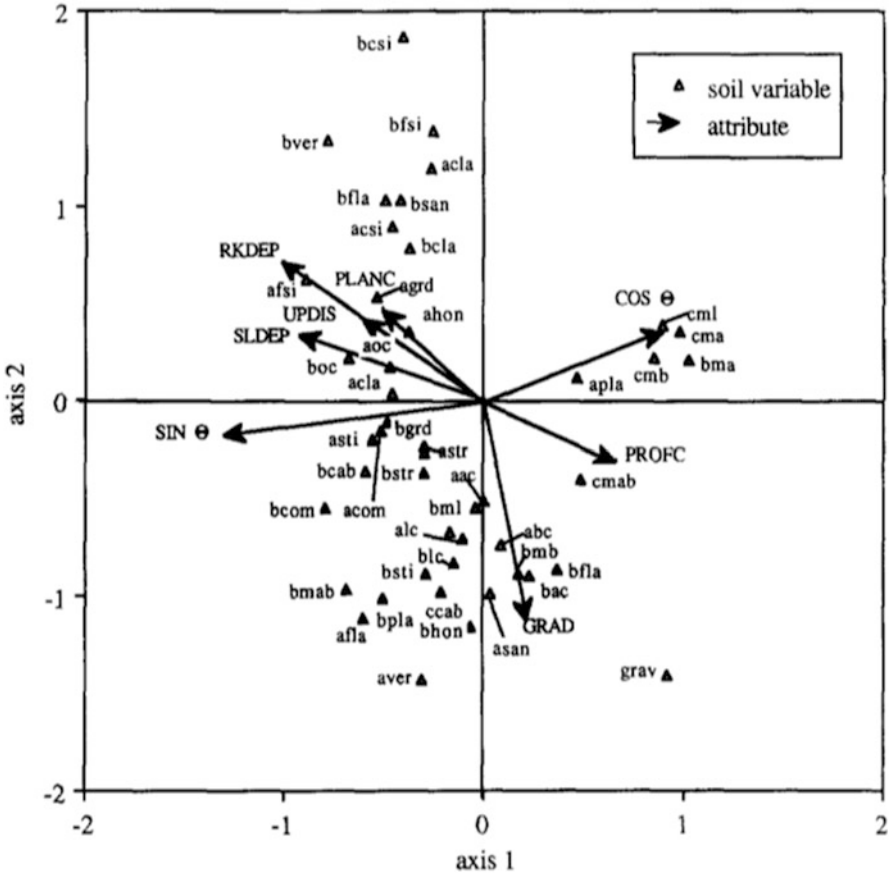


Fig. 4.7 CCA biplot where both soil attributes and environmental attributes are plotted together. Note: The nomenclature is described in Table 4.1 (Source: Odeh et al. 1991)

4.4 Compositional Data

4.4.1 Background

Often we describe a volume of soil with a set of properties that are related to each other that sums to that soil volume or to its mass. Such sets of properties are called compositional variables or compositions. The most common example in soil science would be describing the particle size of the mineral fraction of the soil with the variables sand, silt and clay. The relative proportions of the individual particle-size fraction (PSF) are what constitute the soil texture. The importance of soil texture cannot be overemphasized. The soil texture, and indeed the particle-size distribution, determines, in part, water, heat and nutrient fluxes, water and nutrient holding capacity and soil structural form and stability. In particular the

Table 4.1 Soil variables and environmental variables used

Variable	Symbol	Variable	Symbol	Env. attribute	Symbol
Colour L	*lc*	Soil pH	*ph*	Solum depth	*SLDEP*
Colour a	*ac*	Structure grade	*grd*	Bedrock depth	*RKDEP*
Colour b	*bc*	Structure horizontality	*hon*	Slope gradient	*GRAD*
Mottle colour L	*ml*	Structure verticality	*ver*	Profile convexity	*PROFC*
Mottle colour a	*ma*	Structure face flatness	*fla*	Plan convexity	*PLANC*
Mottle colour b	*mb*	Structure accommodation	*com*	Upslope distance	*UPDIS*
Cutan abundance	*cab*	Gravel	*grav*	Upslope area	*UPA*
Soil strength	*str*	Clay	*cla*	Square root of a	*UPSQ*
Soil plasticity	*pla*	Silt	*sil*	Sine aspect in deg (θ)	*SIN(θ)*
Soil stickiness	*sti*	Sand	*san*	Cosine aspect in deg (θ)	*COS(θ)*
Organic carbon	*om*	Electric Conductivity	*ec*		
		Prefix			
Topsoil	*a*	Subsoil	*b*	Parent material	*c*

Adapted from Odeh et al. (1991). Note: The reader is also referred to Sect. 5.3 of this book for further explanations of the variables in relation to soil structure

clay fraction, as the active constituent of the composition, could be incorporated in pedotransfer functions to predict material fluxes (e.g. Arya et al. 1999) and other soil properties (Sinowski et al. 1997) (refer to Chap. 7 of this book). Other types of compositional soil data could include soil cations that contribute to effective cation exchange capacity, e.g. the relative proportions of calcium Ca^{2+}, magnesium (Mg^{2+}), potassium (K^{+}), sodium (Na^{+}) and aluminium (Al^{3+}). Soil class memberships are further examples of what are considered to be forms of compositional data in that all the memberships will sum to 1 or 100, etc. (McBratney et al. 1992; de Gruijter et al. 1997).

Compositional data have important and particular properties that preclude the application of standard statistical techniques on such data in raw form – although this does this not preclude people from using standard techniques to analyse compositional data, however. Compositional data are vectors of non-negative components showing the relative weight or importance of a set of parts in a total, meaning that the total sum of a compositional vector is considered irrelevant. Another property is that when analysing compositional data, no individual component can be interpreted isolated from the other. For compositional data, the sample space (or set of possible values) is called the simplex, which is the set of vectors of positive (or zero) components which could be described as a proportion, percentage or any other closed-form expression such as parts per million (ppm) or similar. Staying with the soil texture example, mapping the simplex can be illustrated with a ternary plot as shown in Fig. 4.11.

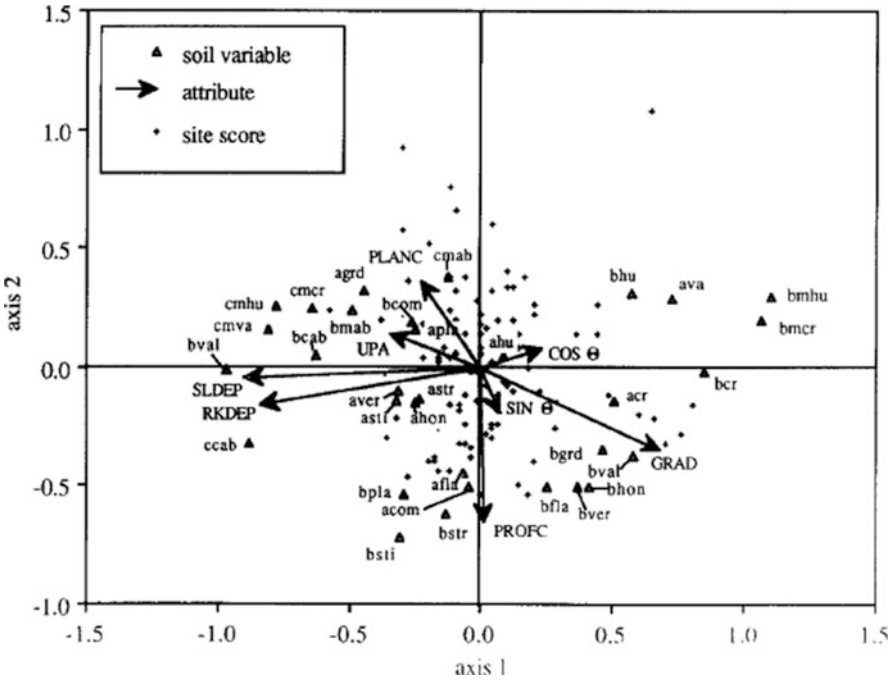


Fig. 4.8 RDA biplot where both soil attributes, environmental attributes and projected site scores, are plotted together (different to Fig. 4.7, this graphic shows standardized data, resulting in higher variance explained on each axis) (Source: Odeh et al. 1991)

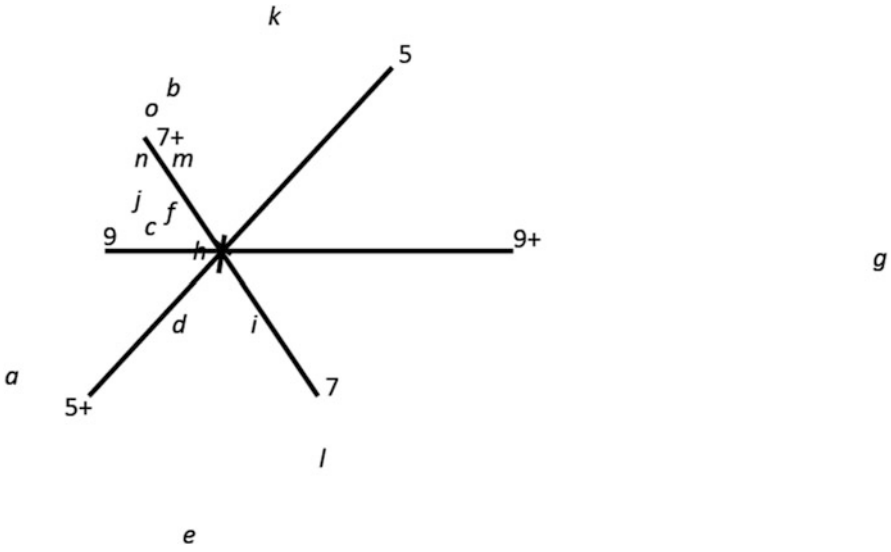


Fig. 4.9 Non-linear biplot using Pythagorean distance. *Lower-case letters* refer to samples (15 Glamorganshire soils, Wales); *numbers* which label linear trajectories refer to variables. The positive end of each trajectory is indicated (Redrawn after Gower and Harding (1988))

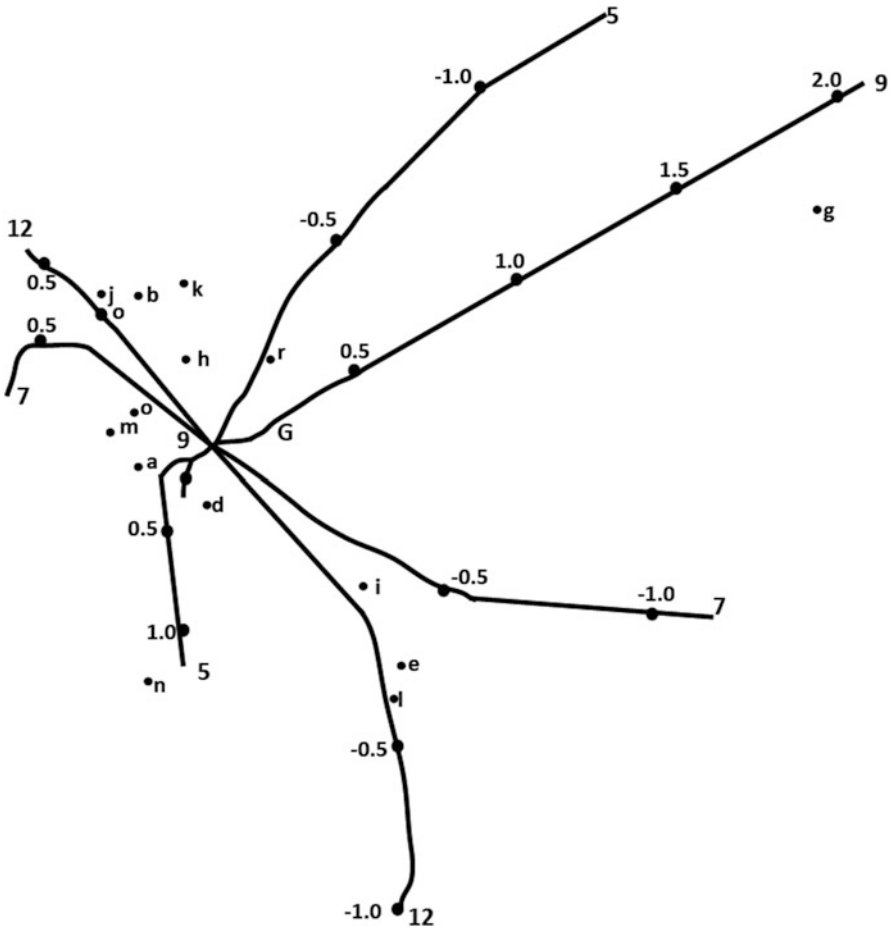


Fig. 4.10 Non-linear biplot using the square root of Minkowski L1 distance. *Lower-case letters* refer to samples (15 Glamorganshire soils, Wales); *numbers* which label linear trajectories refer to variables. The centroid G of the ordination and the common point O, unlabelled, of the trajectories differ (Redrawn after Gower and Hand 1996)

Because of these specific properties, compositional data are not amenable to analysis by common statistical methods designed for use with unconstrained data (Chayes 1960; Aitchison 1986). For example, standard techniques are designed to be used with data that are free to range from $-\infty$ to $+\infty$. A consequence of not treating compositional data in an appropriate manner is likely to result in misleading outcomes. Figure 4.12 shows biplots of soil texture from the Edgeroi dataset (McGarry et al. 1989). The biplot on the left shows the principal components of the soil texture data without recognition of its compositional nature. The biplot on the right is the same data, but they have been transformed using a centre log-ratio (clr) transform for compositional data (more information about this transform

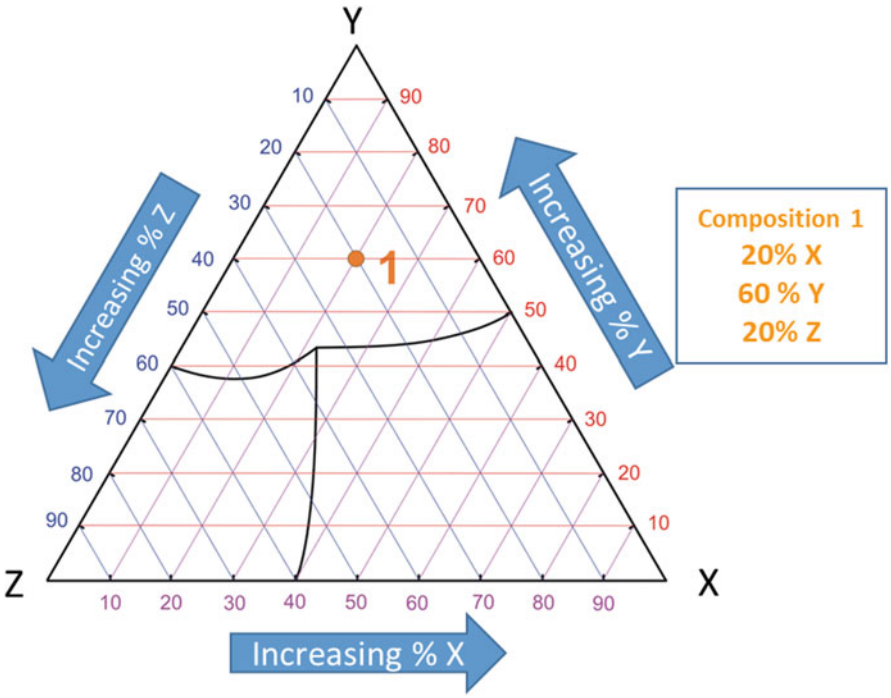


Fig. 4.11 Ternary diagram plotted as an equilateral triangle, with X, Y and Z components defining the compositional system placed at the apices of the triangle. Note that each apex is 100% of that component and 0% of the other two

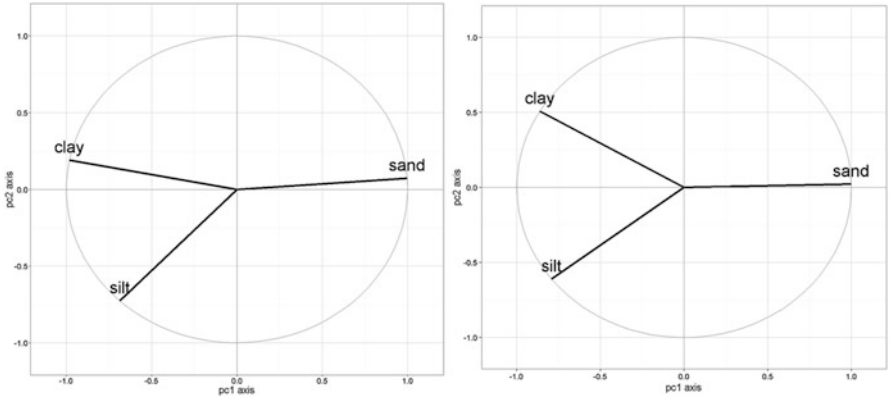


Fig. 4.12 Biplots of untransformed (*left*) and clr-transformed (*right*) compositional soil texture data

is described further on). Despite being the same data, what is clear from these biplots is that the angles of the variable rays are different between them, implying different correlation relationships between the variables when it is and is not transformed. Consequently, analyses performed on untransformed data are likely to reveal outcomes that could be potentially misleading.

4.4.2 *Historical Perspective Behind Compositional Data Analysis*

Historically, the starting point for compositional data analysis is arguably the paper of Pearson (1897), which first identified the problem of *spurious correlation* between ratios of variables. Pearson used the example that if *X*, *Y* and *Z* are uncorrelated, then *X/Z* and *Y/Z* will not be uncorrelated. Pearson then looked at how to adjust the correlations to take into account the *spurious correlation* caused by the scaling. However, this ignores the implicit constraint that scaling only makes sense if the scaling variable is either strictly positive or strictly negative. In short, this approach ignores the range of the data and does not assist in understanding the process by which the data are generated. Tanner (1949) made the essential point that a log transform of the data may avoid the problem and that checking whether the original or log-transformed data follow a normal distribution may provide some guidance as to whether a transform is needed.

Chayes (1960) later made the explicit connection between Pearson's work and compositional data and showed that some of the correlations between components of the composition must be negative because of the unit sum constraint. However, he was unable to propose a means to model such data in a way that removed the effect of the constraint.

The difficulty of interpreting compositional data is well illustrated by the following trivial example: '...If one analyses the contents of a jar half-filled with sand and finds, by a random sample, that it contained (by volume) about 20% quartz, 30% feldspar, 40% rock fragments, and 10% miscellaneous constituents, then, if the volume of the jar were doubled by addition of grains of pure quartz, a second random sample would reveal that the jar contains 60% quartz, 15% feldspar, 20% rock fragments, and 5% miscellaneous. Feldspar, rock fragments, and miscellaneous constituents appear pair-wise positively correlated and all three appear negatively correlated with the quartz abundance. Also, all four components have shifted mean values despite the fact that only the quartz content of the jar changed...' (Woronow 1990).

Although the interdependency of compositional variables has been recognized for over a century (Pearson 1897), appropriate statistical methods for analysing such data have been extremely slow to emerge. Researchers who recognized the problems associated with statistical analysis of compositions were unable to provide feasible solutions. Many scientists therefore decided to abandon statistical analysis

of compositions altogether. Others wished away or ignored the awkward constraints on compositions to justify the use of standard statistical methods for compositional data analysis (Aitchison 1986). A naïve solution and one commonly practised in the analysis of compositional data is to use standard statistical data analysis on *D-1* components (which is of size *D*) and then evaluate the remaining component by difference at the end of the analysis.

4.4.3 A Recent Approach: Log Ratios

John Aitchison is probably recognized as the person to have laid the foundations of a new approach to the statistical analysis of compositional data. The seminal work that is described in Aitchison (1986) is a culmination of all the efforts on this topic and obviously an invaluable resource for those wanting to appreciate it more at the granular level.

It has been found that an appropriate analysis of compositional data is based on a log-ratio representation of the data – and several are available. For example, Aitchison (1982) introduced the additive log-ratio (alr) and centred log-ratio (clr) transformations and Egozcue et al. (2003) the isometric log-ratio (ilr) transformation. Using these transformations, a composition is represented as real vectors, which are often referred to as coordinates, or sometimes coefficients (Pawlowsky-Glahn and Egozcue 2001).

Numerically, coordinates or coefficients are easy to compute. If we consider a sandy clay soil which we will say has 55% sand, 15% silt and 30% clay, this represents a three-part ($D = 3$) composition $x = [55,15,30]$. The three representations would be written as a vector with two components for the alr and ilr coordinates, and with three components for the clr coefficients, as follows (note that numbers are limited to two decimal places throughout):

Additive log ratio

$$\text{alr}(x) = \text{alr}\,[55, 15, 30] = \left[\ln\frac{55}{30}, \ln\frac{15}{30}\right]$$

$$= [\ln(1.83), \ln(0.5)\,] = [\,0.60, -0.69]$$

The inverse of alr(x) is

$$x = \frac{[\exp(0.60); \exp(-0.69)\,; 1]}{\exp(0.60) + \exp(-0.69) + 1}$$

$$= \left[\frac{1.82}{3.32}, \frac{0.50}{3.32}, \frac{1}{3.32}\right]$$

$$= [0.55, 0.15, 0.30]$$

Centred log ratio

$$\mathrm{clr}(x) = \mathrm{clr}\,[55, 15, 30]$$

$$= \left[\ln \frac{55}{(55 \cdot 15 \cdot 30)^{\frac{1}{3}}}, \ln \frac{15}{(55 \cdot 15 \cdot 30)^{\frac{1}{3}}}, \ln \frac{30}{(55 \cdot 15 \cdot 30)^{\frac{1}{3}}} \right]$$

$$= \left[\ln \frac{55}{(24750)^{\frac{1}{3}}}, \ln \frac{15}{(24750)^{\frac{1}{3}}}, \ln \frac{30}{(24750)^{\frac{1}{3}}} \right]$$

$$= \left[\ln \frac{55}{29.14}, \ln \frac{15}{29.14}, \ln \frac{30}{29.14} \right]$$

$$= [\ln(1.89), \ln(0.51), \ln(1.03)\,]$$

$$= [0.64, -0.66, 0.03]$$

The inverse of $clr(x)$ is

$$x = \frac{[\exp(0.64); \exp(-0.66)\,; \exp(0.03)]}{\exp(0.64) + \exp(-0.66) + \exp(0.03)}$$

$$= \left[\frac{1.90}{3.44}, \frac{0.52}{3.44}, \frac{1.03}{3.44} \right]$$

$$= [0.55, 0.15, 0.30]$$

Isometric log ratio

$$\mathrm{ilr}(x) = t\,(\boldsymbol{V}) \cdot \mathrm{clr}\,[55, 15, 30]$$

$\boldsymbol{V}$ is a matrix of D rows and D-1 columns such that $\boldsymbol{V} \cdot t(\boldsymbol{V}) = \boldsymbol{I}_{D-1}$(identity matrix of D-1 elements and $\boldsymbol{V} \cdot t(\boldsymbol{V}) = \boldsymbol{I}_D + a1$, where a may be any value and $\mathbf{1}$ is a matrix full of ones). From Egozcue et al. (2003), the matrix elements of $\boldsymbol{V}$ are the basis elements for the canonical basis of the clr-plane needed for the ilr transform:

$$\boldsymbol{V}\ [55, 15, 30] = \begin{bmatrix} -0.71 & -0.41 \\ 0.71 & -0.41 \\ 0 & 0.82 \end{bmatrix}$$

$$\text{ilr}(x) = t \begin{bmatrix} -0.71 & -0.41 \\ 0.71 & -0.41 \\ 0 & 0.82 \end{bmatrix} \cdot [0.64, -0.66, 0.03]$$

$$= [-0.92, 0.04]$$

The inverse of $ilr(x)$ is performed by converting ilr to clr and then performing the inverse as for clr:

$$\text{ilr}(x) \rightarrow \text{clr}(x) = \text{ilr}(x) \cdot t\,(\boldsymbol{V})$$

$$\text{clr}(x) = [-0.92, 0.04] \cdot t \begin{bmatrix} -0.71 & -0.41 \\ 0.71 & -0.41 \\ 0 & 0.82 \end{bmatrix}$$

$$x = \frac{[\exp(0.64); \exp(-0.66); \exp(0.03)]}{\exp(0.64) + \exp(-0.66) + \exp(0.03)}$$

$$= \left[\frac{1.90}{3.44}, \frac{0.52}{3.44}, \frac{1.03}{3.44} \right]$$

$$= [0.55, 0.15, 0.30]$$

The general equations for alr and clr can be found in Aitchison (1986) and those for the ilr are in Egozcue et al. (2003). The three representations have different properties. As described in Pawlowsky-Glahn and Egozcue (2001) from Aitchison (1986), the alr coordinates are D-1 components which are divided by the remaining component and logarithms taken. In the above example, the last component is used as the denominator, but it could be any of the components. The resulting log ratios are real variables that can be analysed using standard statistical techniques. The clr coefficients are obtained by dividing the components by the geometric mean of the parts and then taking logarithms. While clr coefficients are useful in the computation of biplots, their one major drawback is that they necessarily sum to zero. This means that clr-transformed observations lie on a plane in D-dimensional real space. Consequently, care should be taken in ensuing analysis because covariance and correlation matrices are singular (their determinant is necessarily zero) (Pawlowsky-Glahn and Egozcue 2001). Expressions for the calculation of ilr coordinates are more complex, and there are different rules on how to generate those (Egozcue et al. 2003). Their advantage lies in the fact that they are coordinates in an orthogonal system, and thus, any classical multivariate statistical technique can be used straightway to study them. In summary, log-ratio coordinates and coefficients of random compositions are real random variables, as they are free to range from $-\infty$ to $+\infty$, and thus it is possible to undertake multivariate statistical analysis using them.

4.4.4 Analysing Compositional Data as Regionalized Variables

Some of the statistical techniques of interest when working with compositional data involve those related to geospatial prediction, i.e. the creation of digital soil maps. Detailed discussion on digital soil mapping and geostatistics can be found in Chaps. 12 and 11, respectively. When working with compositional data, most mapping studies are those involving the mapping of soil texture components. Compositional treatment of soil class memberships has also been investigated, e.g. McBratney et al. (1992). The work by McBratney et al. (1992) considered the 'symmetric' or clr transform as a prior step to mapping soil membership classes using ordinary kriging. Also, interested in soil class memberships, de Gruijter et al. (1997) proposed a compositional data kriging algorithm which is for all intents and purposes is an extension of ordinary kriging, but with constraints. The method of de Gruijter et al. (1997) does not actually use a log-ratio transformation as a prior step, rather it imposes conditions, in addition to the unbiasedness conditions upon the ordinary kriging system. The issue with this approach is that it is computationally laborious as the constrained kriging equations must be solved numerically.

With regard to soil texture analysis, Odeh et al. (2003), Lark and Bishop (2007) and Huang et al. (2014) interestingly all used the alr transformation prior to geostatistical analysis. Ordinary kriging was used in Odeh et al. (2003), co-kriging in Lark and Bishop (2007), while Huang et al. (2014) used a regression model, relating the transformed coordinates to a suite of environmental covariates. As described by Lark and Bishop (2007), the preference to use alr is helped by the fact that in the case of co-kriging, the mapped estimates are not affected if the order of the elements in the composition is changed (i.e. changing the denominator to another element that is not the last). This property is called permutation invariance (Pawlowsky-Glahn and Olea 2004). Also beneficial for geostatistical purposes is that the cross-covariance structure of an alr-transformed variable contains all the information on the spatial dependence of the untransformed variable. This is also provided by other transforms, and unlike the clr transform, the covariance matrices of alr-transformed data are not singular. Data transformed with ilr would also have the same ideal properties as for alr, and this has the added advantage of not having to determine if the data exhibits permutation invariance or not.

Considerations about back-transformation of transformed variables are also necessary. The inverse back-transforms as detailed above may be used, but as Pawlowsky-Glahn and Olea (2004) show, this back-transform is necessarily biased, and the unbiased transform is unknown. As Lark and Bishop (2007) point out, this quality of bias was one of the reasons why de Gruijter et al. (1997) developed compositional kriging. Pawlowsky-Glahn and Olea (2004) do indicate that a back-transform can be obtained numerically. Yet Lark and Bishop (2007) were unable to determine any consistent advantage in doing so when comparing results obtained via unbiased back-transformation with those that were obtained using

the simple alr back-transform equation listed above. Other studies such as Huang et al. (2014) did not consider the relative advantages of biased and unbiased back-transformations.

4.5 Conclusions

In this chapter we have described and discussed multivariate methods that have been used in soil science. For some properties, i.e. particle-size fractions, we have shown that we must account for their compositional nature (with a log-ratio transform) as otherwise we may find spurious relationships in our analysis. Another subtlety for multivariate analysis is that in many situations, our data has an associated spatial location, and therefore there may be an implicit spatial autocorrelation between our observations. The concept of spatial autocorrelation will be discussed in later chapters, but in terms of multivariate analysis, future work should consider this. Jombart et al. (2008) in a genetics study modified PCA based on the Moran's I spatial correlation measure. There is potential to adapt this to geographic space encountered in soil science.

Finally, we live in the age of readily accessible data about soil, be it from sensors measuring parts of the electromagnetic spectrum or from DNA sequencing of soil biota. Such datasets can be vast for even just one soil sample, and multivariate methods are crucial for understanding such datasets. Proficiency in there use will be a necessary skill for all soil scientists.

References

Aitchison J (1982) The statistical analysis of compositional data. J R Stat Soc.Series B (Methodological) 44(2):139–177

Aitchison J (1986) The statistical analysis of compositional data. Chapman & Hall, London

Arya LM, Leij FJ, Shouse PJ, van Genuchten MT (1999) Relationship between the hydraulic conductivity function and the particle-size distribution. Soil Sci Soc Am J 63(5):1063–1070

Braimoh AK (2004) Seasonal migration and land-use change in Ghana. Land Degrad Dev 15: 37–47

Chayes F (1960) On correlation between variables of constant sum. J Geophys Res 65(12): 4185–4193

Crave A, Gascuel-Odoux C (1997) The influence of topography on time and space distribution of soil surface water content. Hydrol Process 11:203–210

De Gruijter JJ, Walvoort DJJ, van Gams PFM (1997) Continuous soil maps – a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77(2):169–195

Dray S, Saïd S, Débias F (2008) Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. J Veg Sci 19:45–56

Eckart C, Young G (1936) The approximation of one matrix by another of lower rank. Psychometrika 1:211–218

Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Math Geol 35(3):279–300
Gabriel KR (1971) The biplot graphic display of matrices with application to principal component analysis. Biometrika 58:453–467
Gower JC, Hand DJ (1996) Biplots. Chapman & Hall, London
Gower JC, Harding SA (1988) Nonlinear biplots. Biometrika 75:445–455
Gower JC, Legendre P (1986) Metric and Euclidean properties of dissimilarity coefficients. J Classif 3:5–48
Graffelman J, Tuft R (2004) Site scores and conditional biplots in canonical correspondence analysis. Environmetrics 15:67–80
Hotelling H (1936) Relations between two sets of variates. Biometrika 28(3/4):321–377
Huang J, Subasinghe R, Triantafilis J (2014) Mapping particle-size fractions as a composition using additive log-ratio transformation and ancillary data. Soil Sci Soc Am J 78(6):1967–1976
Islam KR, Weil RR (2000) Soil quality indicator properties in mid-Atlantic soils as influenced by conservation management. J Soil Water Conserv 55:69–78
Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic patterns in genetic variability by a new multivariate method. Heredity 101:92–103
Karunaratne SB, Bishop TFA, Odeh IOA, Baldock JA, Marchant BP (2014) Estimating change in soil organic carbon using legacy data as the baseline: issues, approaches and lessons to learn. Soil Res 52(4):349–365
Kendall MG, Stuart A (1977) The advanced theory of statistics. C. Griffin, London
Kenkel NC (2006) On selecting an appropriate multivariate analysis. Can J Plant Sci 86:663–676
Kenkel NC, Derksen DA, Thomas AG, Watson PR (2002) Multivariate analysis in weed science research. Weed Sci 50:281–292
Lark RM, Bishop TFA (2007) Cokriging particle size fractions of the soil. Eur J Soil Sci 58(3): 763–774
Latty EF, Canham CD, Marks PL (2004) The effects of land-use history on soil properties and nutrient dynamics in northern hardwood forests of the Adirondack Mountains. Ecosystems 7:193–207
Legendre P, Legendre L (2012) Chapter 11 – Canonical analysis. In: Legendre P, Legendre L (eds) Numerical ecology. Developments in Environmental Modelling. Elsevier, Amsterdam, pp 625–710
McBratney AB, De Gruijter JJ, Brus DJ (1992) Spatial prediction and mapping of continuous soil classes. Geoderma 54(1):39–64
McGarry D, Ward WT, McBratney AB (1989) Soil studies in the lower Namoi Valley: methods and data, The Edgeroi Dataset. CSI4RO Division of Soils, Adelaide. 2 vols
McGrath SP, Zhao FJ, Lombi E (2001) Plant and rhizosphere processes involved in phytoremediation of metal-contaminated soils. Plant Soil 232:207–214
Odeh IOA, Chittleborough DJ, McBratney AB (1991) Elucidation of soil-landform interrelationships by canonical ordination analysis. Geoderma 49(1–2):1–32
Odeh IOA, Todd AJ, Triantafilis J (2003) Spatial prediction of soil particle-size fractions as compositional data. Soil Sci 168:501–515
Palmer MW (1993) Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74:2215–2230
Pawlowsky-Glahn V, Egozcue JJ (2001) Geometric approach to statistical analysis on the simplex. Stoch Env Res Risk A 15(5):384–398
Pawlowsky-Glahn V, Olea RA (2004) Geostatistical analysis of compositional data. Oxford University Press, New York
Pearson K (1897) Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc Royal Soc 60:489–498
Rayner JH (1966) Classification of soils by numerical methods. J Soil Sci 17:79–92
Shaw PJA (2003) Multivariate statistics for the environmental sciences. Hodder Arnold, London

Sinowski W, Scheinost AC, Auerswald K (1997) Regionalization of soil water retention curves in a highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function. Geoderma 78(3):145–159

Tanner JM (1949) Fallacy of per-weight and per-surface area standards, and their relation to spurious correlation. J Appl Physiol 2(1):1–15

Ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179

Ter Braak CJF (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetation 69:69–77

Ter Braak CJF (1995) Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. Chemom Intell Lab Syst 28:165–180

Ter Braak CJF, Prentice IC (1988) A theory of gradient analysis. Adv Ecol Res:271–317

Webster R (1977) Canonical correlation in pedology: how useful? J Soil Sci 28(1):196–221

Woronow A (1990) Methods for quantifying, statistically testing, and graphically displaying shifts in compositional abundances across data suites. Comput Geosci 16(8):1209–1233

Zhang J-T, Oxley ERB (1994) A comparison of three methods of multivariate analysis of upland grasslands in North Wales. J Veg Sci 5:71–76

Part III
Soil Measurements and Properties

"Is complexity randomness or is randomness complexity?"

David van der Linden

Soil is described by observing its properties at suitably located places in the landscape. For any hand specimen of soil material, thousands, if not millions, of biological, chemical and physical soil properties can be observed. This number increases with the march of technology. Most countries have developed standard techniques for field description, sampling and laboratory analyses (e.g. in Australia The National Committee on Soil Terrain 2009; Rayment and Higginson 1992). Field description tends to be qualitative, whereas laboratory analyses are largely quantitative. New technology for making observations and the increasing demand for quantitative soil data are agents of change for soil description.

In this book section, we consider the description and improved quantification of key soil properties. We then look at how soil properties can be inferred from others.

Recalling a point from the introduction to this book, it can be argued that what we deal with in pedometrics are called soil attributes. The reason for this being that we deal with numbers or codes that are attributed to soil, they are not intrinsic properties of the soil themselves. So description implies attribution. This might be explained with reference to pH: there are many ways of trying to quantify the pH of a soil by changing the soil:solute ratio and the nature of the electrolyte – these give varying values – so the pH 1:5 in water and pH 1:2.5 in KCl are different attributes of the soil, but they try to measure an intrinsic soil property, its pH. This might be regarded as too pedantic, so the phrases soil attribute and soil property will be used synonymously.

References

Rayment GE, Higginson FR (1992) Australian laboratory handbook of soil and water chemical methods. Inkata Press, Melbourne

The National Committee on Soil and Terrain (2009) Australian soil and land survey field handbook. CSIRO Publishing, Melbourne

Chapter 5
Pedometric Treatment of Soil Attributes

Uta Stockmann, Edward J. Jones, Inakwu O. A. Odeh, and Alex. B. McBratney

> *"The soil itself must be the object of observation and experiment and the facts obtained must be soil facts before they can be incorporated into soil science. The science of zoology was developed through the study of animals, that of botany through the study of plants, and soil science must be developed through the study of the soil".*
>
> C. F. Marbut 1920

5.1 Introduction

There are some universally described soil attributes that are worthy of more detailed pedometric description. Here, we largely concentrate on field properties which have particular issues associated with them. We are not attempting to be exhaustive.

As outlined in the Introduction of Part III of this book, over the years most countries have performed field descriptions and laboratory analysis of soil based on some kind of standard technique. The methods that can be utilized for in-situ field description were already developed in the 1950s and refined and standardized by most countries in the 1970s and 1980s. These efforts have resulted in what is referred to as soil legacy data. The aim of this chapter is to illustrate how these soil legacy data can be modified for pedometric, quantitative and in general terms more objective soil analysis. Here, we will also discuss how we can use these quantitative descriptions to perform a more quantitative analysis of soil attributes, utilizing mathematical descriptors, or how we can achieve a more quantitative measurement and assessment of soil attributes utilizing new technology and computational advances.

U. Stockmann (✉) • E.J. Jones • I.O.A. Odeh • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia
e-mail: uta.stockmann@sydney.edu.au; edward.jones@sydney.edu.au; inakwu.odeh@sydney.edu.au; alex.mcbratney@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_5

5.2 Soil Colour

Soil colour is perhaps the most commonly observed field attribute. Prior to the 1950s, colour was described in common terms such as red, grey and brown without reference to any standards. From then onwards, colour designation has been achieved through standardized colour systems in which an individual colour is represented by a point in a three-dimensional space (Melville and Atkinson 1985). This move from a solely subjective to a more objective description of soil colour was driven by the intention of being able to compare and classify soils.

Colour has been largely described by the three-dimensional Munsell colour system by comparison with chips in the Munsell soil colour book (Munsell Color 2009). The colour designated can be interpolated between the chips and pages of the book. It consists of three axes: the hue (H), value (V) and chroma (C) which describe a cylindrical coordinate system (Fig. 5.1a). The hue characterizes the similarity to a dominant colour (red, yellow and blue) or a combination of any two, the value describes the colour intensity or lightness and the chroma refers to the saturation in colour or relative purity of the dominant wavelength. The Munsell colour chart is the

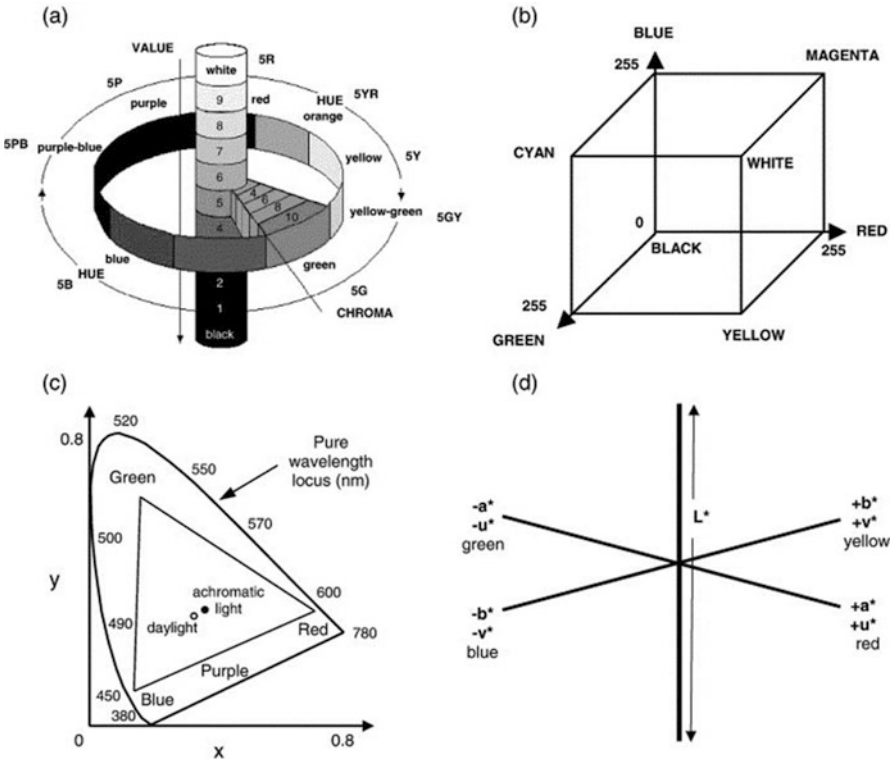


Fig. 5.1 (**a**) The Munsell colour model represented by a cylindrical coordinate system. (**b**) The RGB model, (**c**) the CIE xy chromaticity diagram (**d**) the CIELu*v* and CIELa*b* colour space model (Sourced from Viscarra Rossel et al. (2006a), used with permission)

standard reference for pedologists for describing soil colour in the field. However, a study conducted by Sánchez-Marañón et al. (2005) found that Munsell colour books are subject to variation in themselves due to manufacture differences or effects of colour fading due to prolonged use and that variation is also introduced by subjective colour perception of the individuals describing soil colour in the field. In addition, such a cylindrical coordinate system, sometimes, causes difficulties when it is used for statistical analysis as the Munsell colour space is presented as non-contiguous slices representing each page of the colour book. For quantitative data analysis, further transformations of the Munsell colour notations are therefore required. For such quantitative, numerical or predictive analysis, a three-dimensional Cartesian coordinate system is preferred (Melville and Atkinson 1985; Viscarra Rossel et al. 2006a). Examples of Cartesian colour coordinate systems are CIELab, CIELuv and RGB systems (Fig. 5.1b–d). All have slightly different properties but generally are similar. The Commission Internationale de l'Eclairage (CIE 1931) standardized colour space systems by specifying the light source, the observer and the methodology used to derive the values for describing soil colour. The CIE colour system defines how to map a spectral power distribution to a set of tristimulus coordinates which describe a XYZ coordinate colour space, with Y representing the luminance or brightness component and X and Z representing virtual or 'imaginary' components of the primary spectra (Wyszecki and Stiles 1982). These XYZ tristimulus coordinates can then be transformed to *xy* coordinates of chromaticity: $x = \frac{X}{(X+Y+Z)}$ and $y = \frac{Y}{(X+Y+Z)}$ where *xy* values lie between 0 and 1 (Melville and Atkinson 1985).

Following on, the CIE proposed additional transformation systems to overcome the perceptual non-linearity of the XYZ system, the CIELuv (L*, u*, v*) and CIELab (L*, a*, b*) colour spaces (refer to Fig. 5.1), which represent colour on a continuous numeric colour space that is more suitable for pedometric analysis (Viscarra Rossel et al. 2006a). In these opponent-type colour spaces, u* and a* represent opponent red/green scales (with +u* and +a* representing reds and –u* and –a* representing greens), and similarly v* and b* are opponent yellow/blue scales. The value L* is used in both systems and is representative of the 'metric lightness function' which ranges from 0 (black) to 100 (white) (Melville and Atkinson 1985). For more details on these colour space systems and the derivation of their coordinates, please refer to the books by Wyszecki and Stiles (1982) and Billmeyer and Saltzman (1981).

Another advantage of Cartesian systems is the ability to calculate a colour contrast, for example, for measuring the difference in colour between the A and B horizons or the contrast between soil mottles and the background matrix (Gobin et al. 2000). A colour difference or contrast is simply the Euclidian distance between the two colours (Melville and Atkinson 1985). If p and q are two points in the Euclidean n-space, then the distance (d) between them can be calculated using the Pythagorean formula:

$$d = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \tag{5.1}$$

For example, the overall colour difference between two points in the CIELab colour space (refer to Fig. 5.1) can be calculated using this Euclidean relationship (Melville and Atkinson 1985):

$$\Delta E^*_{ab} = \left[\left(\Delta L^*\right)^2 + \left(\Delta a^*\right)^2 + \left(\Delta b^*\right)^2 \right]^{\frac{1}{2}} \tag{5.2}$$

Viscarra Rossel et al. (2006a) present algorithms (*ColoSol*) for the conversion between colour space models. Recently, however, commercial and open-source coding software has become available that includes packages for transforming colours between different colour spaces. MATLAB, for example, provides the 'colour space transformation' package (Getreuer 2011), whereas the open-source software R provides the 'aqp' package which compiles algorithms for quantitative pedology (Beaudette and Roudier 2016). The latter includes the 'colour space' package which provides colour space manipulations including the mapping between various colour space models including RGB, HSV, HLS, CIE xy, CIELuv, HCL (polar CIELuv), CIELab and polar CIELab (Ihaka et al. 2015). In addition, the R package 'Munsell' (Wickham 2016) provides conversion between the RGB colour space and the Munsell HVC system and also converts a Munsell colour to a hue, chroma and value triplet. Lately, soil colour can also be determined using a (mobile) smartphone as these devices have all the requirements now to capture and process digital images (Gómez-Robledo et al. 2013). Android- or apple-specific application software has been developed that allows transformations between colour spaces on the go.

Soil colour has been used to determine types of soil, and a number of soil properties can also be related to soil colour. In general, dark soil colours, especially of surface soil, have been associated with high organic carbon content and thus good soil fertility, whereas light, pale or bleached colours are associated with a loss of base cations and nutrients and in general poor fertility. Light soil colours are also related to the presence of carbonates and calcium sulphates (gypsum), whereas the yellowish and reddish colour of soil can be related to iron oxides such as goethite and hematite, respectively. Over the years, soil colour has therefore been used to infer other soil constituents. Some of these studies used soil colour measurements to first calculate colour indices and then employed those to quantitatively describe and assess their occurrence (see Table 5.1).

More explicitly, soil colour was successfully used to quantify iron content and in particular hematite and goethite contents in soils (Madeira et al. 1997). Torrent

Table 5.1 Colour indices used to qualitatively describe soil constituents

Index name	Associated soil constituent/use	Formula	Delineation	References
Redness index (RI)	Iron oxide (hematite content)	$\mathrm{RI} = (15 - H)^* \frac{C}{V}$	Conversion of reflectance measurements (400–700 nm) to CIE xyz and Munsell hue (H), value (V), chroma (C)	Torrent et al. (1983)
Redness index (R)	Iron oxide (hematite content) Soil weathering	$R_{\mathrm{CIE}} = \frac{(x-0.34)^2 10^4}{(y-0.34)z^2}$ $R_{\mathrm{MUN}} = \frac{(10-H)^3 \mathrm{C}\ 10^3}{V^6}$	Conversion of reflectance measurements (400–700 nm) to CIE xyz and Munsell hue (H), value (V), chroma (C), 0.34 corresponds to x, y coordinates of an average iron-extracted soil	Barron and Torrent (1986, page 507)
Colour index	Wetness Aeration	C1 C2	C1 based on Munsell chroma C2 based on Munsell hue and chroma	Evans and Franzmeier (1988)
Colour index	Wetness Soil saturation	$\mathrm{RR} = \frac{(45-H)C}{V}$ $C1 = \left(\sum_{i=1}^{m} X_{\mathrm{matrix},i} C_{\mathrm{matrix},i}\right) + \left(\sum_{j=1}^{n} X_{\mathrm{mottle},j} C_{\mathrm{mottle},j}\right)$ $C2 = \left[\sum_{i=1}^{m} X_{\mathrm{matrix},i} \left(45 - H_{\mathrm{matrix},i} + C_{\mathrm{matrix},i}\right)\right] + \left[\sum_{j=1}^{n} X_{\mathrm{mottle},j} \left(45 - H_{\mathrm{mottle,j}} + C_{\mathrm{mottle},j}\right)\right]$	Munsell hue (H), value (V), chroma (C), X is the abundance (fraction), m is the total number of matrix colours, n is the total number of mottle colours	Thompson and Bell (1996)

(continued)

Table 5.1 (continued)

Index name	Associated soil constituent/use	Formula	Delineation	References
Colour index Profile darkness index (PDI) Hydromorphic index (HI)	Soil organic carbon	$\mathrm{PDI} = \sum_{i=1}^{n} \frac{T}{(V \times C_h)+1}$ $\mathrm{HI} = \frac{P}{C_h \times V}$	PDI function of A horizon thickness, n is the number of A horizons, C_h is the Munsell chroma, V is the Munsell value, T is A horizon thicknesses, P (%) is the proportion of total soil profile depth with hydromorphic features	Chaplot et al. (2001)
Redness index (RI) Colour index (CI)	Iron content Soil organic carbon Soil drainage	$P_i = P_i\left(1 - \sum_{i=1}^{m} S_j\right)$ $\mathrm{RI} = \frac{\mathrm{MaRI}}{\sum_{i=1}^{n} P_i\left[\frac{\mathrm{HT}_i C_i}{V_i}\right]} + \frac{\mathrm{MoRI}}{\sum_{i=1}^{n} S_j\left[\frac{\mathrm{HT}_i C_i}{V_i}\right]}$ $\mathrm{CI} = \frac{\mathrm{MaCI}}{\sum_{i=1}^{n} P_i[\mathrm{HT}_i + C_i + V_i]} + \frac{\mathrm{MoCI}}{\sum_{i=1}^{n} S_i[\mathrm{HT}_i + C_i + V_j]}$	P_i is the percentage of each estimated matrix colour, m is the number of mottle colours, S_j is the fraction of each mottle colour, n is the number of matrix colours, V is the Munsell value, C is the Munsell chroma	Gobin et al. (2000)

et al. (1983) and later Barron and Torrent (1986) found a strong relationship between hematite content and colour indices (rating of the degree of redness) that were calculated employing the Munsell notation or CIE chromatic coordinates. Torrent et al. (1983), for example, established a quantitative relationship between soil colour (i.e. redness ratings sourced from visual and spectrophotometric colour measurements) and hematite content. The authors found that the pigmenting power of hematite differed between the geographic regions studied and, for predictive purposes, recommended region-unique calibration models. In addition, soil colour has also been used to infer soil water content. Bedidi et al. (1992) found that hue increased with moisture content and colour purity (saturation) decreased with increasing moisture content. Furthermore, other soil hydrological properties have also been related to soil colour. He et al. (2003), for example, estimated the frequency and duration of soil saturation based on soil colour. They determined the percentage of redoximorphic features at a given soil depth and found that low chroma colours increased the longer the soil studied was saturated and chemically reduced. A difference in colour indices was used by Evans and Franzmeier (1988) to estimate wetness and soil aeration, by Thompson and Bell (1996) to classify seasonal soil saturation and by Gobin et al. (2000), for example, to infer soil drainage. Soil colour has also been used to predict soil (organic) carbon content. Ivey and McBride (1999) related moist soil colour to topsoil organic carbon contents, with the chromaticity coordinate a* (CIELab colour space notation) showing good estimates for medium- to fine-textured soils. Gobin et al. (2000) and Chaplot et al. (2001) used colour indices to quantify differences in soil organic carbon content. Viscarra Rossel et al. (2006a) predicted soil organic carbon content using various colour space models and concluded that the CIELuv and CIELCh models were the most suitable. Following on, Viscarra Rossel et al. (2008) used the CIE colour system as a proxy to measure soil organic carbon and iron contents. More recently, Baumann et al. (2016) used soil colour (CIELab system) to rapidly assess soil organic matter content and found that soil lightness (L* value, refer to Fig. 5.1) was significantly negatively correlated with soil organic matter content. The authors also found that soil colour (L* , a* and b* values) was affected by geographic region and land use, with in general higher L* values for forest soils when compared to grassland soils. In an earlier publication, Spielvogel et al. (2004) established a relationship between soil lightness and soil organic carbon content and the chemical composition of soil organic matter and found that soil lightness decreased with increasing organic carbon content. More recently, Sánchez-Marañón et al. (2011) also used colour to assess aggregate stability, including the effects of fabric on this relationship. An increase in stability was related to increased soil redness, darkness and chromaticity due to an increase in binding agents such as clay, organic C and free forms coating the surfaces of the aggregates successfully used to quantify iron content

Technologies that provide a quantitative measure rather than only a visual matching of colour are available. Soil colour measurements have been made using digital cameras where RGB colour values can be directly estimated from the images made or chromameters, for example (Roudier et al. 2016). Lately, these have also

benefited immensely from advances in hand-held *spectroradiometry* devices (*spectrophotometry technology*). The visible spectrum of the electromagnetic spectrum (refer to Fig. 5.11 in part 5.4) which lies between 380 nm (violet) and 780 nm (red) can be used to calculate soil colour from reflectance spectra (Fernandez and Schulze 1987). Shields et al. (1966) first investigated the capability of measuring soil colour via spectroreflectance. However, this approach only received increased popularity in the 2000s. For example, Barrett (2002) used in-field reflectance spectra from the soil pit face to measure soil colour and found strong correlations between those spectroradiometer measurements and visual estimates of Munsell colours. Viscarra Rossel et al. (2009) estimated soil colour from the visible part of the vis-NIR spectra and also investigated its use for in-field application. The continuum removal technique was used to select characteristic absorption features corresponding to blue (450–520 nm), green (520–600 nm) and red (600–690 nm). The selected RGB values were multiplied by 255 to arrive at an 8 bit pixel colour encoding value that was transformed to the Munsell HVC and CIELab colour models. Comparison of vis-NIR measurements to visual Munsell book estimates was promising, but it was found that those had the tendency to be darker and more yellow in hue. Figure 5.2 exemplifies the use of vis-NIR spectrometry to predict colour of a Red Chromosol, Brown Sodosol and Black Vertosol (Australian Soil Classification) using the R 'Munsell' package (Wickham 2016) and the technique described in Viscarra Rossel et al. (2009). Furthermore, Table 5.2 shows an example of converting the B horizon soil colour values from the RGB to Munsell and CIELab codes using the R colour space package (Ihaka et al. 2015).

As discussed, proximal soil sensors capable of readily taking quantitative soil colour measurements in the field (e.g. new ASD configuration, spectral evolution, in-field vis-NIR profiler) are now available. Such measurements are affected by soil moisture variation. However, correcting algorithms such as external parameter orthogonalization (EPO) are available that can be used to eliminate the moisture effect on scanning soil samples directly in the field as soil moisture generally increases soil darkness (Minasny et al. 2011). This will be discussed further in Sect. 5.5 of this chapter.

Soil colour estimates have not only been made using proximal soil sensors but also by employing remote sensing satellite data as there is also a significant correlation of chromaticity coordinates with reflectance values in the visible bands of remote sensing data (Escadafal et al. 1989). Over the years this relationship has been applied to study the spatial variation of the surface soil colour in remote sensing applications. Different soil types can be mapped based on variations in surface brightness and colour saturation, and soil processes that affect the surface soil may also be inferred using colour differences. For example, topsoil removal can be identified as it results in a colour change because the exposed subsurface soil will have a different colour than the non-eroded soil (Escadafal 1993).

How to use remote sensing data for the measurement of a range of soil attributes will be explained in more detail in the applications of pedometrics on *vis-NIR-SWIR remote sensing products as new soil data for digital soil mapping* (Chap. 13). Here, we only provide a few examples that are related to soil colour inference. Escadafal

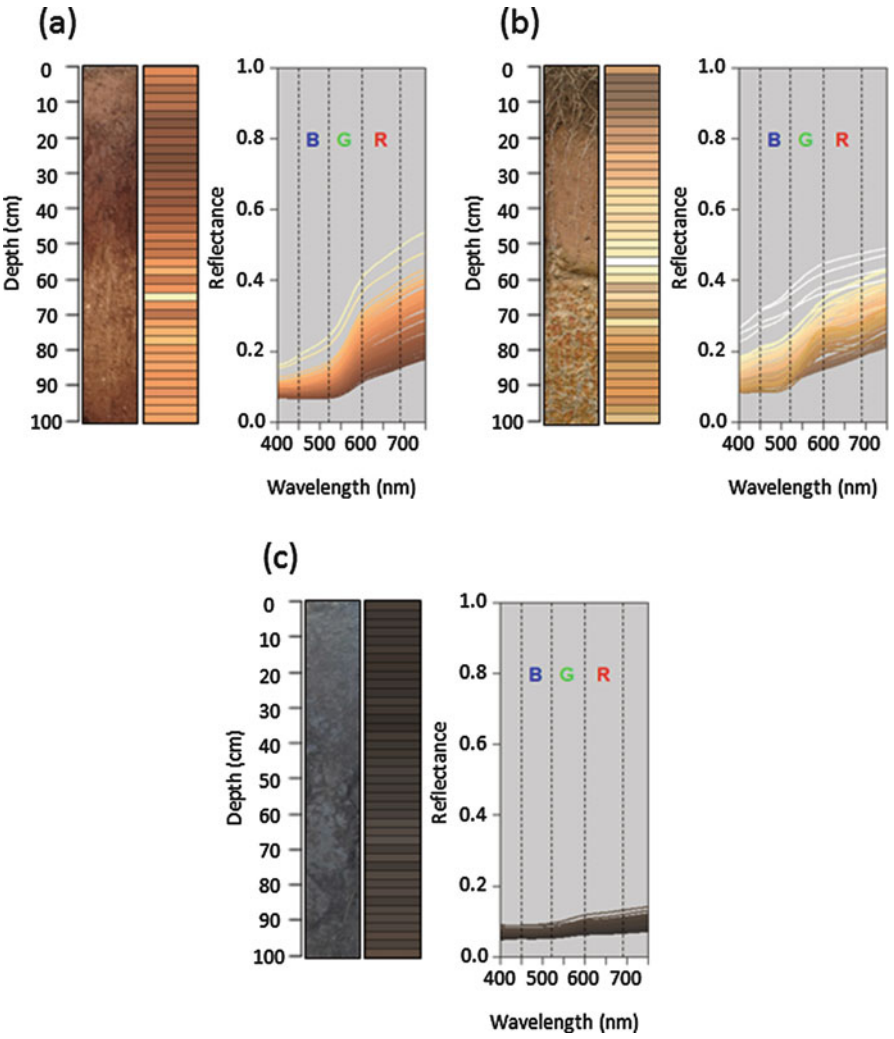


Fig. 5.2 Use of a vis-NIR spectrometer to predict colour of (**a**) Red Chromosol, (**b**) Brown Sodosol and (**c**) Black Vertosol in situ. Each plot contains a digital photograph of the profile, vis-NIR-derived colours in 2.5 cm increments to a depth of 1 m and vis-NIR reflectance plot indicating wavelength bands of *red* (R = 600–690 nm), *green* (G = 520–600 nm) and *blue* (B = 450–520 nm). Colours were predicted in R using the 'Munsell' package (Wickham 2016) after extracting the mean reflectance within each band and scaling appropriately

Table 5.2 Conversion of modal B horizon values (Fig. 5.2) from RGB to Munsell and CIELab through the use of the colour space package (Ihaka et al. 2015)

	Red Chromosol	Brown Sodosol	Black Vertosol
RGB	0.655, 0.402, 0.281	0.934, 0.768, 0.526	0.279, 0.250, 0.229
Munsell	2.5 YR 5/6	10 YR 8/6	N 3/0
CIELab	49.8 22.7 28.2	81.5 7.1 36.6	27.6 2.1 4.4

and Huete (1991) used a redness index from red and green spectral bands to correct vegetation indices (NDVI and SAVI) for 'soil noise' apparent in remote sensing images of low vegetation cover in arid regions. Later, Leone and Escadafal (2001) used the visible bands of the multispectral infrared and visible imaging spectrometer (MIVIS) hyperspectral sensor and the Landsat thematic mapper (TM) to infer soil surface colour. Madeira et al. (1997) performed soil mapping of lateritic soils employing visible spectrometric indices of hematite and goethite content (Landsat TM). Mathieu et al. (1998) studied the relationship between radiometric indices calculated from the visible bands (blue, green and red) of remote sensing data and soil colour (Munsell and Helmholtz data). They found that radiometric indices were good predictors of the soil colour components. Following on, Mathieu et al. (2007) applied radiometric indices (redness index and brightness index) from SPOT imagery for soil erosion mapping in a Mediterranean environment.

5.3 Soil Texture and Particle Size

Particle-size distributions are compositions of solid particles, the building blocks of soils. We commonly describe soils in terms of the mass proportion of sand-, silt- and clay-sized particles, which can be plotted on simplexes or texture triangles. Particle-size distribution is usually measured in the laboratory using mechanical analysis (Gee and Or 2002). In the field, particle size can be estimated using the 'hand feel' method. It is questionable whether texture estimated by hand and in the laboratory measures the same intrinsic soil property. There is little doubt, however, that the two can be calibrated with each other. Hodgson et al. (1976), for example, compared field texture estimates to those derived by the pipette method of particle-size analysis and found that surveyors were able to confidently estimate the particle-size distribution of the soils, although there was a tendency throughout for underestimation of clay content.

Particle-size distribution is useful in pedology to study the development of soil through examining the changes in clay fractions with soil age (Walker and Chittleborough 1986). The distribution of sand, silt and clay in a soil profile can also be used to elucidate soil-forming processes (Legros and Pedro 1985). Particle-size distribution is also a key predictor in pedotransfer functions, especially for soil hydraulic properties. It is surprising then that there is no universal agreement on the size ranges of particle-size distributions. Each country adopted a different system

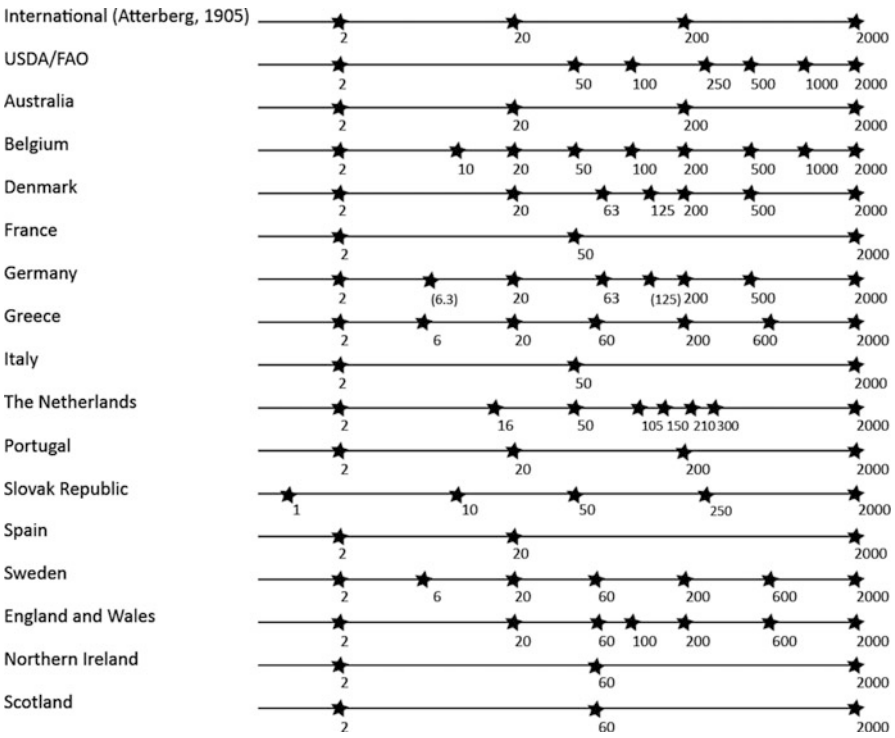


Fig. 5.3 Particle-size limits used in the international and USDA/FAO system and used by Australian and various European countries (Adapted from Minasny and McBratney 2001)

to represent the particle sizes; this resulted in various size ranges used as fraction boundaries (Fig. 5.3).

In addition to the country-specific particle-size systems, there are two systems that are used as international standards, the International system proposed by the Swedish scientist Atterberg (1905) which was endorsed by the IUSS during the first International Congress of Soil Science held in Washington, D.C., 13–22 June, 1927 (Minasny and McBratney 2001), and the USDA system which was adopted by the FAO and whose current particle-size limits were established in 1938 (Knight 1938). It is interesting to note here that not many countries adopted the so-called International system (refer to Fig. 5.3 which shows the international and USDA/FAO system in comparison to Australia and various European countries).

Differing particle-size classification systems can cause particular problems when data from different countries are merged into a single database or when texture estimates are required in continuous pedotransfer functions (refer to Chap. 7), for example, the soil database for Europe (*ESDB* – European Soil Database, European Commission, Joint Research Centre). In such cases empirical equations are required to relate the fractions in one system to the fractions in another. Lack of standardization also creates difficulties when comparing textural classes between

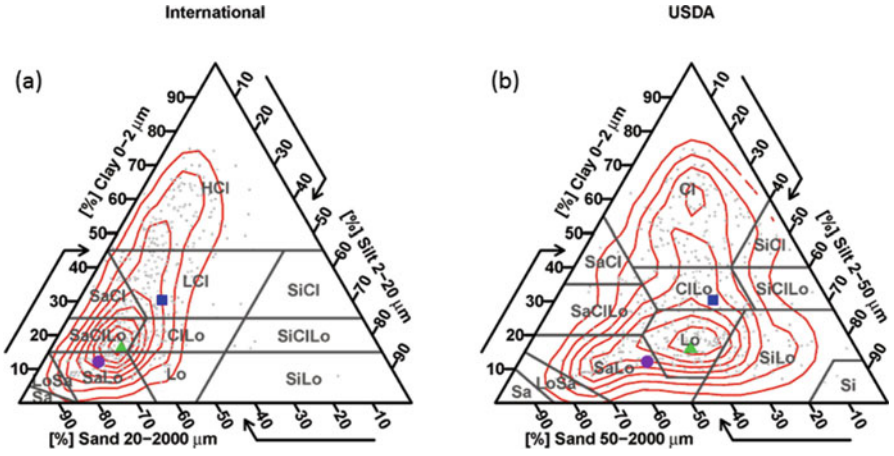


Fig. 5.4 Texture triangles with density contours (*red*) displaying 490 samples (*grey dots*) from 69 Australian soil profiles (Geeves et al. 1995) plotted under the (**a**) international and (**b**) USDA systems. The silt and sand fraction boundary was increased from 20 to 50 µm via a pedotransfer function (Padarian et al. 2012). Texture triangles were created using the soil texture wizard package in R (Moeys 2015). Selected points are indicated with coloured shapes

systems. This incongruity can be demonstrated by plotting texture data under both, the International and USDA systems (Fig. 5.4).

In Fig. 5.4 we can see that increasing the silt-sand fraction boundary from 20 µm (International system) to 50 µm (USDA system) shifts samples to the right of the texture triangle along the fixed clay plane. This results in a more uniform dispersion of points within the texture triangle. However, only 41% of the samples, 202 of 490, remain in their corresponding textural class following conversion to the USDA system. For many texture classes, differing clay content requirements mean that they will never be assigned to the corresponding class. The USDA system generally has a higher clay requirement. For example, the clay loam (ClLo) class requires 27–40% clay in the USDA system versus 15–25% clay in the international system. Clearly these requirements are incompatible, and there are dangers in assuming texture classes in the differing systems correspond to similar soil samples. These issues highlight the need for unifying techniques to facilitate the sharing of data and techniques between countries.

Shirazi and Boersma (1984) proposed a bell-shaped texture triangle representing the mean and variance of an assumed lognormal distribution (Fig. 5.5). The model was later refined using piecewise lognormal curves for each size fraction (Shirazi et al. 1988). This approach is useful for unifying textural classification of differing particle-size systems as it integrates some quantitative statistical information, i.e. any combination of sand, silt and clay can be represented by a geometric mean particle diameter and a standard deviation.

Shirazi et al. (1988) also established relationships between various texture classification schemes adopted by the USDA, the International Union of Soil Science and the American Society of Civil Engineers, based on geometric mean and

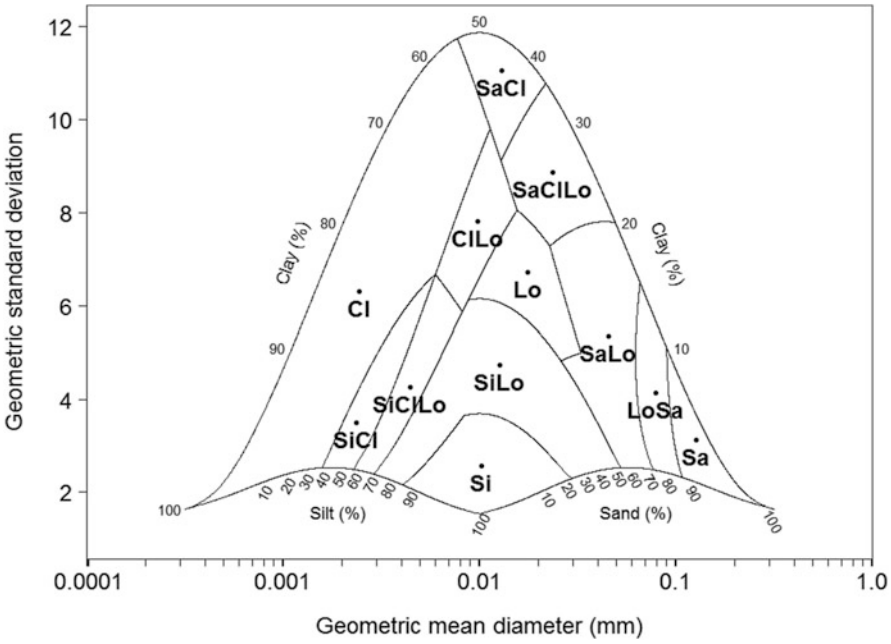


Fig. 5.5 The USDA texture triangle represented as geometric mean diameter and standard deviation (Adapted from Shirazi and Boersma 1984)

standard deviation of the clay, silt and sand fraction size ranges (lognormal particle-size distribution). However, the approach has received criticism for its unrealistic discontinuities at fraction boundaries (Buchan 1989).

The representation of particle-size distribution by a small number of fractions is an oversimplification of reality. Continuous measurements may be made using laser diffraction or Coulter counters, or we may simply isolate a larger number of fractions. This will undoubtedly give us greater insight. Walker and Chittleborough (1986) isolated 18 particle sizes from the fine fraction of Alfisols in south-eastern Australia. The large number of fractions isolated revealed unimodal distributions in A horizons and bimodal particle-size distributions in B horizons (Fig. 5.6). The level of detail allowed the authors to conclude that development of Bt horizons progresses from an initial translocation of the clay followed by intensive weathering and size reduction of clay particles.

Ideally, soil texture should be described by a large number of particle-size fractions and statistical distributions fitted to model cumulative particle-size distributions. Such a mathematical representation of particle size is particularly important for converting between systems. For example, particle-size standardization in accordance with the USDA/FAO system for the European database of hydrologic properties of European soils (HYPRES) required the estimation of data at 50 μm

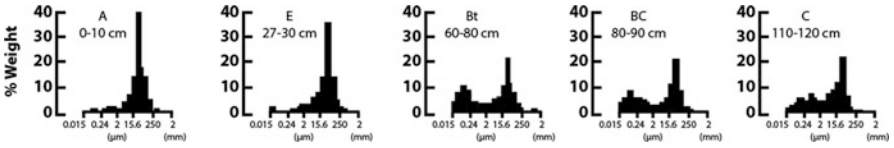


Fig. 5.6 Particle-size histograms for key soil horizons in an Alfisol profile, showing the distinct change in particle-size distributions down the soil profile from the A to the B horizons (Redrawn after Walker and Chittleborough 1986)

from data where soil texture fractions were measured at limits of 60, 63, 200 and 2000 μm above the 20 μm limit (Nemes et al. 1999).

No general statistical distribution has been found to fit all soil particle-size distributions, as they can be of symmetric or nonsymmetric nature as exemplified in Chap. 3. It is also clear that in some symmetric cases, double or even triple lognormal distributions have to be fitted. The amount of model parameters required to describe a cumulative particle-size distribution depends on its complexity; the more complex, the greater the number of model parameters needed.

There are a number of (comparative) studies which addressed this particular subject (Table 5.3). Berezin and Voronin (1981), for example, used probability distribution functions for describing the particle-size distribution of Russian soils where the total weight percentage of particles of a certain diameter was taken as a characteristic of the probability of occurrence of particles of this diameter in the soil. Barndorff-Nielsen (1977) proposed the log-hyperbolic distribution to describe textural characteristics, based on findings of Bagnold (1954) who demonstrated its applicability to describe aeolian sand deposits in his research on sand dune formation. This probability distribution was used to distinguish between dune sides, dune crests and interdunal corridor (Wyrwoll and Smyth 1985) and in a later study to discriminate between coastal sub-environments (Sutherland and Lee 1994).

Buchan et al. (1993) on the other hand compared five lognormal models of particle-size distributions (i.e. Jaky (1944) one-parameter model, simple lognormal model, two adjusted lognormal models, offset-renormalized lognormal model, offset-nonrenormalized lognormal model, bimodal lognormal model) and found that all models accounted for over 90% of the variance in particle-size distribution of most of the New Zealand soils studied. In a different comparative study, Rousseva (1997) found that model suitability was influenced by texture. Closed-form exponential functions represented well the cumulative particle-size distributions of fine-textured soils, whereas closed-form power functions better represented the cumulative particle-size distributions of coarse-textured soils. Following on, Hwang et al. (2002) compared symmetric and asymmetric models for estimating soil particle-size distributions of a range of Korean soils (i.e. five lognormal models, the Gompertz model, the Fredlund model). They identified an effect of texture on model performance, with increased performance of some models with increasing clay content. This dependency of model performance on texture was also found

Table 5.3 Examples of models used to describe and interpolate cumulative particle-size distributions (Revised and updated after Hwang et al. 2002)

Source	Model[a]	Parameters[b]
Jaky one-parameter model (Jaky 1944)	$F(d) = \exp\left\{-\frac{1}{p^2}\left[\ln\left(\frac{d}{d_m}\right)\right]^2\right\}$	p = Particle size distribution (PSD) index d_m = maximum diameter
Simple lognormal (SL) [c](Buchan 1989)	$F(X) = \frac{1}{2}\left(1 + \frac{\mathrm{erf}[X-\mu]}{\sqrt{2}\sigma}\right) \quad d \geq \mu$ $F(X) = \frac{1}{2}\left(1 - \frac{\mathrm{erf}[X-\mu]}{\sqrt{2}\sigma}\right) \quad d < \mu$	$X = \ln(d)$ μ = geometric mean σ = geometric standard deviation
Offset-renormalized lognormal[d] (Buchan et al. 1993)	$G(X) = (1-\varepsilon)F(X) + \varepsilon$	ε = offset or displacement parameter
Offset-nonrenormalized lognormal[d] (Buchan et al. 1993)	$G(X) = F(X) + c$	c = offset or displacement parameter
Double lognormal[d] (Shiozawa and Campbell 1991)	$G(X) = \delta F_1(X) + (1-\delta)F_2(X)$	μ_1, μ_2 = geometric mean of clay and combined sand and silt fraction σ_1, σ_2 = geometric standard deviation clay and combined sand and silt fraction δ = mass fraction of clay
Gompertz (Nemes et al. 1999)	$F(d) = \alpha + \gamma \exp\{-\exp[-\beta(d-\mu)]\}$	$\alpha, \beta, \gamma, \mu$ = shape parameters
Spline interpolation (Nemes et al. 1999)	Given fraction boundaries: $d_0 < d_1 < \cdots < d_i < \cdots < d_m$ Spline is formed as: $S(d) = \begin{cases} F_1(d), & d_0 \leq d \leq d_1 \\ \vdots & \\ F_i(d), & d_{i-1} < d \leq d_i \\ \vdots & \\ F_n(d), & d_{n-1} < d \leq d_n \end{cases}$	d_i = fraction boundaries and knot sites

(continued)

Table 5.3 (continued)

Source	Model[a]	Parameters[b]
	Given: $F_i(d) = b_{0,i} + b_{1,i}d + b_{2,i}d^2 + \ \ldots\ + b_{n,i}d^n$	d_0 = minimum diameter d_m = maximum diameter For internal knots: $\begin{cases} F'_{i-1}(d_i) = F'_i(d_i) \\ F''_{i-1}(d_i) = F''_i(d_i) \end{cases}$ n = degree of polynomial
Fredlund (Fredlund et al. 2000)	$F(d) = \frac{1}{\left\{\ln\left[\exp(1)+\left(\frac{a}{d}\right)^n\right]\right\}^m}\left\{1 - \left[\frac{\ln\left(1+\frac{d_t}{d}\right)}{\ln\left(1+\frac{d_t}{d_0}\right)}\right]^7\right\}$	α, n, m, d_t = shape parameters d_0 = minimum diameter
Skaggs model[e] (Skaggs et al. 2001)	$F(d) = \frac{2}{\left\{1 + \left(\frac{1}{P(d_0)} - 1\right)\exp\left[-u\left(\frac{d-d_0}{d_0}\right)^c\right]\right\}}$	u, c = shape parameters d_0 = minimum diameter $P(d_0)$ = mass fraction below minimum diameter

[a] d, particle diameter in mm
[b] Parameters determining the shape of the curve
[c] erf[], error function
[d] $F(X)$ is defined by the SL model
[e] Model unsuitable for soils with a silt fraction greater than 70%

in a later study that compared a total of nine soil particle-size distribution models (Hwang 2004).

5.4 Soil Structure

Soil structure or pedality is often described in the field, but even with new technological developments, it is still a difficult property to readily quantify. For example, in 1941 Nikiforoff (1941) established a morphological 'standardized' classification of soil structure. In the field soil structure is often described in somewhat vague grades of 'massive', 'weak', 'moderate' or 'strong' (Terrain 2009). These structure grades, however, can be converted into an ordinal system, e.g. '0', '1', '2' and '3', so that they can be used in a more quantitative analysis.

Normally, the soil's structure in relation to the shape of peds present in the soil is also described with respect to a number of classes, e.g. 'blocky', 'prismatic', 'angular', 'subangular', 'columnar', etc. which in essence describe the three-dimensional shape of an individual ped and how the peds relate to each other.

One way of quantifying the pedality of a soil is to consider the individual descriptors or the ped shape as separate variables. When this is done, four variables emerge, i.e. horizontality, verticality, roundness and accommodation. The advantage of describing these four variables as opposed to a single class is that differences between the structures of volumes of soil can be more subtly detected or measured. Table 5.4 shows an example of this method. Here, fuzzy coding was used to describe the structural types of ped shapes. The fuzzy codes show the degree of similarity with respect to the ped face characteristics and account for uncertainty or a lack of a clear boundary between descriptors (Odeh et al. 1991).

Another approach is to characterize soil structure via aggregate shape and surface roughness. Holden (1993), for example, used three indicators (1) circularity or sphericity, (2) angularity or roundness and (3) surface roughness to quantify soil ped shape. Cox's R-statistic was used to quantify circularity or gross shape (R) with

Table 5.4 Fuzzy coding of structural type

Type	Horizontality	Verticality	Flatness	Accommodation
Platy	1.0	0.1	1.0	1.0
Lenticular	1.0	0.3	0.3	1.0
Prismatic	0.2	1.0	1.0	1.0
Columnar	0.2	1.0	0.9	0.9
Angular blocky	1.0	1.0	1.0	1.0
Subangular blocky	0.7	0.7	0.5	0.5
Granular	0.2	0.2	0.0	0.1
Massive	0.0	0.0	0.0	1.0
Single grain	0.0	0.0	0.0	0.0

After Odeh et al. (1991)

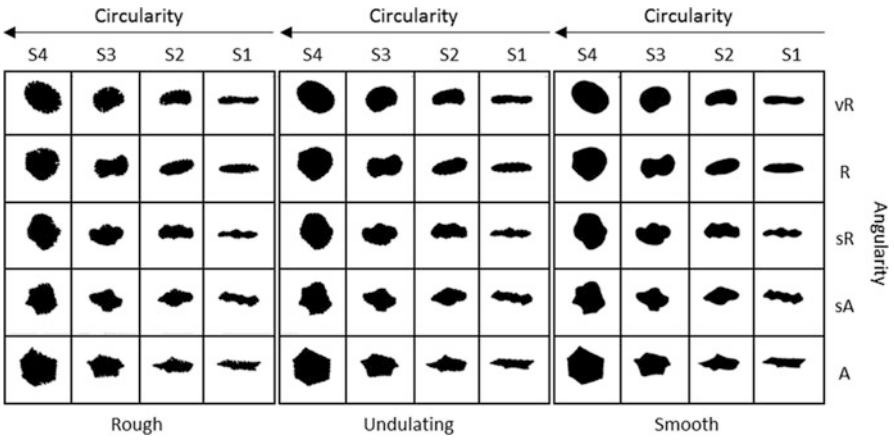


Fig. 5.7 A modified version of Pettijohn's (1957) standard shape chart. *Main boxes* indicate surface roughness classes, *columns* indicate circularity classes, and *rows* indicate angularity classes (Redrawn after Holden 1993)

$$R = \frac{4\pi A}{P^2} \tag{5.3}$$

with A defining the projected area of the particle and P the perimeter of the particle. This measure gives the ratio of the projected area of the particle to the area of a circle that has the same perimeter, and therefore $R = 1$ for a perfect circle and $R = 0$ for a line. The Steinhaus paradox was employed to quantify surface roughness. Accordingly, the length of a perimeter, L, is estimated by the number of steps, F, it takes to measure out the perimeter and their individual length ε:

$$L(\varepsilon) \sim F\varepsilon^{1-D} \tag{5.4}$$

where D is the fractal dimension, i.e. the degree to which a one dimension extends into a two dimension such as a line and its potential to becoming a filled area. Angularity was described according to two classes, angular (0) or rounded (1). Image analysis was used to calculate these measures, and results were compared to six criteria (Clark 1981) that can be used to assess shape measurement (i.e. uniqueness, parsimony, independence, rotation invariance, scale invariance, reflection invariance) and calibrated against Pettijohn's (1957) standard shape chart (Fig. 5.7). This study found that Cox's R-statistic is a powerful measure of circularity and gross shape.

The mathematical technique of fractal geometry which was introduced by Mandelbrot (1989) has also been used to describe soil structure qualitatively, as fractal geometry has the ability to describe rugged or irregular geometries by estimating their fractal dimension (McBratney et al. 2002). In addition, fractal soil models are also capable of relating structure to soil physical processes. Essentially,

fractal objects have no characteristic scale dimension, as it is assumed that their morphological appearance stays the same. Mandelbrot (1989) describes a fractal as 'shapes whose roughness and fragmentation neither tend to vanish, nor fluctuate up and down, but remain essentially unchanged as one zooms in continually and examination is refined'.

McBratney et al. (2002) provide an extensive review about the theory behind fractal geometry and how it can be applied to directly quantify soil structure from soil images, and we will only provide a short summary here. As outlined in this review, fractal objects can be described through fractal dimensions in various ways. One option is to use the *mass fractal dimension*, D_m, which is less than the Euclidean dimension that a fractal is part of (the embedding dimension, d_e), with d_e equal to 2 in a two-dimensional space (e.g. an image of soil from a thin section) and equal to 3 in a three-dimensional space (e.g. a soil aggregate). The mass, M, of a mass fractal inside a radius, r, is scaled according to

$$M \propto r^{D_m} \tag{5.5}$$

Furthermore, the *spectral dimension*, d, can be used as a measure of the connectivity of fractal pathways, with a large value of d equal to a more continuous less winding pathway. The *surface fractal dimension*, D_s, is used to describe the irregularity or ruggedness of a perimeter or surface. For example, in the case of a perimeter, a straight line is non-fractal and has an Euclidean dimension equal to 1, whereas an irregular line that is subject to fractal scaling has a value of D_s that reaches 2 depending on its ruggedness (space filling). Similarly, a flat surface has an Euclidean dimension equal to 2, whereas a fractal surface has a value of D_s that is approaching a value of 3 depending on how volume filling the surface becomes. The *fragmentation fractal dimension*, D_f, is also used which is estimated from a particle-size or aggregate-size distribution and mathematically allows a value range of $0 < D_f > 3$ (refer to Anderson and McBratney (1995) for detailed discussions). This measure of size distribution has been employed, for example, as an indicator of soil fragmentation caused by different agricultural management practices (Anderson and McBratney 1995).

Efforts in directly measuring and quantifying structure are focused on image analysis of two-dimensional images of soil and more recently also on three-dimensional images owing to technological improvements (e.g. Eck et al. 2013). Common practice is to convert the (black and white) digitized images into a binary and to assign each pixel either to the solid soil matrix or the pore space, using, for example, the box-counting method, where each binary soil image is overlain by a grid of square boxes (McBratney et al. 2002). Mathematical techniques such as the discussed fractal geometry are then employed to estimate quantitative measures relevant for soil structure.

Over the years, a concentrated effort, for example, has been the description of soil structure utilizing thin sections and the attempt of a quantitative stereological and topological measurement of the pore phase and solid structure, e.g. Ringrose-Voase (1987). Another approach is to digitize an intact soil cylinder prepared with

resin, section by section using a camera. Vogel et al. (2002), for example, used this multiple-layer approach and were able to reconstruct the soil's architecture in an almost 3D nature making use of image segmentation into pores and the solid phase based on a grey-scale threshold. Models such as this one can subsequently be used for the quantification of pore geometry and the simulation of gas diffusion. Three-dimensional imaging analysis using sophisticated tomographic techniques can now be utilized to study the soil's physical architecture such as pore-space relations as well as biotic interactions from intact soil cores (Young et al. 2001; Vogel et al. 2010). A detailed description of these techniques and methods of image analysis is beyond the scope of this chapter, and the reader is referred to Chap. 6 of this book.

5.5 Improved (Field) Description

In the previous sections, we discussed how soil (legacy) data can be quantified and applied in pedometric studies. In this section we explore further how field descriptions of soil properties can be improved towards a more quantitative assessment. Here, two main approaches are discussed. The first one entails improving conventional description techniques using fuzzy descriptors. The second one is the technologo-statistico approach using field spectrometry and calibration. Over the years, with both technology and computational methods beckoning rapid developments, this area has gained immense popularity and is lately also referred to as the field of digital soil morphometrics which is defined as the 'application of tools and techniques for measuring, mapping and quantifying soil profile attributes and deriving continuous depth functions' (Hartemink and Minasny 2014).

5.5.1 Improved Field Description Using Fuzzy Descriptors

In the example of structure given in Sect. 5.4, an attempt was made to improve the traditional description by firstly defining the variables that are actually being observed and secondly putting those variables on a continuous scale. This seems to be a general approach to improving conventional field descriptors. The benefit of this approach is that we will carry more information about the soil description than is done conventionally.

5.5.1.1 Fuzzy Sets and Numbers

Fuzzy sets and fuzzy logic (Zadeh 1965; Kosko 1994) are a generalization of hard sets. The field of fuzzy logic can thus be used to improve soil morphological descriptions as it was developed to model, classify or categorize something whose membership may be inherently variable or uncertain. Fuzzy logic also allows for a quantitative assessment through numerical coding. In conventional classification systems, an element either has membership (1) or does not have membership to a class (0). In a fuzzy set, membership is fuzzy and not binary, with a point –

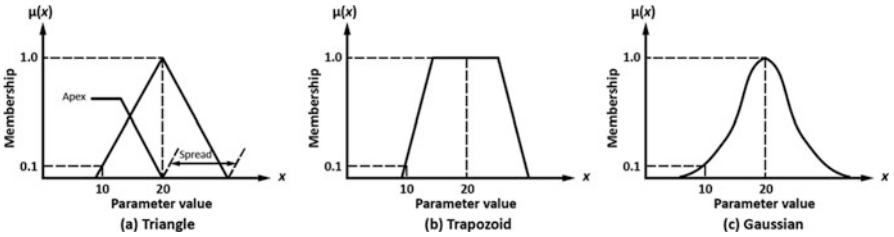


Fig. 5.8 Representation of different fuzzy numbers (After Liu and Samal 2002b)

an element, object or property, for example – having membership to a class A but to a differing degree of belonging (Liu and Samal 2002b). This degree of belongingness of an object to a class can then be formalized by a membership characteristic function, $\mu_A(x)$, with the value of $\mu_A(x)$ at x representative of the degree of membership of x in class A, with x having a value between and including 0 (not belonging) and 1 (completely belonging) (Liu and Samal 2002a).

A formalization of a fuzzy set can be represented as follows (McBratney and Odeh 1997); if we assume that $X = \{x\}$ is a finite set of points, a fuzzy subset A of X is represented by a function $\mu_A(x)$, in the ordered pairs:

$$A = \{x, \mu_A(x)\} \qquad \text{for each } x \in X$$

$$A = x_1, \mu_A(x_1) + x_2, \mu_A(x_2) + \cdots + x_n, \mu_A(x_n) \tag{5.6}$$

where $\mu_A(x)$ is the membership function that defines the belongingness of a point x in A and x$\in X$ expresses that x is an element of X.

Fuzzy numbers are fuzzy subsets of a set of regular, real numbers that describe a set of possible values through a possibility distribution, with each value having its own membership function. A fuzzy number is a convex and normalized fuzzy set. Fuzzy numbers can be of various types and shapes, for example, the most common are triangular-shaped, trapezoidal-shaped or Gaussian fuzzy numbers. Liu and Samal (2002b) give an example of these three types of fuzzy numbers (Fig. 5.8). In this example each number shows a central value of 20, but the variability around this central value differs. The graphs also show the membership function for different values in the fuzzy set, i.e. here the number 20 has a membership of 1, whereas the number 10 has a membership of 0.1 in each case; however, the type of fuzzy numbers and their parameters are determined by the actual distribution of observations.

Liu and Samal (2002a) give an example of a Gaussian membership function which represents the amount of rainfall (Fig. 5.9). Here, it is exemplified that the occurrence of 30 in. of rain is most likely but that 25 and 35 in. of rainfall are about half likely.

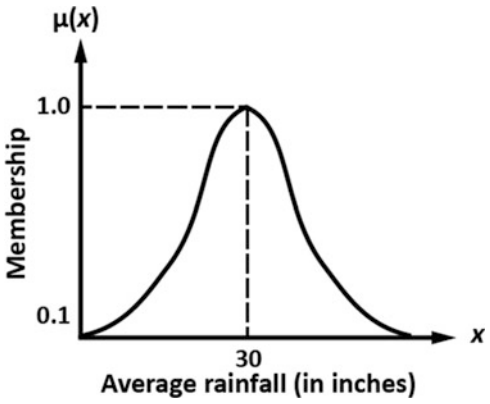


Fig. 5.9 A Gaussian membership function (After Liu and Samal 2002a)

McBratney and Odeh (1997) review the application of fuzzy sets in soil science and provide examples of how fuzzy logic could be used for improving field soil descriptions that are somewhat vague. One of these examples is to put a quantitative measure behind the classification of soil thickness which is often described as 'shallow', 'deep' or 'very deep'. In this example Gaussian membership functions were fitted to three assumed fuzzy subsets with a total range of soil depths between 0 and 200 cm (Fig. 5.10). Here, the fuzzy subset 'B' of 'deep soil' is represented by soil depths of 100 cm. Soils of 100 cm depth therefore fully belong to set B and have a membership of 1, whereas other depths of $0 \leq x < 100$ cm or 100 cm $< x \leq$ 200 cm have only partial membership. The fuzzy membership function, $\mu_B(x)$, for the subset of 'deep soil' is then defined as

$$\mu_B(x) = e^{-}\left(\frac{(x-c)}{1.44\sigma^2}\right) \qquad 0 < x \leq 200 \text{ cm} \tag{5.7}$$

where $c = 100$ cm represents the centroid of fuzzy subset B and σ is the lower crossover point (50 cm), the point at which x has a membership of 0.50. In this case, the higher crossover point is where $x = 150$ cm (Fig. 5.10).

5.5.2 Field Electromagnetic Spectra

As already discussed in Sect. 5.2, the whole electromagnetic spectrum intrinsically holds information about a wide range of biological, chemical and physical soil properties, for example, it contains information about the mineral organic components of the soil (Huete and Escadafal 1991). Several portions of the electromagnetic spectrum are now readily measurable in the field. Conventionally, we use our eyes to 'measure' the visible part of the electromagnetic spectrum, but as discussed in Sects. 5.2, 5.3 and 5.4, this is associated with a degree of subjectiveness. More

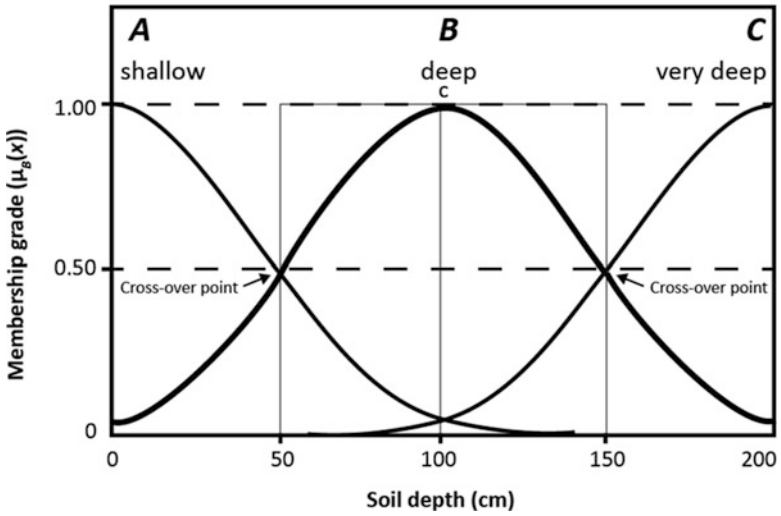


Fig. 5.10 Gaussian membership functions fitted to fuzzy subsets of 'shallow soil' (**a**) (asymmetrical), 'deep soil' (**b**) (symmetrical) and 'very deep soil' (**c**) (asymmetrical). The Gaussian (fuzzy) membership fitted to the subset of 'deep soil' (**b**) is also shown (Redrawn and changed after McBratney and Odeh (1997))

and more, these measurements have thus been replaced by sensors, for example, by digital colour CCD cameras which measure the red, green and blue parts of the spectrum. In the language of remote sensing, this can be called *multispectral*. Furthermore, portable field spectrometers are now readily available that can measure the electromagnetic spectrum (1) in the visible and near-infrared (vis-NIR) range, i.e. from 400–700 to 2500 nm; (2) in the mid-infrared (MIR) range, i.e. from 2500 to 25,000 nm which is still in its infancy; and (3) in the X-ray range in the 0.02–45 kEV region of the electromagnetic spectrum (Fig. 5.11). In the language of remote sensing, this is called *hyperspectral.*

Table 5.5 shows a range of soil chemical and physical properties that can be directly measured or inferred from these spectrometers (also refer to McBratney et al. 2006; Stenberg et al. 2010; Viscarra Rossel et al. 2011; Soriano-Disla et al. 2014; Horta et al. 2015). Good estimates for soil attributes have been achieved from scanning soils in field condition, and formal methods have also been developed that can correct for environmental effects (Horta et al. 2015). At the moment, vis-NIR and pXRF have the most prospect for in-field quantitative soil description, and in the following we will therefore focus on how these sensors can be used to do so. Vis-NIR is known to estimate well the soil's organic component and mineralogy suites (also refer to Table 5.3, Chap. 13), whereas XRF is known to accurately measure the soil's inorganic elemental concentration. The advantage is that these soil properties, some not previously measurable in the field, can be measured simultaneously from the relatively quick, cheap and reproducible observation systems. The disadvantage

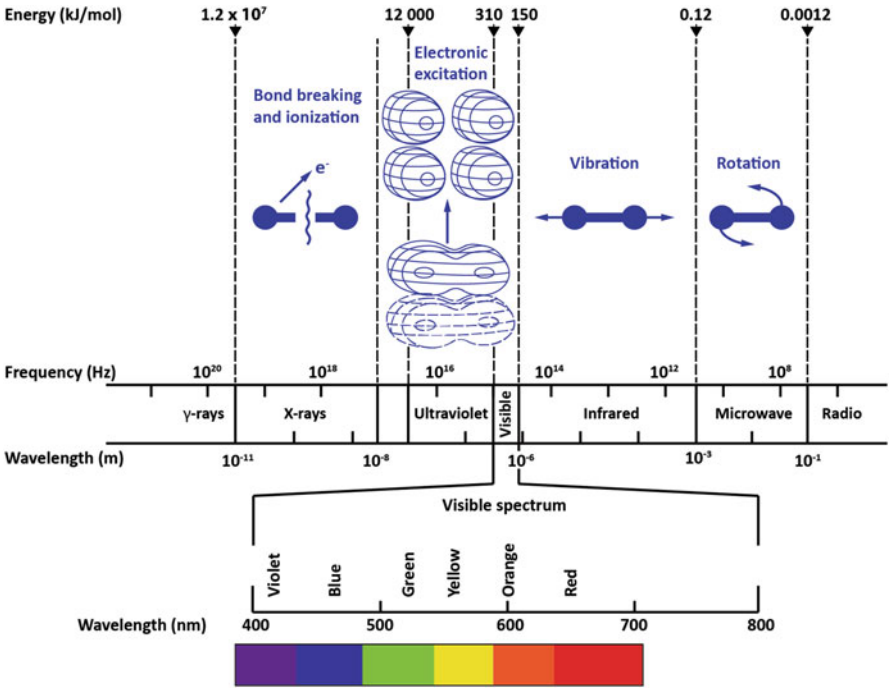


Fig. 5.11 Electromagnetic spectrum showing the regions of interest for field spectroscopy. It is also illustrated how radiation affects molecules and atoms (Redrawn after Harris D (2007))

is that each individual soil property will be measured less precisely than laboratory analysis; however, on the other hand, more measurements can be taken which may compensate for the less accurate in situ predictions (Horta et al. 2015)).

5.5.2.1 Handheld, Portable Sensors

Visible Near-Infrared Spectrometry (vis-NIR)

How are field observations of the visible NIR spectrum that consists of about 2000 different wavelengths (at 2 nm of resolution) used for improved in-field soil description? There are several possibilities. The first approach that comes to mind is to compare the spectra at each of the wavelengths; however this entails a huge amount of data storage. A more feasible approach is to perform a multivariate analysis and measure the spectra and a suite of soil properties at a number of sites and to develop calibration models for those soil properties with respect to the spectra. Principal component regression (PCR) analysis is one method that can be used here. First, a principal component analysis (PCA) is conducted which

Table 5.5 Electromagnetic sensors that can be utilized for quantitative soil field description

Proximal soil sensor	Sensed data	Soil information
Vis-NIR spectrometer[a,b]	Soil reflectance	Soil morphological (colour), chemical (SOC, SOM, TC, pH, CEC, nutrients) and physical properties (clay and sand content, soil mineralogy)
MIR spectrometer[a]	Soil reflectance	Soil chemical and physical properties
pXRF spectrometer[a]	X-ray fluorescence	Soil's elemental concentration ($Z > 11$) (used to identify concentrations of soil contaminants and soil nutrients and to populate soil profile development indices), soil chemical (pH, CEC, soil organic carbon, total carbon) and physical properties (clay and sand content)
Ground-penetrating radar[b]	Radiance energy	Soil conductivity, soil layers, soil structure
Gamma radiometer[a,b]	Gamma radiation	K, Th and U radioisotopes
Electromagnetic induction[a, b]	Apparent conductivity	Soil moisture, salinity, clay content

[a]Portable/handheld
[b]On the go

plots the spectra in a low-dimensional space followed by a regression analysis (Leone and Sommer 2000). As a rule of thumb, the first few principal components usually explain about 99% of the entire spectral variance. Islam et al. (2003) used this method of reducing the amount of predictor variables and then employing stepwise multiple linear regression to estimate a suite of soil chemical and physical properties.

Another well-established chemometric method that can deal with a large amount of spectral data is partial least squares regression (PLSR) which combines the reduction of predictor variables and multiple regression in one single step (Viscarra Rossel et al. 2006b). Furthermore, PLSR can handle multicollinearity, data noise and missing values and is more robust than PCR as it explains predictor as well as response variation (Wold et al. 2001). More recently, data mining techniques have been used for establishing a calibration model as they also allow for non-linear relationships between the spectra and soil attributes of interest. One such technique is Cubist (Quinlan 1993; Kuhn et al. 2013), a rule-based regression tree model which builds a set of rules with each rule representing a linear model of the predictor. These rules then relate the independent variables (soil spectra) to a dependent variable (soil attribute of interest). Minasny and McBratney (2008) tested the accuracy of Cubist compared to PLSR for predicting a range of soil properties and found that Cubist gives higher accuracy, is parsimonious and considers the upper and lower boundary values of the predictant which avoids the prediction of negative values and that the resulting calibration model is easier to interpret.

Related to chemometric methods is the rapid assessment of soil properties via vis-NIR spectrometry in the field, which requires spectral libraries (Shepherd and Walsh 2002) that are representative of the soils. Studies have shown that the geographical range of spectral libraries has an impact on the calibration performance; a reduction of prediction accuracy was observed when spectral libraries are not representative of the soils of the region studied (Sudduth and Hummel 1996). One way of dealing with this problem is to 'spike' the spectral libraries used with some regional spectra with known properties (Viscarra Rossel et al. 2009) or to compile a global vis-NIR spectral library representative of a variety of soil types and environments (Viscarra Rossel et al. 2016). Here, PCA can be used to assess whether the obtained spectra are similar to the spectral library used to predict soil attributes. By fitting a convex hull around the first two principal components which describes the spectral domain of the library used, it can be assessed where the obtained spectra lie within the library domain.

In the early 1990s, Sudduth and Hummel (1993) described a prototype of a portable near-infrared spectrometer for rapid soil analysis, designed for rapid estimation of soil organic matter. However, using vis-NIR to estimate soil attributes (in the field) only received increased popularity in the mid-2000s, with a noticeable, constant increase of published papers on that topic (Bellon-Maurel and McBratney 2011). For example, Ben-Dor et al. (2008) modified an ASD vis-NIR field spectrometer to perform in-field soil profile measurements. They developed an accessory to fit the spectral head of the instrument that was able to penetrate the subsoil for in situ measurements after the drilling of a soil core. As discussed in Sect. 5.2 of this chapter, Viscarra Rossel et al. (2009) reported good estimates of soil colour, mineral composition and clay content from in situ vis-NIR measurements along the pit face of ten soil profiles. Furthermore, Kusumo et al. (2010) achieved good calibrations (RPD = 2.90–5.80) for C and N for scanning intact soil cores.

When scanning the soil in its field condition, we are however dealing with a range of environmental factors influencing the spectra, e.g. soil moisture, surface roughness or smearing (Stenberg et al. 2010). There is a selection of studies that compared predictions from laboratory and field spectra for a variety of soil properties, and the majority of studies found that in general predictions improved for the air-dried spectra (e.g. Chang et al. 2005; Mouazen et al. 2006; Waiser et al. 2007; Fontán et al. 2010). Correcting for the effect of environmental factors, especially soil moisture, is particularly important for in-field estimation of soil attributes using vis-NIR because most spectral libraries used for predictions were established under laboratory conditions, using ground and dried soil.

Several ways have been tested and used to correct for soil moisture as it seems to affect the behaviour of the spectra the most. Soil moisture effects in general result in a higher absorbance of the spectra with increased moisture content. Viscarra Rossel et al. (2009) used the approach of augmenting or 'spiking' a large spectral library containing laboratory-collected spectra with a small proportion of field-collected spectra to improve the prediction performance by introducing spectral

variation to the calibration model. This method is similar to the approach of global moisture modelling (GMM) where variation is also created in the spectral library by including 'differently behaving spectra' (spectra at different soil water content) (Wijewardane et al. 2016). A so-called selective wavelength modelling approach on the other hand was trialled by Wu et al. (2009). Essentially, in this case, certain bands of the spectra that are highly influenced by moisture are not considered in the calibration models. In addition, formal chemometric techniques have been introduced to correct for moisture effects. These usually require a set of soils scanned in field and in standardized air-dried condition from which correction matrixes can be produced. One of the most common approaches is external parameter orthogonalization (EPO) which was developed in the food industry to perform temperature-independent measurement of the sugar content of intact fruits (Roger et al. 2003). Minasny et al. (2011) successfully applied this technique to remove the moisture effect on soil spectra. The algorithm separates a vis-NIR spectrum into two orthogonal components: a useful one that has a direct relationship with the response variable and an unuseful one that is affected by an external variable (soil moisture). A transformation matrix P is then utilized to project spectra into a space less affected by moisture content (Fig. 5.12). Ackerson et al. (2015) tested the capability of the EPO method on field-moist analysis of intact soil spectra for the prediction of clay content, employing a dried and ground vis-NIR spectral library, and found that predicted clay content improved after EPO transformation of the spectra. Another approach is direct standardization (DS) which has been used to standardize spectra acquired from multiple spectrometers (Ge et al. 2011) and was also applied by Ji et al. (2015) to remove the effect of soil moisture from the spectra. The DS method establishes a relationship between the set of laboratory spectra and their corresponding field spectra, employing a transfer set which accounts for in this case the effect of water. Recently, Wijewardane et al. (2016) performed a comparative study including the methods discussed above and tested their effectiveness for correcting the effect of soil moisture on spectral data, concluding that EPO, DS and GMM are all viable approaches for moisture correction of vis-NIR data.

Portable X-Ray Fluorescence Spectrometry (pXRF)

Different to vis-NIR, handheld pXRF field spectrometers directly and simultaneously measure the total concentration of a range of elements in the soil (in mg kg^{-1} and %). This approach is based on factory-installed internal calibration methods. In addition, the EM spectrum resulting from pXRF can also be used in similar ways as described above for the case of vis-NIR spectra, as the spectrum itself might also contain some inherent information of soil chemical and physical properties. O'Rourke et al. (2016) tested this approach and found that using the XRF spectra was more effective to predict total carbon, soil organic carbon and pH

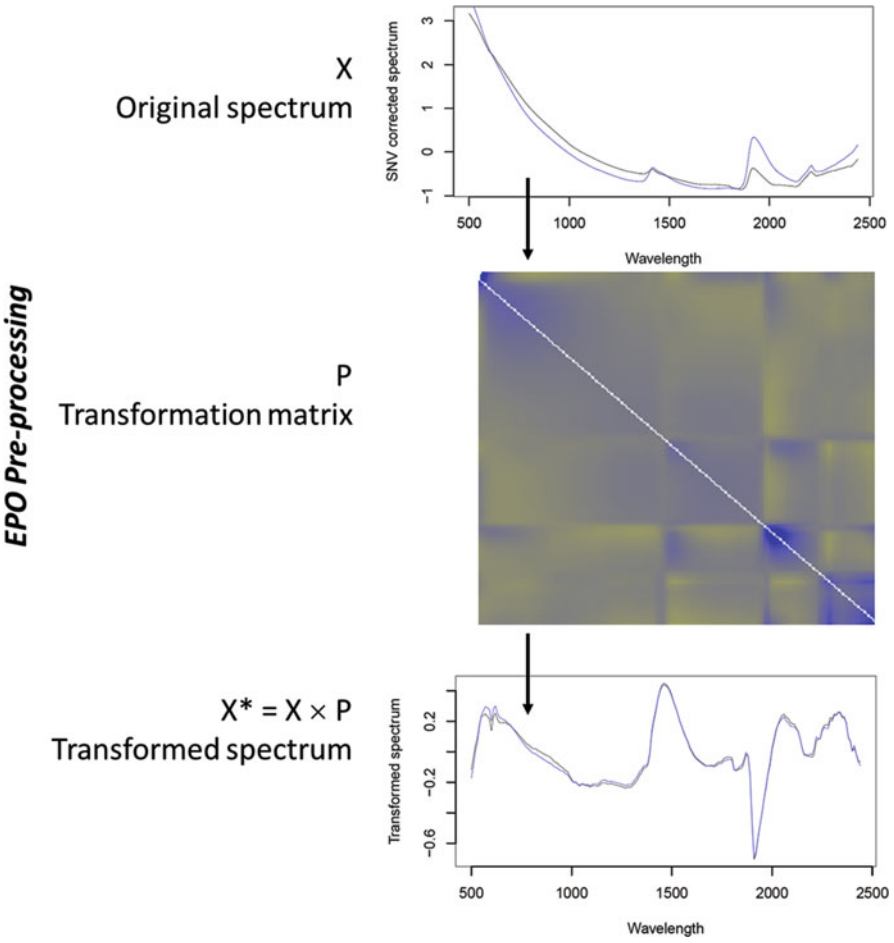


Fig. 5.12 Spectral transformation using the external parameter orthogonalization (EPO) method

when compared to the more conventional use of XRF in soil analysis so far which uses the elemental concentrations directly measured by XRF to predict other soil properties, generally applying multiple linear regression models. For example, the latter approach has been applied to infer a suite of soil properties from laboratory-treated soil data in air-dried and ground condition (soil pH, total C and N, CEC and gypsum content, clay and sand content) (Weindorf et al. 2013, 2014a; Sharma et al. 2014, 2015; Wang et al. 2015; O'Rourke et al. 2016).

Similar to vis-NIR, in-field pXRF measurements are influenced by environmental effects. These are mainly related to soil matrix heterogeneities, i.e. the less homogenous the soil sample, the less representative is the estimate, including soil moisture (Laiho and Perämäki 2005; Markowicz 2008). Both can cause the attenuation or increase of the X-ray signal received from the detector through increased absorption or scatter of the X-rays. Generally, this results in an underestimation or in some circumstances also overestimation of the elemental data and therefore a shift in the intensity of the raw spectral response which is related to the quantity of elements in the soil sample of interest. It has therefore been recommended to not scan soil samples that have a gravimetric moisture content of more than 20% (Laiho and Perämäki 2005). Correction for these effects seems to be not as straightforward as for vis-NIR reflectance spectra because of the above discussed different behaviour of X-rays in relation to the soil matrix. So far, attempts have been made to correct between in-field and laboratory-based pXRF measurements but not to correct the spectral response directly using formal algorithms. Ge et al. (2005) proposed a method for correcting for the effect of moisture, based on the assumption that the intensity of scattered X-rays is directly proportional to the gravimetric moisture content ($w \quad \log(I_o/I_w)$ where I_o represents the X-ray intensity of air-dried samples and I_w the X-ray intensity of field-moist samples). To apply this method, soil moisture also needs to be documented in the field using a soil moisture probe, for example. To correct for soil moisture, Stockmann et al. (2016b) proposed to use a correction factor stemming from the fit of a linear relationship which was found between the in-field and laboratory-based measurement of the iron (Fe) content. In this case, a subset of samples scanned at field-moist and air-dried ground condition is required that represents the soils studied.

A few studies compared in-field and laboratory-based pXRF performance and found that elemental values acceptably compare (Weindorf et al. 2014b; Stockmann et al. 2016a, b). It was noted, however, that this varies element specifically, with higher variability between field-moist and air-dried ground conditions for lighter elements with lower atomic numbers such as Al and Si. This is related to the limit of detection of the pXRF detector which varies between 5 mg kg^{-1} and 1% depending on the atomic number of the elements for an Olympus Delta Premium pXRF device, for example, and therefore also the abundance of these elements naturally in the soil.

Synergetic Use of Handheld, Portable Sensors for Predictive Power Improvement

Methods have been introduced to improve the estimation of soil attributes through the combination of data from multiple sensors which has high potential for in-field analysis. Wang et al. (2015) used concatenation for fusing the principal

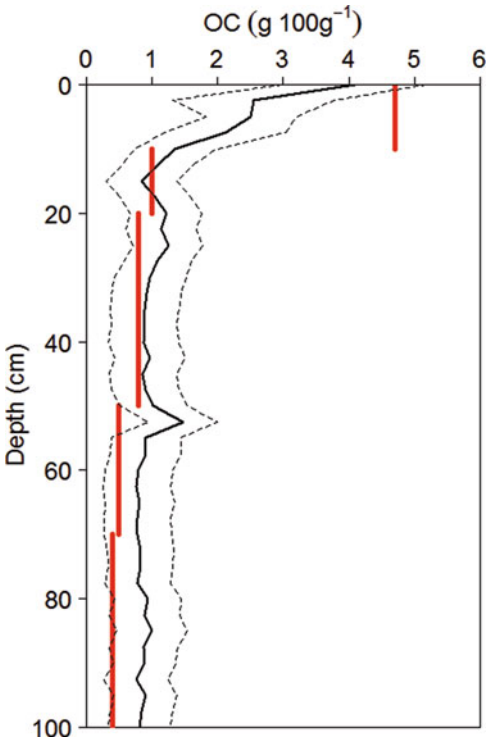


Fig. 5.13 Comparison of soil organic carbon levels measured using conventional laboratory techniques versus in situ, diffuse reflectance vis-NIR spectrometry. *Vertical red lines* represent the observed OC content of dried and ground bulk samples for each horizon as measured using the dry combustion method. In situ results are displayed as the mean (*solid black line*) and 95% confidence interval (*dotted black lines*) based on predictions of 50 Cubist models. Measurements were taken every 2.5 cm to a depth of 1 m

components of NIR spectral data from PCA with pXRF-measured elemental data which improved the accuracy of soil texture prediction. Viscarra Rossel et al. (2006b) used concatenation, i.e. in this case the spectral fusion of the vis-NIR and MIR region which resulted in a minor improvement of the prediction of soil texture compared to single sensor predictions. This approach only shows potential for spectral data that use the same principle, soil reflectance, but seems not sufficient when fusion of spectral data from sensors with differing measuring principles is desired. The following approaches seem to be more applicable. O'Rourke et al. (2016) were able to improve the predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties through methods of model averaging, whereas Jones and McBratney (2016a) used an integrated chemometric and mass balance approach to estimate the soil's mineral component, from vis-NIR and pXRF data. Together with the prospects of improving predictions from spectra in field condition through data assimilation methods, focus is now also directed towards the quantification of prediction uncertainties (Fig. 5.13).

5.5.2.2 On Digital Soil Morphometrics (DSMorph)

The availability of portable, handheld devices that measure various parts of the electromagnetic spectrum and are robust enough to be taken to the field in combination with advances in mathematical approaches and computational power has warranted the introduction of digital soil morphometrics or in short DSMorph (Hartemink and Minasny 2014). The establishment of DSMorph as part of the pedometric family has potential as it promises to introduce a scientific and technological overhaul of conventional soil field descriptions by "digitally enriching the toolkit of the field pedologist" (Jones and McBratney 2016b). DSMorph introduces quantitative data collection directly in the field and thus also offers the possibility of describing and measuring soil attributes that are otherwise unattainable such as the in situ analysis of the soil's mineralogical composition. Jones and McBratney (2016a) demonstrate how the conjoint use of vis-NIR and pXRF could achieve this. In conventional field descriptions, the pedologist can only approximate the presence of certain clay minerals from experience; a soil that shows evidence of shrink-and-swell behaviour in the form of pronounced cracks through the profile and the presence of lenticular peds will most certainly contain the secondary clay mineral smectite, for example. Furthermore, DSMorph has the potential to increase sampling intervals and thus to quantify the spatial variation of soil attributes more readily as compared to conventional methods (Jones and McBratney 2016b) and in this regard offers the prospect of adaptive sampling and analysis as outlined in Horta et al. (2015). Jones and McBratney (2016b) argue however that the "marriage of digital data collection with morphometric approaches is going to be crucial to the success of DSMorph".

Examples of applications of DSMorph techniques to revolutionize traditional soil description, including novel and potential applications, are given in Hartemink and Minasny (2014) and Hartemink and Minasny (2016). To name an example here, recently, Fajardo et al. (2016) used fuzzy clustering of vis-NIR spectra to assess soil properties of soil profiles and to identify zones of similarity, i.e. zones that were spectrally homogenous, to identify soil morphological horizons (Fig. 5.14). This approach could enhance and/or change the way we describe soil layers or horizons in the field.

Stockmann et al. (2016a) calculated geochemical indices from pXRF-derived elemental data acquired from the soil pit face and used these to investigate parent material origin and pedological processes (Fig. 5.15). DSMorph can therefore also help to explore soil-forming processes directly in the field which complements and advances conventional soil field descriptions.

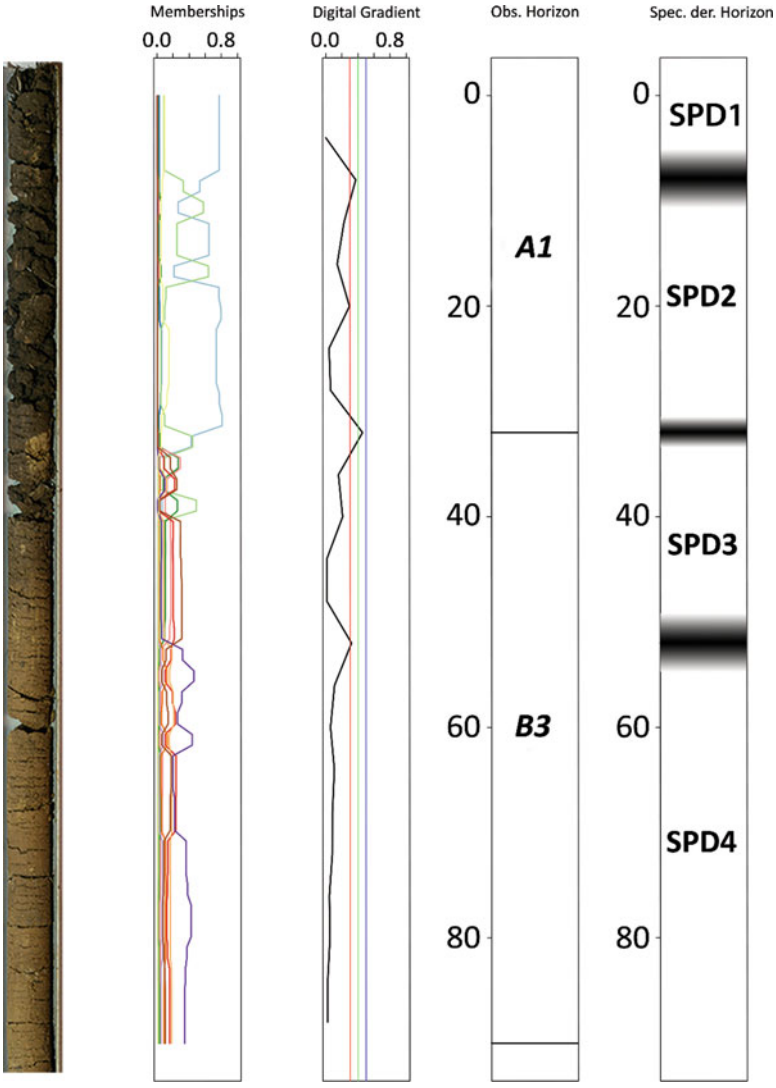


Fig. 5.14 *Left* to *right*. Photograph of soil core, fuzzy membership classes, digital gradient, horizons observed using conventional techniques, spectrally derived horizons (Changed after Fajardo et al. (2016))

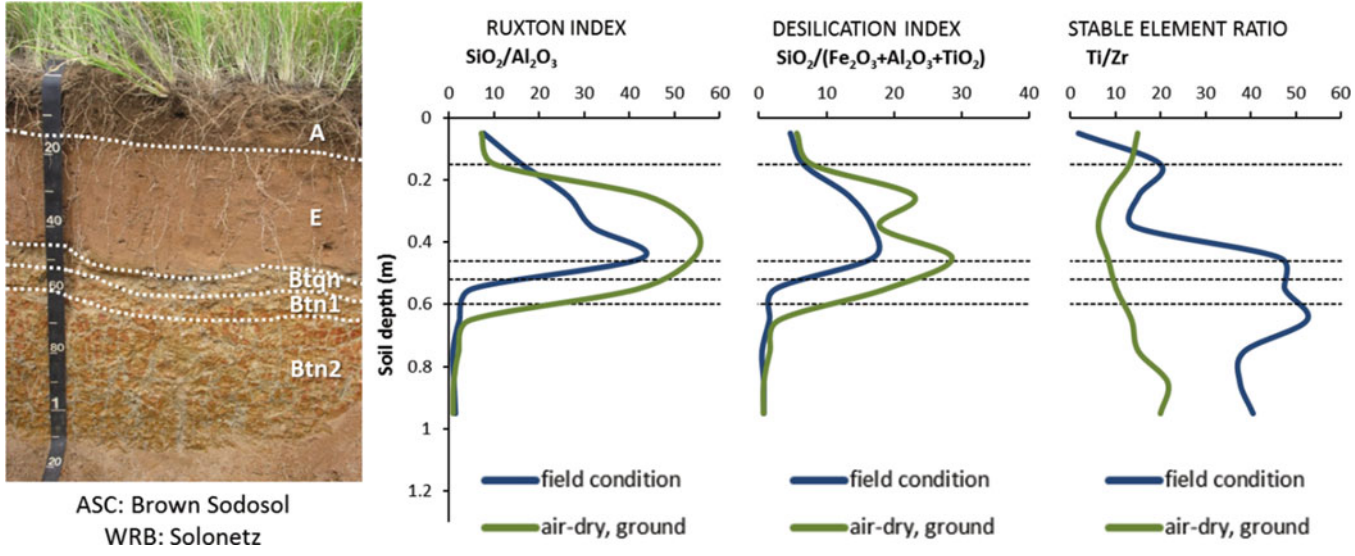


Fig. 5.15 Pattern of geochemical indices and the elemental ratio of Ti/Zr with increasing depth in a Sodosol profile, Nowley Farm, Liverpool Plains, NSW, Australia, assessed in both field condition and air-dried and ground condition. The abrupt change in indices behavior down the profile indicates the presence of multiple parent materials (Taken from Stockmann et al. (2016a), used with permission)

References

Ackerson JP, Demattê JAM, Morgan CLS (2015) Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. Geoderma 259–260:196–204

Anderson A, McBratney A (1995) Soil aggregates as mass fractals. Soil Res 33:757–772

Atterberg A (1905) Die rationelle Klassifikation der Sande und Kiese. Chemiker-Zeitung 29: 195–198

Bagnold RA (1954) The physics of blown sand and desert dunes. Methuen, London

Barrett LR (2002) Spectrophotometric color measurement in situ in well drained sandy soils. Geoderma 108:49–77

Barron V, Torrent J (1986) Use of the Kubelka – Munk theory to study the influence of iron oxides on soil colour. J Soil Sci 37:499–510

Baumann K, Schöning I, Schrumpf M, Ellerbrock RH, Leinweber P (2016) Rapid assessment of soil organic matter: soil color analysis and Fourier transform infrared spectroscopy. Geoderma 278:49–57

Beaudette D, Roudier P (2016) aqp: algorithms for quantitative pedology. A toolkit for soil scientists. R package version 1.9.10. URL http://aqp.r-forge.r-project.org/

Bedidi A, Cervelle B, Madeira J, Pouget M (1992) Moisture effects on visible spectral characteristics of lateritic soils. Soil Sci 153:129–141

Bellon-Maurel V, McBratney A (2011) Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – critical review and research perspectives. Soil Biol Biochem 43:1398–1410

Ben-Dor E, Heller D, Chudnovsky A (2008) A novel method of classifying soil profiles in the field using optical means. Soil Sci Soc Am J 72:1113–1123

Berezin PN, Voronin AD (1981) Use of probability distribution functions for describing the particle-size distribution of soils. Moscow University Soil Science Bulletin (English Translation of Vestnik Moskovskogo Universiteta 36, 28–34

Billmeyer FW, Saltzman M (1981) Principles of color technology. Wiley, New York

Buchan GD (1989) Applicability of the simple lognormal model to particle-size distribution in soils. Soil Sci 147:155–161

Buchan GD, Grewal KS, Robson AB (1993) Improved models of particle-size distribution: an illustration of model comparison techniques. Soil Sci Soc Am J 57:901–908

Chang C-W, Laird DA, Hurburgh CRJ (2005) Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. Soil Sci 170:244–255

Chaplot V, Bernoux M, Walter C, Curmi P, Herpin U (2001) Soil carbon storage prediction in temperate hydromorphic soils using a morphologic index and digital elevation model. Soil Sci 166:48–60

CIE (1931) Commission Internationale de l'Eclairage (CIE) Proceedings. Cambridge University Press, Cambridge

Clark MW (1981) Quantitative shape analysis: a review. Math Geol 13:303–320

Eck DV, Hirmas DR, Giménez D (2013) Quantifying soil structure from field excavation walls using multistripe laser triangulation scanning. Soil Sci Soc Am J 77(4):1319–1328

Escadafal R (1993) Remote sensing of soil color: principles and applications. Remote Sens Rev 7:261–279

Escadafal R, Huete A (1991) Improvement in remote sensing of low vegetation cover in arid regions by correcting vegetation indices for soil noise. C R Acad Sci 312:1385–1391

Escadafal R, Girard M-C, Courault D (1989) Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. Remote Sens Environ 27:37–46

Evans CV, Franzmeier DP (1988) Color index values to represent wetness and aeration in some Indiana soils. Geoderma 41:353–368

Fajardo M, McBratney A, Whelan B (2016) Fuzzy clustering of Vis-NIR spectra for the objective recognition of soil morphological horizons in soil profiles. Geoderma 263:244–253

Fernandez RN, Schulze DG (1987) Calculation of soil color from reflectance Spectra1. Soil Sci Soc Am J 51:1277–1282
Fontán JM, Calvache S, López-Bellido RJ, López-Bellido L (2010) Soil carbon measurement in clods and sieved samples in a Mediterranean vertisol by visible and near-infrared reflectance spectroscopy. Geoderma 156:93–98
Fredlund MD, Fredlund DG, Wilson GW (2000) An equation to represent grain-size distribution. Canadian Geotechnical Journal 37(4):817–827. doi:10.1139/t00-015
Ge L, Lai W, Lin Y (2005) Influence of and correction for moisture in rocks, soils and sediments on in situ XRF analysis. X-Ray Spectrom 34:28–34
Ge Y, Morgan CLS, Grunwald S, Brown DJ, Sarkhot DV (2011) Comparison of soil reflectance spectra and calibration models obtained using multiple spectrometers. Geoderma 161:202–211
Gee GW, Or D (2002) 2.4 Particle-size analysis. In: Methods of soil analysis. Part 4. Science Society of America, Madison, pp 255–293
Geeves GW, Cresswell HP, Murphy BW, Gessler PE, Chartres CJ, Little IP, Bouwman GM (1995) The physical, chemical and morphological properties of soils in the wheat-belt of southern NSW and northern Victoria. NSW Department of Conservation and Land Management/CSIRO Aust. Division of Soils Occasional report
Getreuer P (2011) Colorspace transformations. Matlab package
Gobin A, Campling P, Deckers J, Feyen J (2000) Quantifying soil morphology in tropical environments: methods and application in soil classification. Soil Sci Soc Am J 64:1423–1433
Gómez-Robledo L, López-Ruiz N, Melgosa M, Palma AJ, Capitán-Vallvey LF, Sánchez-Marañón M (2013) Using the mobile phone as Munsell soil-colour sensor: an experiment under controlled illumination conditions. Comput Electron Agric 99:200–208
Harris D (2007) Quantitative chemical analysis, 7th edn. WHFrrman&Co, New York
Hartemink AE, Minasny B (2014) Towards digital soil morphometrics. Geoderma 230–231:305–317
Hartemink AE, Minasny B (2016) Digital soil morphometrics. Springer International Publishing, Cham
He X, Vepraskas MJ, Lindbo DL, Skaggs RW (2003) A method to predict soil saturation frequency and duration from soil color. Soil Sci Soc Am J 67:961–969
Hodgson JM, Hollis JM, Jones RJA, Palmer RC (1976) A comparison of field estimates and laboratory analyses of the silt and clay contents of some west midland soils. J Soil Sci 27: 411–419
Holden NM (1993) A two-dimensional quantification of soil ped shape. J Soil Sci 44:209–219
Horta A, Malone B, Stockmann U, Minasny B, Bishop TFA, McBratney AB, Pallasser R, Pozza L (2015) Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: a prospective review. Geoderma 241–242:180–209
Huete AR, Escadafal R (1991) Assessment of biophysical soil properties through spectral decomposition techniques. Remote Sens Environ 35:149–159
Hwang SI (2004) Effect of texture on the performance of soil particle-size distribution models. Geoderma 123:363–371
Hwang SI, Lee KP, Lee DS, Powers SE (2002) Models for estimating soil particle-size distributions. Soil Sci Soc Am J 66:1143–1150
Ihaka R, Murrell P, Hornik K, Fischer JC, Zeileis A (2015) Colourspace: color space manipulation. R package version 1.2-6. URL http://CRAN.R-project.org/package=colorspace
Islam K, Singh B, McBratney A (2003) Simultaneous estimation of several soil properties by ultraviolet, visible, and near-infrared reflectance spectroscopy. Soil Res 41:1101–1114
Ivey JL, McBride RA (1999) Delineating the zone of topsoil disturbance around buried utilities on agricultural land. Land Degrad Dev 10:531–544
Jaky J (1944) Soil mechanics. Egyetemi Nyomda, Budapest (In Hungarian). Cited it Probalistic solutions in geotechnics. Dev Geotech Eng 46:157
Ji W, Viscarra Rossel RA, Shi Z (2015) Accounting for the effects of water and the environment on proximally sensed vis–NIR soil spectra and their calibrations. Eur J Soil Sci 66:555–565

Jones EJ, McBratney AB (2016a) In situ analysis of soil mineral composition through conjoint use of visible, near-infrared and X-ray fluorescence spectroscopy. In: Hartemink EA, Minasny B (eds) Digital soil morphometrics. Springer International Publishing, Cham, pp 51–62

Jones EJ, McBratney AB (2016b) What is digital soil morphometrics and where might it be going? In: Hartemink EA, Minasny B (eds) Digital soil morphometrics. Springer International Publishing, Cham, pp 1–15

Knight HG (1938) New size limits for silt and clay. Soil Sci Soc Am Proc 2:502

Kosko B (1994) Fuzzy systems as universal approximators. IEEE Trans Comput 43:1329–1333

Kuhn M, Weston S, Keefer C, Coulter N (2013) Cubist package. Version 0.0.13

Kusumo BH, Hedley MJ, Tuohy MP, Hedley CB, Arnold GC (2010) Predicting soil carbon and nitrogen concentrations and pasture root densities from proximally sensed soil spectral reflectance. In: Viscarra Rossel RA, McBratney AB, Minasny B (eds) Proximal soil sensing. Springer, Switzerland

Laiho JVP, Perämäki P (2005) Evaluation of portable X-ray fluorescence (PXRF) sample preparation methods. Spec Paper Geol Surv Finland 38:73–82

Legros JP, Pedro G (1985) The cause of particle-size distribution in soil profiles derived from crystalline rocks, France. Geoderma 36:15–25

Leone AP, Escadafal R (2001) Statistical analysis of soil colour and spectroradiometric data for hyperspectral remote sensing of soil properties (example in a southern Italy Mediterranean ecosystem). Int J Remote Sens 22:2311–2328

Leone AP, Sommer S (2000) Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the southern Apennines (Italy). Remote Sens Environ 72:346–359

Liu M, Samal A (2002a) Cluster validation using legacy delineations. Image Vis Comput 20: 459–467

Liu M, Samal A (2002b) A fuzzy clustering approach to delineate agroecozones. Ecol Model 149:215–228

Madeira J, Bedidi A, Cervelle B, Pouget M, Flay N (1997) Visible spectrometric indices of hematite (Hm) and goethite (Gt) content in lateritic soils: the application of a Thematic Mapper (TM) image for soil-mapping in Brasilia, Brazil. Int J Remote Sens 18:2835–2852

Mandelbrot BB (1989) Fractal geometry: what is it, and what does it do? Proc R Soc Lond A Math Phys Sci 423:3–16

Markowicz AA (2008) Quantification and correction procedures. In: Potts PJ, West M (eds) Portable X-ray fluorescence spectrometry: capabilities for in situ analysis. The Royal Society of Chemistry, Cambridge, pp 13–38

Mathieu R, Pouget M, Cervelle B, Escadafal R (1998) Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and Typic soil color of an arid environment. Remote Sens Environ 66:17–28

Mathieu R, Cervelle B, Rémy D, Pouget M (2007) Field-based and spectral indicators for soil erosion mapping in semi-arid mediterranean environments (coastal cordillera of central Chile). Earth Surf Process Landf 32:13–31

McBratney AB, Odeh IOA (1997) Fuzzy sets in soil science application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77:85–113

McBratney AB, Anderson AN, Lark RM, Odeh IO (2002) 1.8 newer application techniques. In: Dane JH, Topp CG (eds) Methods of soil analysis: part 4 physical methods. Soil Science Society of America, Madison, pp 159–200

Melville MD, Atkinson G (1985) Soil colour: its measurement and its designation in models of uniform colour space. J Soil Sci 36:495–512

Minasny B, McBratney AB (2001) The Australian soil texture boomerang: a comparison of the Australian and USDA/FAO soil particle-size classification systems. Soil Res 39:1443–1451

Minasny B, McBratney AB (2008) Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom Intell Lab Syst 94:72–79

Minasny B, McBratney AB, Bellon-Maurel V, Roger J-M, Gobrecht A, Ferrand L, Joalland S (2011) Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167–168:118–124

Moeys J (2015) Soiltexture: functions for soil texture plot, classification and transformation. R package version 1.3.3. http://CRAN.R-project.org/package=soiltexture

Mouazen AM, De Baerdemaeker J, Ramon H (2006) Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. J Near Infrared Spectrosc 14:189–199

Munsell Color (2009) Munsell soil-color charts with genuine Munsell color chips. Munsell Color, Grand Rapids

Nemes A, Wösten JHM, Lilly A, Oude Voshaar JH (1999) Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. Geoderma 90:187–202

Nikiforoff CC (1941) Morphological classification of soil structure. Soil Sci 52:193–212

Odeh IOA, Chittleborough DJ, McBratney AB (1991) Elucidation of soil-landform interrelationships by canonical ordination analysis. Geoderma 49:1–32

O'Rourke SM, Stockmann U, Holden NM, McBratney AB, Minasny B (2016) An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. Geoderma 279:31–44

Padarian J, Minasny B, McBratney A (2012) Using genetic programming to transform from Australian to USDA/FAO soil particle-size classification system. Soil Res 50(6):443–446

Pettijohn FJ (1957) Sedimentary rocks. Harper and Row, New York

Quinlan JR(1993) Combining instance-based and model-based learning. Proceedings of the tenth international conference on machine learning, San Mateo, CA, pp 236–243

Ringrose-Voase AJ (1987) A scheme for the quantitative description of soil macrostructure by image analysis. J Soil Sci 38:343–356

Roger J-M, Chauchard F, Bellon-Maurel V (2003) EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemom Intell Lab Syst 66:191–204

Roudier P, Manderson A, Hedley C (2016) Advances towards quantitative assessments of soil profile properties. In: Hartemink EA, Minasny B (eds) Digital soil morphometrics. Springer International Publishing, Cham, pp 113–132

Rousseva SS (1997) Data transformations between soil texture schemes. Eur J Soil Sci 48:749–758

Sánchez-Marañón M, Huertas R, Melgosa M (2005) Colour variation in standard soil-colour charts. Soil Res 43:827–837

Sánchez-Marañón M, Martín-García JM, Delgado R (2011) Effects of the fabric on the relationship between aggregate stability and color in a Regosol–Umbrisol soilscape. Geoderma 162:86–95

Sharma A, Weindorf DC, Man T, Aldabaa AAA, Chakraborty S (2014) Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction (pH). Geoderma 232–234:141–147

Sharma A, Weindorf DC, Wang D, Chakraborty S (2015) Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). Geoderma 239–240:130–134

Shepherd KD, Walsh MG (2002) Development of reflectance spectral libraries for characterization of soil properties. Soil Sci Soc Am J 66:988–998

Shields JA, St. Arnaud RJ, Paul EA, Clayton JS (1966) Measurement of soil color. Can J Soil Sci 46:83–90

Shiozawa S, Campbell G (1991) On the calculation of the mean particle diameter and standard deviation from sand, silt, and clay fractions. Soil Sci 152(6):427–431

Shirazi MA, Hart JW, Boersma L (1988) A unifying quantitative analysis of soil texture: improvement of precision and extension of scale. Soil Sci Soc Am J 52:181–190

Shirazi MA, Boersma L (1984) A unifying quantitative analysis of soil texture. Soil Sci Soc Am J 48:142–147

Skaggs TH, Arya LM, Shouse PJ, Mohanty BP (2001) Estimating particle-size distribution from limited soil texture data. Soil Sci Soc Am J 65(4):1038–1044

Soriano-Disla JM, Janik LJ, Viscarra Rossel RA, MacDonald LM, McLaughlin MJ (2014) The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl Spectrosc Rev 49:139–186

Spielvogel S, Knicker H, Kögel-Knabner I (2004) Soil organic matter composition and soil lightness. J Plant Nutr Soil Sci 167:545–555

Stenberg B, Viscarra Rossel RA, Mouazen AM, Wetterlind J (2010) Visible and near infrared spectroscopy in soil science. Adv Agron 107:163–215

Stockmann U, Cattle SR, Minasny B, McBratney AB (2016a) Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. Catena 139:220–231

Stockmann U, Jang HJ, Minasny B, McBratney AB (2016b) The effect of soil moisture and texture on Fe concentration using portable X-ray fluorescence spectrometers. In: Hartemink EA, Minasny B (eds) Digital soil morphometrics. Springer International Publishing, Cham, pp 63–71

Sudduth KA, Hummel JW (1993) Portable, near-infrared spectrophotometer for rapid soil analysis. Trans Am Soc Agric Eng 36:185–193

Sudduth KA, Hummel JW (1996) Geographic operating range evaluation of a NIR soil sensor. Trans Am Soc Agric Eng 39:1599–1604

Sutherland RA, Lee C-T (1994) Discrimination between coastal subenvironments using textural characteristics. Sedimentology 41:1133–1145

Terrain TNCFSA (2009) Australian soil and land survey field handbook. CSIRO Publishing, Melbourne

Thompson JA, Bell JC (1996) Color index for identifying hydric conditions for seasonally saturated Mollisols in Minnesota. Soil Sci Soc Am J 60:1979–1988

Torrent J, Schwertmann U, Fechter H, Alferez F (1983) Quantitative relationships between soil color and hematite content. Soil Sci 136:354–358

Viscarra Rossel RA, Minasny B, Roudier P, McBratney AB (2006a) Colour space models for soil science. Geoderma 133:320–337

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO (2006b) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59–75

Viscarra Rossel RA, Fouad Y, Walter C (2008) Using a digital camera to measure soil organic carbon and iron contents. Biosyst Eng 100:149–159

Viscarra Rossel RA, Cattle SR, Ortega A, Fouad Y (2009) In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150:253–266

Viscarra Rossel RA, Adamchuk VI, Sudduth KA, McKenzie NJ, Lobsey C (2011) Proximal soil sensing: an effective approach for soil measurements in space and time. Adv Agron 113: 243–291

Viscarra Rossel RA, Behrens T, Ben-Dor E, Brown DJ, Demattê JAM, Shepherd KD, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H, Barthès BG, Bartholomeus HM, Bayer AD, Bernoux M, Böttcher K, Brodský L, Du CW, Chappell A, Fouad Y, Genot V, Gomez C, Grunwald S, Gubler A, Guerrero C, Hedley CB, Knadel M, Morrás HJM, Nocita M, Ramirez-Lopez L, Roudier P, Campos EMR, Sanborn P, Sellitto VM, Sudduth KA, Rawlins BG, Walter C, Winowiecki LA, Hong SY, Ji W (2016) A global spectral library to characterize the world's soil. Earth Sci Rev 155:198–230

Vogel HJ, Cousin I, Roth K (2002) Quantification of pore structure and gas diffusion as a function of scale. Eur J Soil Sci 53:465–473

Vogel HJ, Weller U, Schlüter S (2010) Quantification of soil structure based on Minkowski functions. Comput Geosci 36:1236–1245

Waiser TH, Morgan CLS, Brown DJ, Hallmark CT (2007) In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. Soil Sci Soc Am J 71: 389–396

Walker PH, Chittleborough DJ (1986) Development of particle-size distributions in some Alfisols of southeastern Australia. Soil Sci Soc Am J 50:394–400

Wang D, Chakraborty S, Weindorf DC, Li B, Sharma A, Paul S, Ali MN (2015) Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. Geoderma 243–244:157–167

Weindorf DC, Herrero J, Castañeda C, Bakr N, Swanhart S (2013) Direct soil gypsum quantification via portable x-ray fluorescence spectrometry. Soil Sci Soc Am J 77:2071–2077

Weindorf DC, Bakr N, Zhu Y (2014a) Chapter one – advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. In: Donald LS (ed) Advances in agronomy. Academic, Amsterdam, pp 1–45

Weindorf DC, Bakr N, Zhu Y, McWhirt A, Ping CL, Michaelson G, Nelson C, Shook K, Nuss S (2014b) Influence of ice on soil elemental characterization via portable X-ray fluorescence spectrometry. Pedosphere 24:1–12

Wickham C (2016) Munsell: utilities for using Munsell colours. R package version 0.4.3. http://CRAN.R-project.org/package=munsell

Wijewardane NK, Ge Y, Morgan CLS (2016) Prediction of soil organic and inorganic carbon at different moisture contents with dry ground VNIR: a comparative study of different approaches. Eur J Soil Sci 67:605

Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130

Wu C-Y, Jacobson AR, Laba M, Baveye PC (2009) Alleviating moisture content effects on the visible near-infrared diffuse-reflectance sensing of soils. Soil Sci 174:456–465

Wyrwoll KH, Smyth GK (1985) On using the log-hyperbolic distribution to describe the textural characteristics of eolian sediments. J Sediment Petrol 55:471–478

Wyszecki G, Stiles WS (1982) Color science: concepts and methods, quantitative data and formulae. Wiley, New York

Young IM, Crawford JW, Rappoldt C (2001) New methods and models for characterising structural heterogeneity of soil. Soil Tillage Res 61:33–45

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

Chapter 6
Scaling Characteristics of Soil Structure

Ana M. Tarquis, Iván G. Torre, Juan J. Martín-Sotoca, Juan C. Losada, Juan B. Grau, Nigel R. A. Bird, and Antonio Saa-Requejo

"For a complex natural shape, dimension is relative. It varies with the observer. The same object can have more than one dimension, depending on how you measure it and what you want to do with it. And dimension need not be a whole number; it can be fractional. Now an ancient concept, dimension, becomes thoroughly modern".

Benoît B. Mandelbrot, The (Mis)Behaviour of Markets (2004)

A.M. Tarquis (✉) • J.J. Martín-Sotoca
Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Complex Systems Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain
e-mail: anamaria.tarquis@upm.es

I.G. Torre
Complex Systems Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

Department of Applied Mathematics and Statistics, Universidad Politécnica de Madrid (UPM), Madrid, Spain

J.C. Losada
Complex Systems Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

J.B. Grau
Group of Automation of Signals and Communications, Universidad Politécnica de Madrid (UPM), Madrid, Spain

N.R.A. Bird
Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

A. Saa-Requejo
Research Center for the Management of Agricultural and Environmental Risks (CEIGRAM), Universidad Politécnica de Madrid (UPM), Madrid, Spain

Resource Assessment Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_6

6.1 Introduction

As previously discussed in Chap. 5, soil structure is defined by the spatial arrangement of soil primary particles and aggregates. There is increasing evidence that quantitative characterization of the soil structure and of its heterogeneity and complexity holds the key to a deeper understanding on physical, chemical, and biological processes that take place within the soil (Vogel 2000; Rockhold et al. 2004; Young et al. 2008; Blair et al. 2007; Pajor et al. 2010; Kravchenko et al. 2010; Dullien 2012). Therefore, it is very important to obtain an accurate description of soil structure which best approximates to reality. Although many parameters may be used to attempt to describe irregular morphology, the spatial arrangement of the most prominent features is a challenging problem across a wide range of disciplines (Ripley 1988; Griffith 1988; Baveye and Boast 1988).

In this chapter we will present scaling and multiscaling analysis on soil images and summarize the work in this area that has been going on for several years until now.

6.2 Image Acquisition

The first image analysis systems were available in the early 1970s and have since become increasingly accessible to soil scientists. Currently, low-cost software-based image analysis systems make automated analysis of soil pore space very easy and can be done in a two-dimensional (2D) as well as a three-dimensional (3D) way.

6.2.1 Soil Thin Sections

2D images of the soil pore space can be obtained when thin sections are cut from polished faces of undisturbed soil blocks, previously impregnated with a resin containing a fluorescent dye (Fig. 6.1). Subsequently, these thin sections are illuminated with UV light or viewed with a fluorescence microscope (Geyger and Beckman 1967; Murphy et al. 1977; Ringrose-Voase and Bullock 1984; Bouabid et al. 1992). Another method is to view the pore space with a scanning electron microscope (SEM) on finely polished thin sections in backscattered mode (Fies and Bruand 1990).

Significant contributions to the characterization of soil pores by 2D image analysis procedures are those by Jongerius et al. (1972), Ismail (1975), Murphy et al. (1977), Bullock and Murphy (1980), Ringrose-Voase and Bullock(1984), Moran et al. (1989), McBratney et al. (1992), Terribile and Fitzpatrick (1992), Protz et al. (1992), and Protz and Van den Bygaart (1998).

The 2D morphological image analysis of undisturbed samples was the most typical technique for many years. The interpretation of micromorphometric results from two-dimensional sections is often biased by anisotropy of the investigated structural elements (Vogel et al. 1993). Stereological methods consist in extracting quantitative

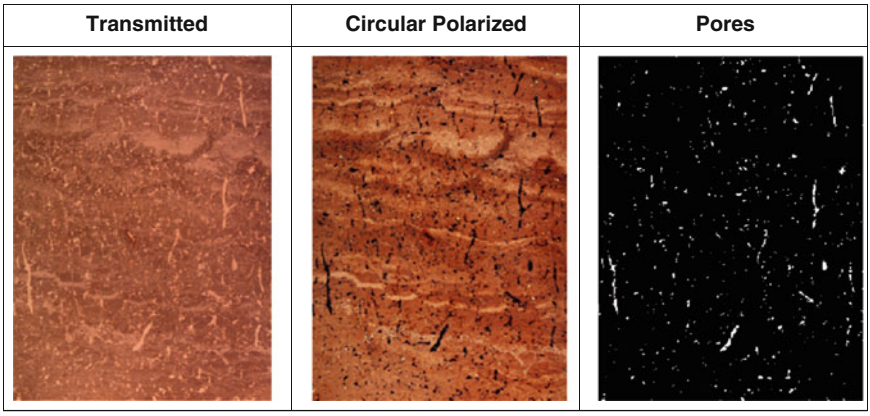


Fig. 6.1 Sample prepared for image analysis following the procedure described by Protz and Van den Bygaart (1998). Imaging thin sections obtained with a Kodak 460 RGB camera using transmitted and circularly polarized illumination. Then EASI/PACE software was used to classify the original transmitted and circularly polarized data and to separate the void bitmap

information about a three-dimensional material from measurements made on two-dimensional planar sections. These have been applied in soil science, but they require strong assumptions about the shape of pores and particles (McBratney and Moran 1990).

6.2.2 Computer Tomography (CT) and Image Processing

X-ray computed tomography (CT scans) is a relatively recent nondestructive testing method which offers an attractive opportunity for the three-dimensional insight of the inner structure of objects and materials. Nowadays, due to the great technological advances and the computational power of modern calculators, CT systems are massively employed for a wide range of purposes in the scientific and industrial sectors, e.g., flaw detection, failure analysis, metrology, and reverse engineering.

The basic components forming a CT system are the X-ray source, characterized by a micrometric or sub-micrometric focal spot size; the detection system, to collect the transmitted radiation emerging from the sample; and the sample positioning stages. The CT scan is done through recording on the detector, which is placed behind the sample, a set of planar projections while the sample rotates inside the incident beam over approximately the angular range between 0 and 180/360°. A sufficient number of those angular views should be acquired at regular or known steps, in order to efficiently reconstruct a set of horizontal cross sections (the slices) of the object by means of a well-established mathematical procedure known as the filtered altered back-projection (FBP) algorithm (Torre 2014). Figure 6.2 illustrates the main components needed.

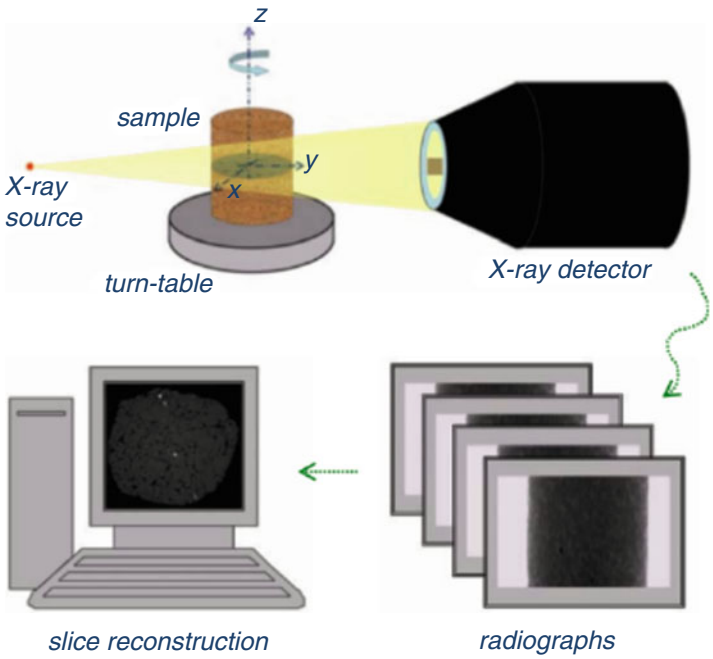


Fig. 6.2 Main parts of CT scan (From Favretto 2008)

CT scanning has provided an alternative for observing intact soil structure (Anderson et al. 1988; Warner et al. 1989; Grevers and de Jong 1994; Peyton et al. 1994; Perret et al. 1997, 1998, 1999, 2003; Rasiah and Aylmore 1998a, b; Rogasik et al. 1999; Gantzer and Anderson 2002; Pierret et al. 2002; Anderson et al. 2003; Rachmant et al. 2005; Gibson et al. 2006). The principal benefits of CT techniques are reducing the physical impact of sampling, providing three-dimensional (3D) information (analysis of images), and allowing rapid scanning to study sample dynamics in near real time (Rasiah and Aylmore 1998b; Rasband 2006). Elliot and Heck (2007a) have compared the void set determination from thin section samples imaged for optical and from CT image acquisition on identical region. They concluded that CT identified a category of void often overlooked by optical imagery classification and therefore would make an excellent complementary technique to soil micromorphology.

In this chapter we will focus on 2D soil images. In general, 2D images from stacked slices that the CT scan captures are analyzed to reconstruct the 3D information with the aim to simplify the analysis. Common concepts are highlighted and explained.

6.2.3 *Image Binarization*

Segmentation is one of the most critical steps in processing images prior to analysis. Segmentation techniques are generally used to define the pore space in soil thin sections and CT images (Elliot and Heck 2007b). In case we want to distinguish between two different objects (object and background), we will usually talk about binarization instead of segmentation. Binarization techniques are mainly based on thresholding methods (Sezgin and Sankur 2004) in which global or local thresholds are calculated to separate pore space (object) from solid space (background). Global thresholding calculates a unique threshold which is applied to the whole image. By contrast, local thresholding adapts the threshold depending on local image characteristics, and so, values change over the image.

Thresholding methods follow different strategies to calculate an optimal threshold. Sezgin and Sankur (2004) categorize thresholding in six groups: histogram shape-based methods, clustering-based methods, entropy-based methods, object attribute-based methods, spatial methods, and local methods (the first five groups define global thresholding methods). The most well-known clustering-based methods are the iterative method (Ridler and Calvard 1978) and Otsu's method (Otsu 1979). Entropy-based methods are maximum entropy (Kapur et al. 1985), maximum Renyi's entropy (Sahoo et al. 1997), and minimum cross entropy (Li and Lee 1993). The most well-known local thresholding method is indicator kriging (Oh and Lindquist 1999) based on local spatial correlations. Recently, a new version of indicator kriging called "adaptive-window indicator kriging" (Houston et al. 2013a) has become available that improves significantly the segmentation results. Some other local strategies followed, e.g., extending Otsu's method by minimizing grayscale intra-class variance (Hapca et al. 2013) to obtain better performance when compared to the original method.

At this time, a lack of general agreement on the most appropriate pore-solid soil images threshold remains. Figure 6.3 shows an original soil thin section and several threshold methods results to show their possible differences.

6.3 Mathematical Morphology

The development of computer technologies had a positive effect on image acquisition and processing, but also a profound effect in the development of mathematical techniques that are known as mathematical morphology (MM). The central idea of MM is to examine the geometrical structure of an image by matching it with small patterns (structuring elements) at various locations in the image. By varying the size and the shape of the structuring elements, one can extract useful information. The seminal work of Matheron and Serra in 1964 (Serra 1982) quantified the mineral characteristics from thin cross sections applying new concepts in integral geometry and topology. MM was originally developed for binary images and was extended later to grayscale images.

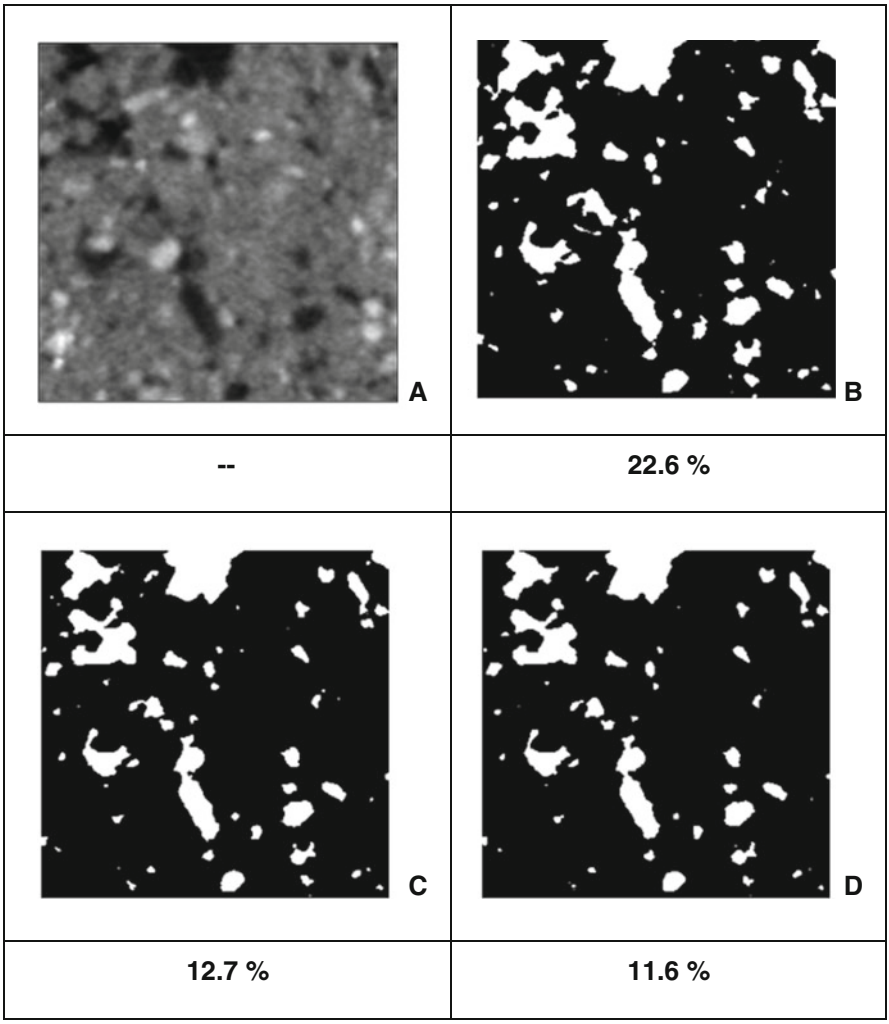


Fig. 6.3 2D soil thin sections (**a**) and the binary image and porosity obtained applying Otsu's method (**b**), maximum entropy (**c**), and Shanbhag method (**d**). Thresholding using ImageJ. *White* pixels are the pores and *black* pixels are the void space

The application of MM for the analysis of soil structure in two dimensions has been presented in several works (2D soil images) that are summarized here. Moran et al. (1988) evaluated mathematically the macropore structure to evaluate the long-term effect of several tillage systems. Moran and McBratney (1992a) presented a method to analyze pore structure with three components: field-impregnated pore space, laboratory-impregnated pore space, and soil solid. They applied this method to study the total and connected macroporosity on several trials with different soil management (Moran and McBratney 1992b). Dullien (1992) studied the relationship between soil pore structure and fluid flow phenomena using these

techniques on 2D soil thin sections. Moran (1994) made an interesting review on the use of digital images to study the form of soil features. Ringrose-Voase (1996) introduced several measurements to characterize soil pore structure such as area, perimeter density, and numerical density. He showed examples of how to apply these measures in 2D images and an adaptation to use in a 1D dimension. Horgan (1998) reviews the basic MM ideas and then applies it to studies of the size distribution of soil pores, the lengths and geometric patterns of cracks in drying soil, and the growth of fungal hyphae. He shows how gray-level images can be handled as binary images in three dimensions. Stewart et al. (1999) showed how the K function could be applied to determinate the spatial relation between roots and macropores.

Later on, studies by Mecke and Arns (2005), Lehmann (2005), and Lehmann et al. (2006), among others, have been applying MM in 3D soil images. All of them pursue the linkage between geometrical parameters with flow parameters as an attempt to relate structure and functioning in soil.

There are several applications of MM tools. However, here we will focus our attention on Minkowski functionals applied in binary images.

6.3.1 Minkowski Functionals

One suitable approach to describe the distribution of elements within structures is Minkowski functionals (MFs), a family of topological and geometrical descriptors belonging to methods of integral geometry (Michielsen and De Raedt 2001).They are invariant under rigid motions and additive. MFs embody information from every order of the correlation functions, are numerically robust even for small samples, and yield global as well as local morphological information (Kerscher et al. 2001).

Within a two-dimensional binary image, three MFs can be defined: covered area (A), boundary length (B), and Euler characteristic. Of particular interest is the latter parameter that contains information related to pattern connectivity as is normally denoted as Euler-Poincarè characteristic (EPC). The MFs can be calculated following the relations:

$$\begin{aligned} A &= n_b \\ B &= -4n_b + 2n_e \\ \mathrm{EPC} &= n_b - n_e + n_v \end{aligned} \tag{6.1}$$

where n_b is the number of boxes (pixels), n_e is the number of edges, and n_v is the number of vertices, with the common edges and vertices of connected pixels counted only once. Three examples of these MFs are shown in Fig. 6.4.

Mecke (1998) showed that integral geometry provides formulae which make the calculus convenient for many models of stochastic geometries. He points out an example using the Boolean grain model and applied morphological measures in the context of percolation.

In three dimensions, MFs correspond to the enclosed volume (V), surface area (S), mean breadth (B), and connectivity (EPC). To compute these functionals of the image of the pore space, it is necessary to count the number of vertices (n_v),

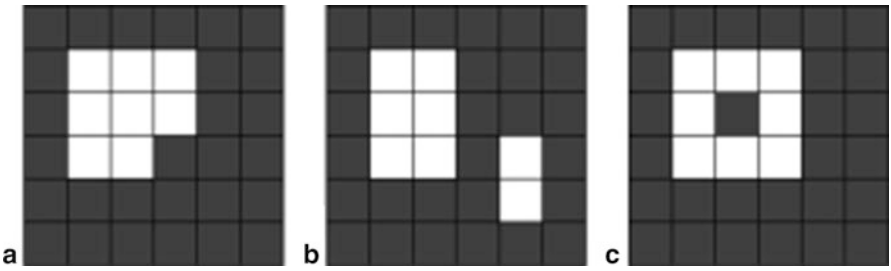


Fig. 6.4 The 2D MFs: area (**a**), boundary length (**b**), and Euler-Poincarè characteristic (EPC). Considering the white pixels as pores and the black pixels void, we have in (**a**) $A = 8$, $B = 12$, and $EPC = 1$; (**b**) $A = 8$, $B = 16$, and $EPC = 2$; and (**c**) $A = 8$, $B = 16$, and $EPC = 0$ (Adapted from Larkin et al. 2014)

faces (n_f), and cubes (n_c) of all voxels of the pore space object in the 3D image (Michielsen and De Raedt 2001):

$$\begin{aligned} V &= n_c \\ S &= -6n_c + 2n_f \\ 2B &= 3n_c - 2n_f + n_c \\ \text{EPC} &= -n_c + n_f - n_c + n_v \end{aligned} \tag{6.2}$$

Vogel and Kretzschmar (1996) estimated EPC analyzing pairs of parallel images (dissectors). This was the first time that EPC was applied to characterize the pore space in soils. Vogel et al. (2005) provided a quantitative description through MFs to explain the crack patterns at the soil surface and their dynamics. In this work they also employed MM tools such as dilations and erosions.

6.3.2 Minkowski Fractal Analysis

In the next section, we will talk about soil pore fractal geometry. However, we prefer to include this concept here to point out how the scaling behavior of soil binary images has been studied with different subdivision techniques and not only the well-known box-counting method mentioned in the next section.

Minkowski fractal dimensions (DS) of binary images can be calculated following the Bouligad-Minkowski protocol (Russ 1995). The method consists of producing a series of images displaying only the boundaries of the structure. Boundary images are formed by adding the complementary part of a dilated image (following the "OR" Boolean logic) to an eroded image of the same picture and gradually increasing dilatation/erosion degrees. The effective boundary width (EBW) is defined as the area of the boundary divided by the dilation/erosion degree and the number of iterative cycles.

To calculate DS, the EBW is calculated for each boundary image and plotted vs the number of dilation/erosion cycles on a log-log scale. For isotropic dilation, the circle diameters may be directly used as the scaling parameter (Dathe et al. 2001).

Minkowski fractal dimensions are calculated from the slope of the curves as DS = 1–slope.

Dathe et al. (2001) found deviation from linearity dividing it in two linear regressions obtaining two DS values: textural (DS1) and structural (DS2) fractal dimensions. The crossing point (CP) between both lines was determined minimizing the combined mean quadratic error of both regressions.

6.4 Pore Soil Fractal Geometry

Fractal geometry has been increasingly applied to quantify soil structure using fractal parameters due to the complexity of the soil structure, and thanks to the advances in computer technology (Tarquis et al. 2003 and references therein). The value of fractal parameters can be derived through indirect methods, such as water retention curves (Crawford et al. 1995), or direct ones through image analysis.

For many years 2D images of soil thin sections have been used in several works to describe the spatial structure extracting the mass fractal dimension (D), pore-solid interface, and lacunarity (Brakensiek et al. 1992; Pachepsky 1996; Giménez et al. 1997, 1998; Oleschko et al. 1998a, b; Bartoli et al. 1999; Dathe et al. 2001; Dathe and Thulner 2005).

6.4.1 Box-Counting Mass Dimension

We consider a binary image displaying two phases – pore and solid. In executing a fractal analysis, one of the two phases for analysis needs to be chosen. Without the loss of generality, we shall here choose the pore phase.

The image is covered by a grid of boxes of size δ. The number of boxes which cover the pore phase is recorded. This is repeated for different size boxes $N(\delta)$. For a fractal set

$$N(\delta) \propto \delta^{-D} \tag{6.3}$$

A log-log plot of $N(\delta)$ against δ then yields a line of slope equal to –D. This is named as fractal dimension or mass dimension in which the box-counting (BC) algorithm is used. As an illustrative example, in Fig. 6.5, we show the calculated mass fractal dimension of the binary images presented previously in Fig. 6.3.

Given the relation in Eq. 6.3, it is now instructive to seek bounds on the values the function $N(\delta)$ can take (Bird et al. 2006). In doing so, we look to see what behavior we can expect from an arbitrary image including or not including any element of self-similar scaling. We denote the fraction of the image occupied by pore phase with f (porosity). If $f = 1$, the number of boxes required to cover the set is $\left(\frac{L}{\delta}\right)^2$, and this is trivially an upper bound for $N(\delta)$ when f <1. In order to derive a lower bound, we consider the situation in which the boxes cover the set, but no part of the

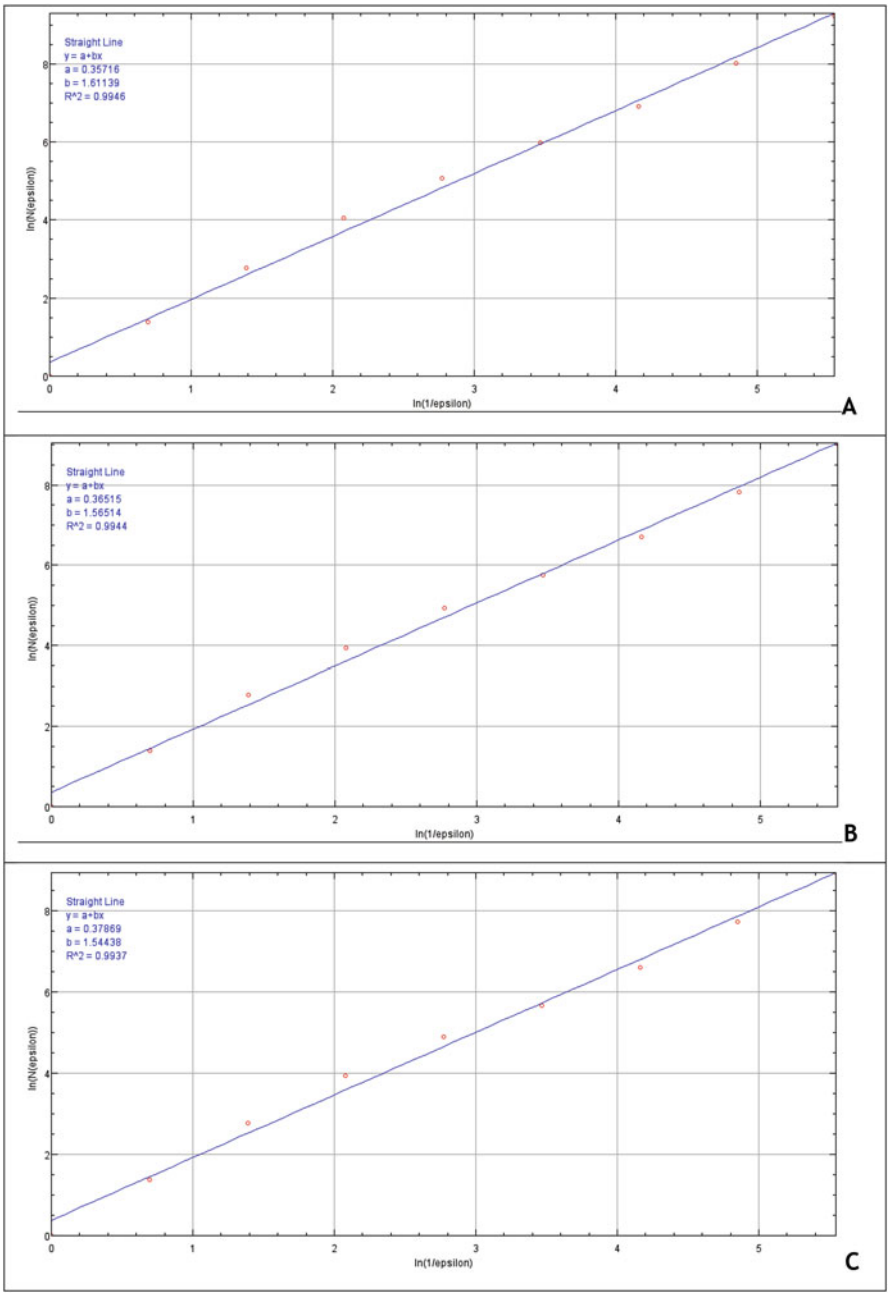


Fig. 6.5 Mass fractal dimension (slope b from the *straight line*) calculated on a binary image obtained applying Otsu (**a**), maximum entropy (**b**), and Shanbhag method (**c**). Bi-log plot and calculations have been made with the plug-in for ImageJ "Multifractal Soil Analysis" (Torre et al. 2014)

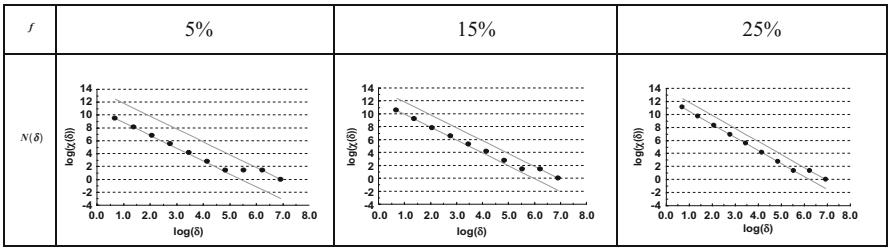


Fig. 6.6 Boundaries of the $N(\delta)$ function in 2D image case at different porosities (f)

complementary set. If this were to occur, then the number of boxes is equal to $\left(\frac{L}{\delta}\right)^2 f$. In general, the boxes will cover part of the complementary set, and consequently the former number of boxes is a lower bound for $N(\delta)$. We may now write that $N(\delta)$ must satisfy the inequalities:

$$\left(\frac{L}{\delta}\right)^2 f \leq N(\delta) \leq \left(\frac{L}{\delta}\right)^2 \tag{6.4}$$

In terms of the log-log plot used to extract a fractal dimension, these inequalities result in two parallel lines of slope – 2, between which the measured data must lie. The vertical spacing between these lines is equal to $\log(f)$. For all images sharing a common value of f, the box-counting data will lie between these bounding lines. Figure 6.6 presents a graphical example of how these two parallel lines reduce the distance as the porosity of an image increases.

We shall consider presenting the data in an alternative way. Namely, we may transform the data by expressing $N(\delta)$ relative to the result obtained for a uniform distribution of pixels. That is, we compute $N'(\delta)$ given by

$$N'(\delta) = N(\delta)\left(\frac{L}{\delta}\right)^{-2} \tag{6.5}$$

This yields a measure of the area of the boxes used to cover the pore space. This is bounded by the inequalities:

$$f \leq N'(\delta) \leq 1 \tag{6.6}$$

For a fractal pore space, this must still produce a power-law relationship. In terms of a log-log plot, we now look for a linear trend in the data with slope equal to 2D constrained to lie between 0 and $\log(f)$. A graphical example of bi-log plots of $N(\delta)$ and $N'(\delta)$ is given in Fig. 6.7.

As we can see, this fractal approach can be applied in a binary image to void or pore sets. The question of fractal nature of the void set and/or the complement set was discussed by Crawford and Matsui (1996). Recently, Perfect and Donelly

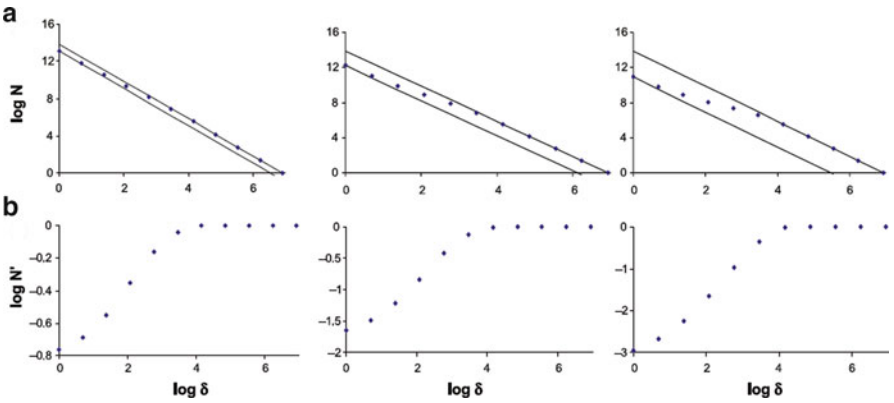


Fig. 6.7 Bi-log plots of the $N(\delta)$ function (**a**) and the $N'(\delta)$ function (**b**) from three different 2D soil thin sections presenting different porosity: *left*, 46.67 %; *center*, 19.17 %; and *right*, 5.25 %. *Solid lines* correspond to the upper and lower bound of $N(\delta)$ function (Figure adapted from Bird et al. (2006))

(2015) have introduced a new method ("biphase box counting") for discriminating between void fractal, pore fractal, and Euclidean scaling in images that display apparent two-phase fractal behavior when analyzed using the traditional method. It consists of box counting the selected phase and its complement conjointly. They present several examples of fitting both datasets conjointly to fractal and/or Euclidean scaling relations.

A key observation regarding this analysis is that a box is counted regardless of the pore area contained within as long as there is at least one bit or pixel representing pore space. But what happens in the case when there is variability in the local porosity? This question can be addressed using a multifractal analysis (MFA) which we describe in Sect. 6.5.

6.4.2 Pore-Solid Fractal Model

Once that the concept of fractal dimension has been described, and the restrictive bounds that it presents, we can summarize the two forms of mass fractal that were proposed yielding opposing scaling behavior. The solid (void) mass fractal describes a porous material with a density which decreases and a porosity which increases with increasing sample size. In particular, mass and density scale as power laws. This model was used as a descriptor of soil aggregates (Young and Crawford 1991; Rieu and Sposito 1991; Anderson and McBratney 1995). Conversely, the pore mass fractal describes a porous material with a density which increases and a porosity which decreases with increasing sample size. In this case, the porosity scales as a power law. The solid mass fractal exhibits a power-law pore-size distribution,

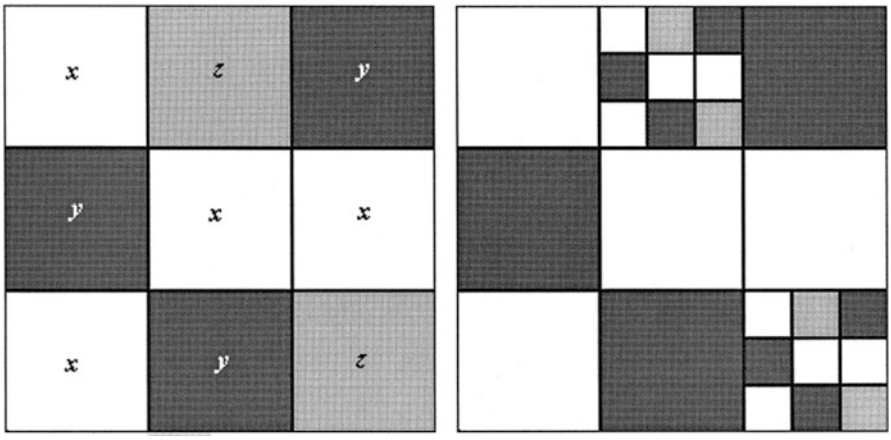


Fig. 6.8 Pore-solid fractal (PSF) model in 2D (After Perrier and Bird 2002): *x*, solid; *y*, pore; *z*, part that continues subdividing

whereas the pore mass fractal exhibits a power-law particle-size distribution, but neither can represent both solid and void scaling distributions. The two models need to account for a lower cutoff of scale since they are seen to fail immediately if this lower bound is not present, yielding porosities of one and zero, respectively.

The pore-solid fractal (PSF) model (Perrier et al. 1999; Bird et al. 2000) of soil structure is an extension and generalization of the fractal approach to modeling soil structure, in which a range of particle sizes and a range of pore sizes are incorporated in a common geometric model (see Fig. 6.8). Solid and pore mass fractal models appear as special cases of the PSF model. Besides it, PSF can be used to model several other scaling properties such as fractal pore-solid interfaces (Perrier et al. 1999) as well as fractal pore-size distributions (Perrier et al. 1996) or observed distributions of aggregates in a fragmentation process (Perrier and Bird 2002).

Several authors have found that the general expression for the retention curve derived from the PSF model (Bird et al. 2000) is quite accurate with experimental data although some parameters should have a previous estimation (Fallico et al. 2010; Behzad et al. 2011; Ghanbarian-Alavijeh et al. 2011).

6.4.3 Lacunarity

Mandelbrot (1983) pointed out that fractal dimensions would not suffice to provide a satisfactory description of the geometry of lacunar fractals and that at least one other parameter, which he termed "lacunarity," is necessary. He introduced the concept of lacunarity as a measure of the distribution of gap sizes in a given geometric object, and later on a more precise definition was introduced by Gefen et al. (1983). He viewed it as a measure of the deviation of a geometric object from translational

invariance. In a low lacunarity object, where gap sizes are relatively homogeneous, different regions of the object tend to be similar to each other. In contrast, in a high lacunarity object, different regions may be very dissimilar and cannot be made to coincide by simple translation. This characterization is highly scale dependent; objects that are heterogeneous at small scales can be homogeneous when examined at larger scales or vice versa (Gefen 1983).

A commonly used method for calculating the lacunarity in binary images was outlined by Allain and Cloitre (1991) based on "gliding boxes" (GB) of increasing sizes. A box of side length l is placed in the upper left-hand corner of an image of side length L (being $1 \leq L$). The calculation algorithm records the number or "mass" (m) of pixels that are associated with the solid underneath the moving box. The box is then translated by one pixel to the right, and the underlying mass is again recorded. When the moving box reaches the right side of the image, it is moved back to its starting point at the left side of the image and is translated by one pixel downward. The calculation proceeds in this fashion until the box reaches the lower right-hand corner of the image, at which point it has explored every one of its possible positions on the image ($(L-1+1)^2$). Then, the number $n(m, l)$ of times a particular value of mass (m) has been recorded with the moving box of side length l is computed. Division of $n(m, l)$ by the total number $(L-1+1)^2$ of possible positions of the moving box above the image yields the probability distribution function ($Q(m, l)$):

$$Q(m,l) = \frac{n(m,l)}{(L-l+1)^2} \tag{6.7}$$

The statistical moments ($Z_Q(q,l)$) of $Q(m,l)$ are defined as

$$Z_Q(q,l) = \sum_{m=0}^{l^2} m^q Q(m,l) \tag{6.8}$$

Based on the first and second moment (q =1 and 1 = 2, respectively), the lacunarity ($\Lambda_{GB}(1)$) with a box size length l is defined as

$$\Lambda_{GB}(l) = \frac{Z_Q(2,l)}{\left(Z_Q(1,l)\right)^2} \tag{6.9}$$

For a translationally invariant set, $Z_Q(2,l) = (Z_Q(1,l))^2$, and then $\Lambda_{GB} = 1$ and is independent of l (Allain and Cloitre 1991). Equation 6.9 can be rewritten in a more clear and common used expression taking into account that $Z_Q(1,l) = \mu_Q$ and $Z_Q(2,l) = \mu_Q{}^2 + \sigma_Q{}^2$:

$$\Lambda_{GB}(l) = \frac{\sigma_Q{}^2}{\mu_Q{}^2} + 1 \tag{6.10}$$

Equation 6.8 shows that $\Lambda_{GB}(1)$ is mainly a measure of the width of the distribution $Q(m, l)$. In Fig. 6.9 we show an example of lacunarity functions of the binary soil images obtained in Fig. 6.3.

Several authors have pointed out that the joint distribution of lacunarity and other fractal parameters can better discriminate soil architecture (Chun et al. 2008; Zamora-Castro et al. 2008; Luo and Lin 2009; Cumbrera et al. 2012).

Lacunarity is not confined to binary configurations, but can also be used with grayscale data (Plotnick et al. 1996). Voss (1986) proposed a probability approach to estimate the fractal dimension and lacunarity of image intensity surfaces. Let us begin to define $P(m, l)$ as the probability that there are m intensity points within a box size of l centered about an arbitrary point in an image. Intensity points are referred to as the number of points filled in a cube box. Hence, we have

$$\sum_{m=1}^{N} P(m, l) = 1 \tag{6.11}$$

where N is the number of possible points in the box of l. Suppose that the total number of points in the image is M. If one overlays the image with boxes of side l, then the number of boxes with m points inside the box is $\left(\frac{M}{m}\right) P(m, l)$. Hence,

$$M(l) = \sum_{m=1}^{N} \mathrm{mP}(m, l) \tag{6.12}$$

and

$$M^2(l) = \sum_{m=1}^{N} m^2 P(m, l) \tag{6.13}$$

Lacunarity can be computed from the same probability distribution $P(m, l)$. Hence, lacunarity ($\Lambda(1)$)is defined as

$$\Lambda(1) = \frac{M^2(l)}{(M(l))^2} + 1 \tag{6.14}$$

This type of analysis hasn't been so popular in soil science. However, several authors have begun to use it and relate it with other fractal dimensions in grayscale images, focusing on soil grayscale images (Roy and Perfect 2014).

6.4.4 Configuration Entropy and Local Porosity

The first works found in the literature using this parameter are found in planar thin sections of gold and sandstone to characterize the spatial patterns of binary images (Andraud et al. 1994, 1997). It was until much later, however, that this concept of configuration entropy began to be applied to pore-soil images. Tarquis et al.

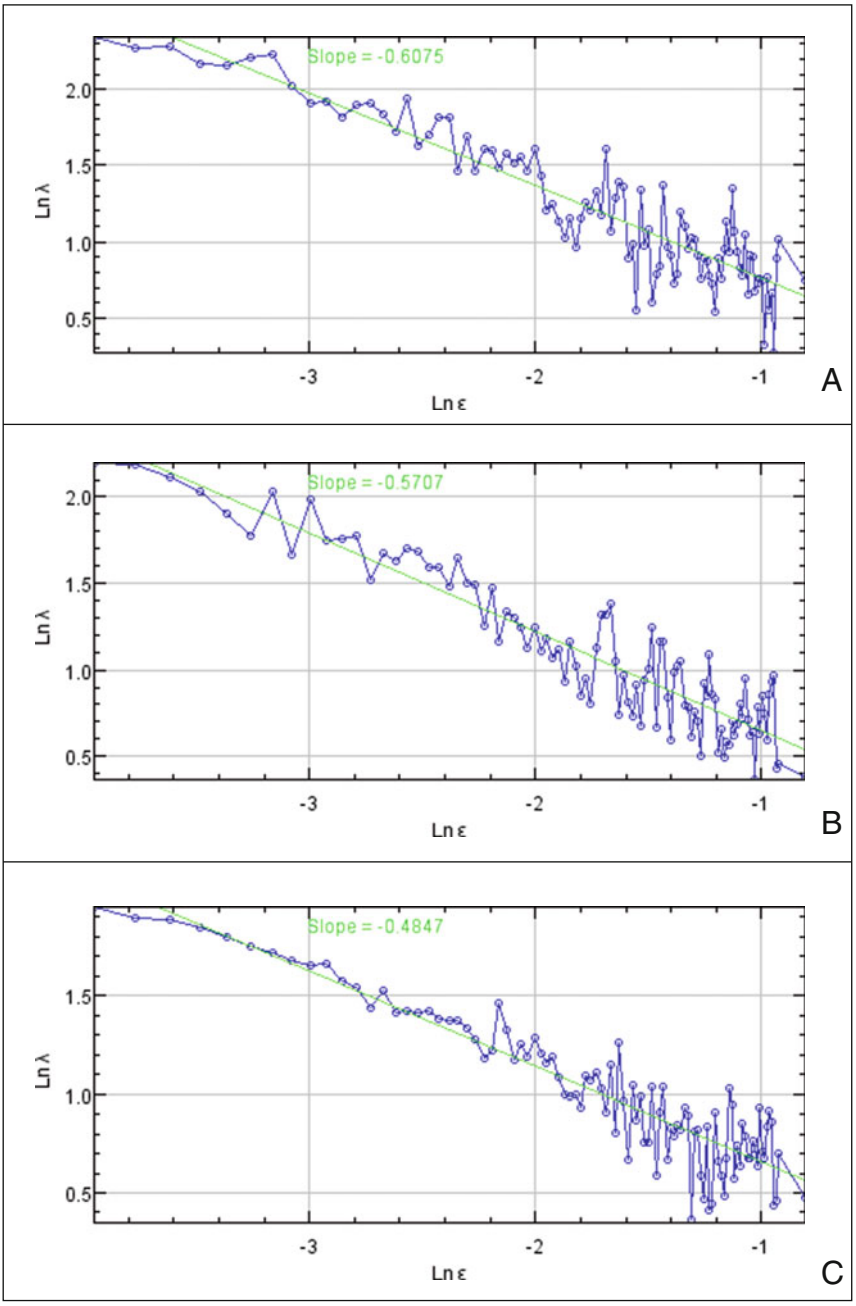


Fig. 6.9 Lacunarity function (bi-log plot of $Ln(\lambda)$ vs $Ln(\varepsilon)$) calculated on binary images obtained applying Otsu (**a**), maximum entropy (**b**) and Shanbhag method (**c**). Bi-log plot and calculations have been made with the plug-in for ImageJ "FracLac" (Karperein 2013)

(2003) illustrated the method showing three images of soil samples representing contrasting soil void patterns. These gray images were converted to binary images by visually comparing the derived binary image with the original gray image. The normalized configuration entropies (H*(l)) were calculated at different l on the binary images following the methodology outlined in Andraud et al. (1994). The authors showed that there is a range of l values where H*(l) reaches a maximum value after which it decreases and tends to stabilize around 0.2 (Tarquis et al. 2003). The maximum values of the normalized configuration entropy $H^*(l_c)$ and of the characteristic length or the entropy length (l_c) were different for each image, and their shapes were similar to the ones that were obtained from an image created by simulating a Poisson grain distribution model (Andraud et al. 1997).

The basis of this method is to study the effect of scale in porosity. Estimation of porosity from a binary image implies counting the pixels representing pores and expressing this count as a percentage of the total number of pixels in the image. If an image is divided in an arbitrary number of smaller areas (e.g., boxes) and porosity is estimated in each subarea, a distribution of the measure "porosity" is obtained for the image. The basic idea of local porosity distributions (local is related to each box the image was divided to) is to turn a global measure into a distribution of local measures (Beghdadi et al. 1993).

An image is formed by pixels arranged in a square lattice of side L. A distribution of local porosity is obtained if this lattice is subdivided in $n(l)$ boxes of size l, from $l = 1$ to $l = L/4$. In every box the number of pixels belonging to the pore class, N_j, is recorded.

The probability associated with a case of j pore pixels in a box of size $l(p_j(l))$ is defined as

$$p_j(l) = \frac{N_j(l)}{n(l)} \tag{6.15}$$

where $N_j(l)$ is the number of boxes with j pore pixels and $n(l)$ is the number of boxes of size l. In this probability function, the value $j = 0$ has a meaning and should be considered. The configuration entropy, H, (Andraud et al. 1994) is defined as

$$H(l) = -\sum_{\mathrm{j}=0}^{\delta x \partial} \mathrm{p_j}(l) \ \log\left(\mathrm{p_j}(l)\right) \tag{6.16}$$

and measures the uncertainty associated with the porosity that can be attained by a set of boxes of size l. Andraud et al. (1997) provided a rigorous connection between the configuration entropy $H(l)$ and the local porosity concept.

Since the underlying probability changes with the number of pixels inside the box (lxl), $H(l)$ needs to be normalized for comparing entropy values corresponding to different sizes l. This is done through (Andraud et al. 1994)

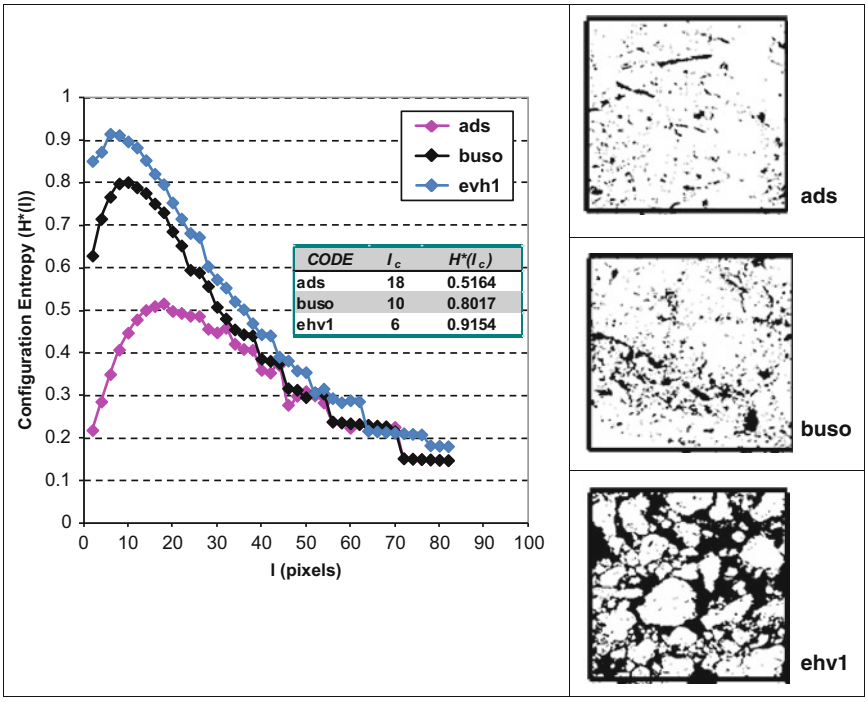


Fig. 6.10 Configuration entropy (*left column*) from three different binary soil images: ads, buso, and evh1. The inner table shows the characteristic length (l_c) and the maximum configuration entropy $H^*(l_c)$ of each soil image

$$H^*(l) = \frac{H(l)}{H_{\max}(l)} \tag{6.17}$$

where $H_{\max}(l) = \log(l^2 + 1)$.

$H^*(l)$ is called the normalized configuration entropy of the two-dimensional morphology of the binary image. An example of these calculations is shown in Fig. 6.10. From this plot two parameters can be obtained, the characteristic length (l_c) and the maximum configuration entropy ($H^*(l_c)$).

Several authors have applied those as a useful morphology parameter (Tarquis et al. 2006; Dathe et al. 2006; Chun et al. 2008).

6.5 Pore Multiscale Analysis

In the mass dimension, a box is counted regardless of the pore area contained, which is a too coarse approach when local porosity shows certain variability. The approach of presence or absence of pores can therefore not be used, and to overcome

this, multifractal analysis (MFA) is conducted. This type of analysis has been increasingly applied to quantify soil structure (Posadas et al. 2003; Bird et al. 2006; Dathe et al. 2006 among others) and is becoming increasingly popular.

MFA in 2D images involves partitioning the plane into boxes to construct samples with multiple scales. The box-counting (BC) method combines pixels to form larger, mutually exclusive boxes each containing a different set of pixels. Given an $L \times L$-pixel image partitioned to a box size of $\delta \times \delta$, the number of boxes that will follow the proportion of line size δ ($n(\delta)$) is

$$n(\delta) \propto \left(\frac{L}{\delta}\right)^2 \tag{6.18}$$

If δ is overly large, the resulting number of samples may not be sufficient for statistical analysis.

The gliding box (GB) method was originally used in lacunarity analysis (Allain and Cloitre 1991) and later modified by Cheng (1997a, b) for application to MFA. The GB method essentially constructs samples by gliding a box of a certain line size (δ) over the grid map in all possible directions. An "up-scaling" partitioning process begins with a minimum line size box ($\delta_{\min}$), which is steadily enlarged to a specific size smaller than L. The proportionality of the number of boxes of linear size δ ($n^*(\delta)$) is

$$n*(\delta) \propto \left(\frac{L-\delta+\delta_{\min}}{\delta_{\min}}\right)^2 \text{ or } n*(\delta) \propto \left(\frac{L}{\delta_{\min}} - \frac{\delta}{\delta_{\min}} + 1\right)^2 \tag{6.19}$$

where $\delta_{\min} \leq \delta \leq L$. This relationship is often expressed as

$$n*(\delta) \propto \left(\frac{L}{\delta_{\min}} - \varepsilon + 1\right)^2 \tag{6.20}$$

where ε is the ratio between the box line size and the minimum size chosen $\left(\frac{\delta}{\delta_{\min}}\right)$.

One advantage of the GB method is that, generally, the larger the sample size, the better the statistical results. Because this partitioning overlaps, the measures defined by these boxes are not statistically independent, and the definition of the measure in gliding boxes differs, as explained below. This type of analysis has been applied in binary images in different fields (Grau et al. 2006; Saa et al. 2007; Tarquis et al. 2007).

6.5.1 Box-Counting Algorithm

The generalized dimension calculated using the BC technique essentially reflects the mass contained in each box. An image is divided into $n(\delta)$ boxes, and the fraction of pore space (μ_i) in each box is calculated from

$$\mu_i = \frac{m_i}{m_T} = \frac{m_i}{\sum_{i=1}^{n(\delta)} m_i} \tag{6.21}$$

where m_i is the number of pore class pixels in box i and m_T is the total number of pore class pixels in an image. In this case, the pore space area is the measure and the square grid the support. The generating function ($\chi(q, \delta)$) is then defined as:

$$\chi(q,\delta) = \sum_{i=1}^{n(\delta)} [\chi_i(q,\delta)] \qquad q \in \Re \tag{6.22}$$

where

$$\chi_i(q,\delta) = \mu_i^q = \left(\frac{m_i}{\sum_{i=1}^{n(\delta)} m_i}\right)^q \tag{6.23}$$

χ_i is a weighted measure that represents the percentage of pore space in box i, and q is the weight or moment of the measure. When computing the number of boxes of size δ, the possible values of m_i range from 0 to $\delta \times \delta$. So if $N_j(\delta)$ is the number of boxes containing j pixels of pore space in a given grid, Eqs. 6.5 and 6.6 can then be written as follows (Barnsley et al. 1988):

$$\chi(q,\delta) = \sum_{i=1}^{n(\delta)} (\mu_i)^q = \sum_{i=1}^{n(\delta)} \left(\frac{m_i}{m_T}\right)^q = \sum_{j=1}^{\delta x \delta} N_j(\delta) \left(\frac{j}{m_T}\right)^q \tag{6.24}$$

where $m_T = \sum_{j=1}^{\delta \times \delta} \mathrm{j}\ \mathrm{N}_j(\delta)$

Using the distribution function, $N_j(\delta)$ simplifies calculations and reduces computational errors (Barnsley et al. 1988).

A log-log plot of a self-similar measure, χ(q, δ), vs δ at various values for q gives

$$\chi(q,\delta) \sim \delta^{\tau(q)} \tag{6.25}$$

where $\tau(q)$ is the mass exponent for q (Hentschel and Procaccia 1983; Feder 1989). $\tau(q)$ can be written as

$$\tau(q) = \lim_{\delta \to 0} \frac{\log(\chi(q,\delta))}{\log(\delta)} \tag{6.26}$$

The generalized dimension, D_q, can then be introduced through the following scaling relationship (Hentschel and Procaccia 1983; Feder 1989):

$$D_q = \lim_{\delta \to 0} \frac{\log[\chi(q,\delta)]}{(q-1)\log\delta} \tag{6.27}$$

And, therefore

$$\tau(q) = (q-1)\, D_q \tag{6.28}$$

for $q \neq 1$. The case D_1 is defined as the limit $D_1 = \lim_{q \to 1} D_q$. The dimension D_1 can then be extracted from a plot of entropy ($S(\delta)$) against $\log(\delta)$.

$$D_1 = \lim_{\delta \to 0} \frac{\sum_{i=1}^{n(\delta)} \chi_i\,(1,\delta) \log\left[\chi_i\,(1,\delta)\right]}{\log \delta} \tag{6.29}$$

$$\mathrm{S}\,(\delta) = -\sum_{i=1}^{n(\delta)} \chi_i\,(1,\delta) \log\left(\chi_i\,(1,\delta)\right) \sim D_1 \log\,(\delta) \tag{6.30}$$

The generalized dimensions, Dq, for $q = 0$, $q = 1$, and $q = 2$ are known as the capacity, information, and correlation dimensions, respectively (Hentschel and Proccacia 1983). The capacity dimension is the box-counting or fractal dimension. The information dimension is related to system entropy, whereas the correlation dimension computes the correlation of the measures contained in boxes of various sizes (Posadas et al. 2003).

6.5.2 Gliding Algorithm

As in the BC method, a definition of the mass contained in each box (number of pore pixels) must be given in the gliding box procedure. Let $n^*(m, \delta)$ be the number of gliding boxes of size δ and mass m. Dividing by $n^*(\delta)$ yields the probability function, $\beta(m, \delta)$, for a gliding box of size δ and mass m.

The statistical moment of this distribution is

$$\chi * (q,\delta) = \sum_{m=1}^{\delta \times \delta} \beta\,(m,\delta)\, m^q = \frac{1}{n*(\delta)} \sum_{i=1}^{n^*(\delta)} m_i^q\,(\delta) \approx E\,\{m^q\} \tag{6.31}$$

where $\chi^*(q, \delta)$ is the moment of order q of $\beta(m, \delta)$ and the sum includes all gliding boxes with m_i>0.

The difference between $\chi^*(q, \delta)$ (Eq. 6.29) and $\chi(q, \delta)$is that

$$\chi\,(q,\delta) \approx \left(\frac{L}{\delta}\right)^2 E\,\{m^q\} \tag{6.32}$$

Combining Eqs. 6.13 and 6.14 yields

$$\chi\,(q,\delta) \approx \left(\frac{L}{\delta}\right)^2 \chi * (q,\delta) \tag{6.33}$$

Substituting this relation in Eqs. 6.25 and 6.26 gives

$$\tau(q) = \lim_{\delta \to 0} \frac{\log\left(\chi^*\left(q, \delta\right)\right)}{\log\left(\delta\right)} + \lim_{\delta \to 0} \frac{2\log\left(\frac{L}{\delta}\right)}{\log\left(\delta\right)} \tag{6.34}$$

The second limit is readily solved, giving a value of −2. It is concluded that (Cheng 1999a)

$$\tau(q) = \lim_{\delta \to 0} \frac{\log\left(\chi^*\left(q, \delta\right)\right)}{\log\left(\delta\right)} - 2 \tag{6.35}$$

where 2 is the value of the dimension of the Euclidean plane containing the image.

In the present study, the multiplier method, based on Eq. 6.35 below, was applied to estimate $\tau(q)$ (see Cheng 1999a, b for further details):

$$< \tau(q) > +2 = -\frac{\log\left(< M\left(q, \delta\right) >\right)}{\log\left(\varepsilon\right)} \tag{6.36}$$

where <> stands for statistical moment and M represents the multiplier measured on each pixel as

$$M\left(q, \delta\right) = \left(\frac{\mu\left(\delta_{\min}\right)}{\mu\left(\delta\right)}\right)^q \tag{6.37}$$

The advantage of using Eq. 6.36 compared to Eq. 6.26 is that, since the estimation is independent of box size δ, successive box sizes may be used to estimate $\tau(q)$. The condition to applying the former, however, is that $\mu(\delta_{\min})$ may not be nil, for otherwise the expression $\left(\frac{\mu(\delta_{\min})}{\mu(\delta)}\right)$ is always 0 or undefined for all values of δ.

The resulting estimate may be applied in Eq. 6.26 to estimate D_q.

As an example to compare both algorithms, the generalized dimension of the 2D binary soil images shown in Fig. 6.10 is estimated. Therefore, a comparison between the D_q values obtained with the two methods is given in Fig. 6.11. The two curves are similar for soil ads. The differences between the two methods for soils buso and evh1 are wider for negative values of q, although decay is greater for the positive values.

Finally, an overall comparison of the D_q values point out a much lower standard error in the GB method. The main differences found in the BC method are in the negative q range, whereas in the GB procedure, differences were also observed in the positive q range. For further details, see Grau et al. (2006).

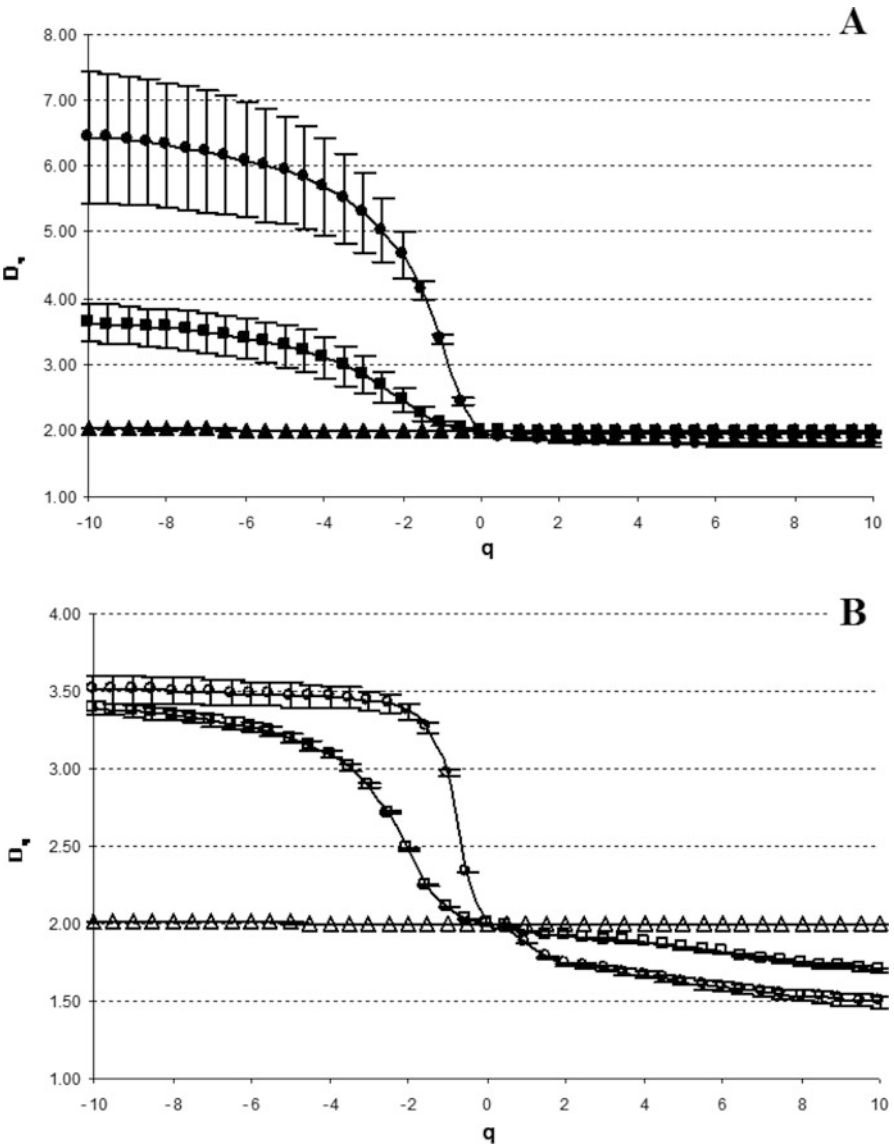


Fig. 6.11 Generalized dimensions (D_q) from $q = -10$ to $q = +10$ for the three soil thin sections (ads, *triangles*; buso, *squares*; and ehv1, *circles*) based on (**a**) the box-counting method and (**b**) gliding box method. The s.e. of D_q values are represented by *line bars* (From Grau et al. 2006)

Table 6.1 Upper and lower bounds of the partition function ($\chi(q, \delta)$) and entropy ($S(\delta)$) based on mass exponent (q), size length of the 2D binary image (L), size length of the box (δ), and porosity percentage (f) (Summarized from Bird et al. 2006)

	$q<0$	$0 \leq q<1$	$q=1$	$1<q$
Bounds	$\chi(q, \delta)$	$\chi(q, \delta)$	$S(\delta)$	$\chi(q, \delta)$
Lower	$\left(\frac{L}{\delta}\right)^{2(1-q)} f^{(1-q)}$	$\left(\frac{L}{\delta}\right)^{2(1-q)} f^{(1-q)}$	$2\ln\left(\frac{L}{\delta}\right) + \ln(f)$	$\left(\frac{L}{\delta}\right)^{2(1-q)}$
Upper	$L^{2(1-q)} f^{(1-q)}$	$\left(\frac{L}{\delta}\right)^{2(1-q)}$	$2\ln\left(\frac{L}{\delta}\right)$	$\left(\frac{L}{\delta}\right)^{2(1-q)} f^{(1-q)}$

6.5.3 Limitations

Given all these definitions, we will see that the partition function presents restrictive bounds. In establishing bounds for $\chi(q, \delta)$ and $S(\delta)$, we must consider five separate ranges of values of the parameter q. All cases are shown in Table 6.1, summarized by Bird et al. (2006).

Having defined these bounds, we now seek to examine their significance in terms of extracting generalized dimensions from image data. For 1<q and for $0 \leq q<1$, the bounding functions when plotted on the log-log plot used to extract the dimension yield two parallel lines with a vertical separation of $(1-q)\ln(f)$. The higher the porosity, the lower the separation between these lines, and consequently a very good straight line fits for the image data.

For $q = 1$, the bounding functions when included on the plot of entropy against $ln(\delta)$ again yield two parallel lines of slope 2 with separation of $ln(f)$. Again, fits to image data lying between these reference lines are obliged to be good. For $q<0$, our bounding functions do not yield parallel lines, and there exists scope for the image data points to deviate from linear behavior within the log-log plot of $\chi(q, \delta)$. For further details, see Bird et al. (2006).

Several authors have pointed out several more restrictions in this type of analysis of binary images. Kravchenko et al. (2009) pointed out that, because their Lipschitz-Hölder exponents are either 0 or 2 at the pixel scale, binary images are not multifractal in a strict mathematical sense. In the same direction, Zhou et al. (2011) contend that the interpretation of the generalized dimensions estimated from binary images should be made very cautiously. These authors indicate that further exploration of the use of grayscale images for multifractal characterization of soil structure should be encouraged.

At this point we would like to recall the work of Buczkowski et al. (1988) in which they comment that the main difficulty in using MFA is that the ideal limit cannot be reached in practice, in this context the pixel.

Observing bi-log plots of the partition function ($\chi(q, \delta)$) in several works (Tarquis et al. 2008, 2009; Perrier et al. 2006), we can observe the existence of a plateau phase at the smallest scales that can be explained by the nature of the measure we are studying. At δ values close to 1, the variation in number of black pixels is based on a few pixels, having the most simplicity when $\delta = 1$ where the measurecan only

have a value of 0 or 1. Thus, for small boxes of size δ, the proportions among their values are mainly constant. However, when the box size passes a certain threshold, a scaling pattern begins.

Several authors have applied the characteristic length (l_c), based on the configuration entropy concept, to determine the minimum size δ that should be used in the MFA (Dathe et al. 2006; Tarquis et al. 2006). At $\delta = l_c$, the distribution of the measure (number of black pixels) is the most uniform we can find in the image. Increasing δ value is when configuration entropy begins to reduce and a scaling pattern can be detected if it is present.

6.6 CT-Scan Image Multiscale Analysis

Taina et al. (2008) did a cursory examination of recent research, involving the application of CT to soil. Their review reveals a considerable diversity of methodology and a lack of standardization. An immediate consequence is the impediment to the comparison of results from different studies. One of the methodologies that presents a lack of general agreement is in the appropriate pore-solid CT threshold (Cortina-Januchs et al. 2011; Wang et al. 2011; Ojeda-Magaña et al. 2014), which is used to obtain a black and white image from the original grayscale data, before calculating any of these scaling parameters. At the same time, the effect of settings of all these processes in scanning and image reconstructions has been deeply studied by Houston et al. (2013b). At this point we will clarify that pore space determined by image analysis is air-filled pore space (hereinafter named pore space). The variety of thresholding methods has already been briefly discussed in Sect. 6.2.3.

Gibson et al. (2006) compared three fractal analytical methods to quantify the heterogeneity within soil aggregates. In this work, the frequency distribution of pore and solid components was clearly dependent on thresholding, which could not be generalized. Tarquis et al. (2008, 2009) point out that a practical problem in the MFA of binary images is that the thresholding methods have a pronounced effect on the porosity and resulting generalized dimensions. It has been suggested to further study grayscale soil images for multifractal characterization of soil structure, avoiding any intermediate thresholding step (Zhou et al. 2010, 2011). This MFA applied in grayscale images is very common in other types of images (Lovejoy et al. 2008; Tarquis et al. 2014). In the next section, we will show some examples of using MFA in gray soil images.

6.7 Box-Counting and Gliding Algorithms

As in the case of binary images, gray images can be analyzed with both algorithms. All the definitions explained in Sect. 6.4 can be applied except that, now our measure is not the number of black pixels representing pores, it is the gray value of the pixel (m_i), and therefore Eq. 6.22 doesn't apply.

Based on the mass exponent function ($\tau(q)$), the singularity index or Hölder exponent (α) can be determined by Legendre transformation of the $\tau(q)$ curve as (Evertsz and Mandelbrot 1992)

$$\alpha(q) = \frac{d\tau(q)}{dq} \tag{6.38}$$

The number of cells of size δ with the same α, $N_\alpha(\delta)$, is related to the cell size as $N_\alpha(\delta) \propto \delta^{-f(\alpha)}$, where $f(\alpha)$ is a scaling exponent of the cells with common α. Parameter $f(\alpha)$ can be calculated as

$$f(\alpha) = q\alpha(q) - \tau(q) \tag{6.39}$$

Multifractal spectrum (MFS), a graph of α vs $f(\alpha)$, quantitatively characterizes variability of the measure studied with asymmetry to the right and left indicating domination of small and large values, respectively. The width of the MF spectrum indicates overall variability (Tarquis et al. 2001). It could be interpreted also as the relationship between a Hausdorff dimension f and an average singularity strength α as implicit functions of the parameter q.

A direct way to calculate the MFS is to use the method of Chhabra and Jensen (1989) following the relations:

$$\left\{ \begin{array}{l} f(q) = \lim\limits_{\delta \to 0} \frac{\sum_{i=1}^{n(\delta)} \eta_{i(q,\delta)} \log(\eta_i(q,\delta))}{\log \delta} \\ \alpha(q) = \lim\limits_{\delta \to 0} \frac{\sum_{i=1}^{n(\delta)} \eta_{i(q,\delta)} \log(\mu_i(\delta))}{\log \delta} \end{array} \right\} \tag{6.40}$$

where $\mu_i(\delta)$is defined in Eq. 6.19 and η_i is

$$\eta_i(q,\delta) = \frac{(\mu_i(\delta))^q}{\sum_{i=1}^{n(\delta)} (\mu_i(\delta))^q} \tag{6.41}$$

One can regard multifractal measures as the union of fractal sets, each one characterized by different scaling exponents α that have been named "singularity exponents." This interpretation is important to understand the concept of singularity maps explained in Sect. 6.6.

From a CT-scan 3D image of a soil aggregate sample, three slices have been selected perpendicular to the XY plane, as shown in Fig. 6.12. For each of the three slices, we will compute, using both algorithms (BC and GB), first the mass function (Fig. 6.13), second the generalized dimension (Fig. 6.14), and third the multifractal spectrum (Fig. 6.15).

Comparing the box-counting method and the gliding method, the errors in $\tau(q)$ (Fig. 6.13) are diminished in the negative q values which is relevant for slice 192. On the other hand, slices 64 and 128 present similar structures; meanwhile 192 is significantly different. At the same time, the curvature of $\tau(q)$ is smoother in the gliding method. However, the differences are retained.

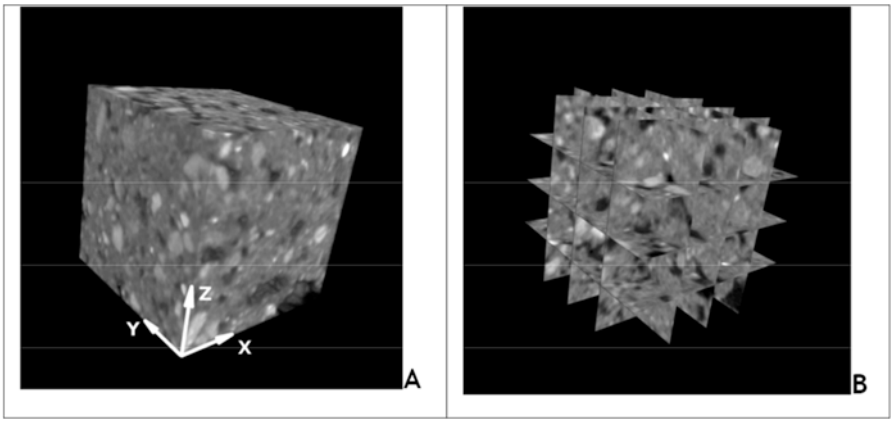


Fig. 6.12 Original CT-scan 3D image of a sample of soil aggregate (**a**) and the selection of three slices in Z direction (**b**)

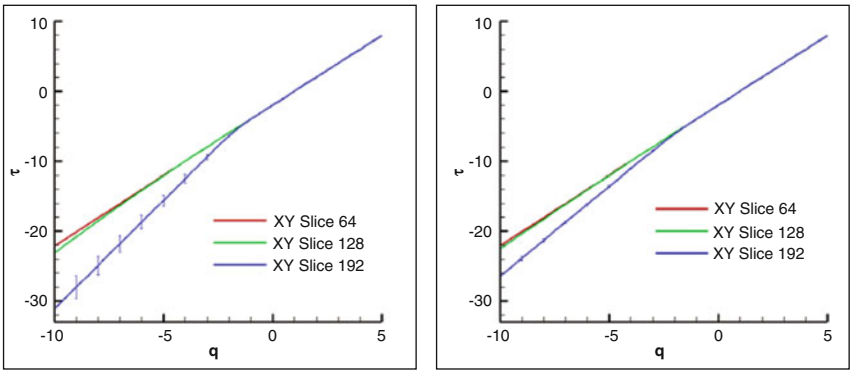


Fig. 6.13 Mass exponent function $\tau(q)$ for each of the three slices selected estimated by *box*-counting method (*left*) and gliding method (*right*) (Adapted from Torre et al. 2016)

$D(q)$, or Dq as named earlier (Fig. 6.14), shows the same results as $\tau(q)$ as both are related by Eq. 6.28. For negative q values, the differences are clearer between slice 128 and slice 64.

The multifractal spectrum (Fig. 6.15) gives us complete information of the hierarchical soil structure. Slice 64 shows a weak multifractal character in concordance with previous results. Slice 128 presents a stronger scaling in the low values that correspond to the right part of the spectrum.

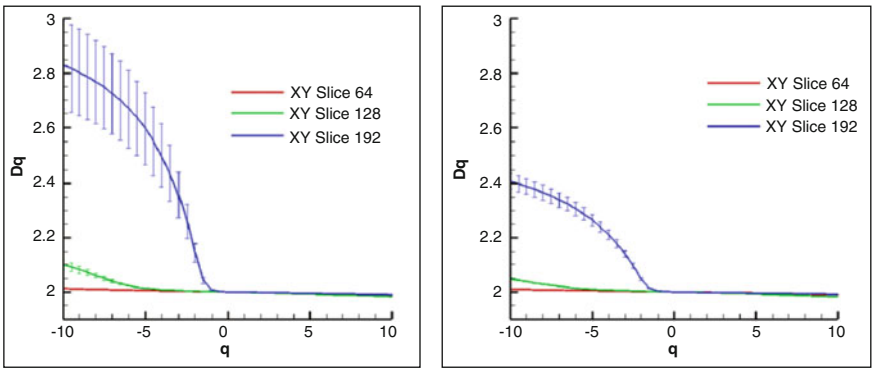


Fig. 6.14 Generalized dimension ($D(q)$) for each of the three slices selected estimated by, *left*, *box*-counting method, and, *right*, gliding method (Adapted from Torre et al. 2016)

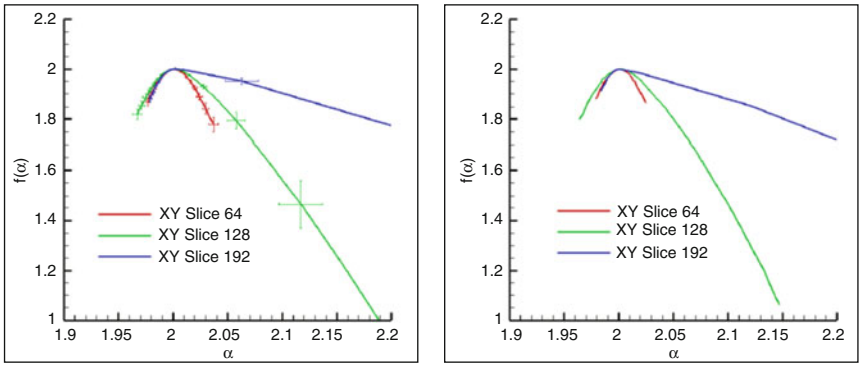


Fig. 6.15 Multifractal spectrum (α vs $f(\alpha)$) for each of the three slices selected estimated by, *left*, *box*-counting method, and, *right*, gliding method (Adapted from Torre et al. 2016)

6.8 Multifractal Sierpinski Carpet

A Sierpinski multifractal can be obtained by a multiplicative cascade or "canonical curdling" process (Perfect et al. 2006) with eight iterations and a multifractal generator. The chosen generator probabilities are $p_1 = 0.1875$, $p_2 = 0.2250$, $p_3 = 0.2875$, and $p_4 = 0.3000$. The order of probabilities, p_i, was randomized before being applied in each iteration.

Grayscale soil images and the multifractal Sierpinski carpet are shown in Fig. 6.16, as well as their respective grayscale histograms. Applying MFA to this synthetic image, we obtain a multifractal spectrum that reminds us of the scaling properties observed already in a CT-scan image used in this chapter. This implies that to create synthetic images that resemble a CT-scan image, one should include scaling properties of the gray values in the methodology.

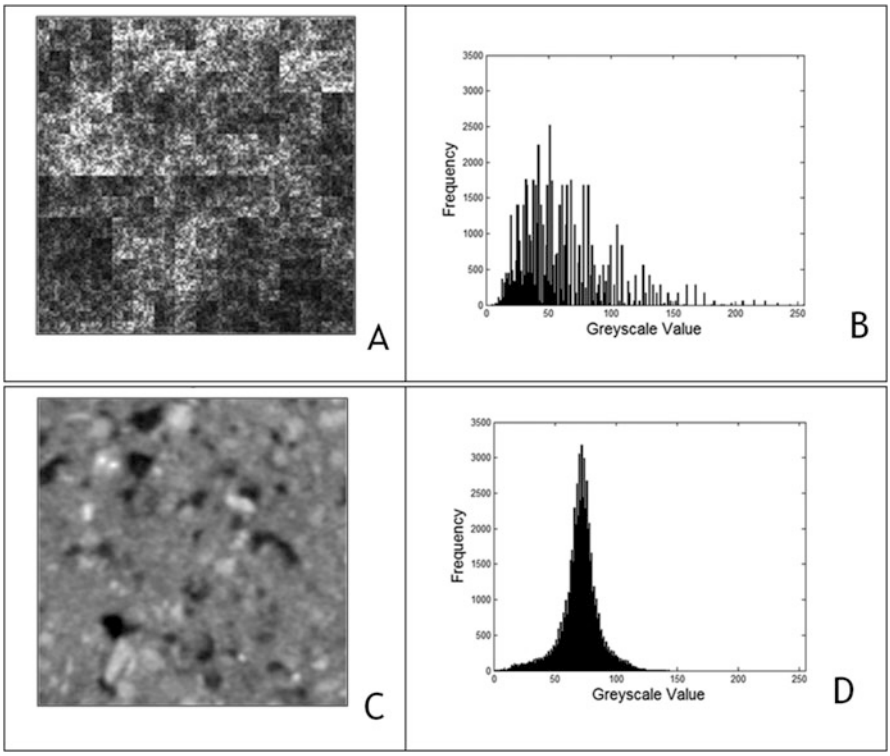


Fig. 6.16 Sierpinski multifractal of 246 × 246 pixels (**a**) and the corresponding gray histogram (**b**). In the second line, slice 128 from the aggregate soil sample (**c**) and the corresponding gray histogram (**d**)

6.9 Singularity Maps and Binarization

A "singularity map" is defined as the locus of the points $x \epsilon$ *support* (μ) with the same Hölder exponent $\alpha(x)$:

$$\alpha = T(x) \tag{6.42}$$

where T(x) is the function that gives us the Hölder exponent of each point x.

To calculate "singularity maps," we will follow Cheng (2001) and Falconer (2003). In multifractal measures the values of the local mass, $\mu(B(x, r))$, calculated for various cell sizes r centered at x, follow a power-law relationship with r:

$$\mu\left(B\left(x, r\right)\right) \sim r^{\alpha(x)} \tag{6.43}$$

where $B(x, r)$ is a ball centered at x with radius r. Locations $\alpha(x) \neq E$, where E is the topological dimension of the measure support, are called "singular locations." Moreover, we can distinguish positive singularities $\alpha(x) < E$ and negative singularities $\alpha(x) > E$. Hölder exponents $\alpha(x)$ can be calculated by the expression:

$$\alpha(x) = \lim_{r \to 0} \frac{\ln \mu\,(B\,(x, r))}{\ln r} \tag{6.44}$$

Cheng (2001, 2007) applied these concepts to a concentration map of a certain mineral. He states that concentration or density mass $\rho(x)$, calculated for various cell sizes r centered at x, obeys a power-law relationship with r:

$$\rho(x) = \frac{\mu\,(B\,(x, r))}{r^E} \sim r^{\alpha(x)-E} \tag{6.45}$$

where $E = 2$ for 2D support and $E = 1$ for 1D support. This power-law is fulfilled in a certain range of r, $[r_{\min}, r_{\max}]$.

If the spatial variability of the exponent $\alpha(x)$ is a constant, then it means we have obtained a monofractal distribution; otherwise, if $\alpha(x)$ has multiple values, then we have got a multifractal distribution.

Positive singularities with $\alpha(x)$ <2 correspond to high values of concentration (enriched values) in a geochemical map, while negative singularities with $\alpha(x)$ <2 correspond to low concentration values (depleted values). Therefore, calculating the "singularity map" for an ore concentration map may be used to characterize concentration patterns (Cheng 2007).

6.9.1 Concentration-Area Method (C-A)

The C-A method (Cheng et al. 1994) establishes power-law relationships between the concentration of a variable and the area enclosed by this concentration:

$$A\,(\rho \geq C) \propto C^{\beta} \tag{6.46}$$

where A($\rho \geq$C) is the area constituted by concentrations (ρ-values) greater than a given value C and β is the fractal dimension of the C-A method.

In the context of fractal/multifractals, concentration is related to singularity exponents through the expression (6.43), so in this case the C-A method tells us

$$A\,(\alpha \geq C) \propto C^{\beta} \tag{6.47}$$

where A($\rho \geq$ C) is the area constituted by singularity exponents (α-values) greater than a given value C and β is the fractal dimension of the C-A method (Liu et al. 2013).

When the singularity distribution follows a fractal model, this power-law relation has only one exponent. Sometimes the C-A method does not meet the simple fractal model because different power-laws apply to the study area. In this case we can assume spatial variability is multifractal instead of fractal, and different slope-change values in the log-log plot can be considered as a separation among different sets. The implications and interpretations in the field of spatial statistics can be found in Cheng (1999b).

6.9.2 *Image Binarization*

The new "singularity C-A" method is mainly based on the existence of self-similar properties in the concentration or density variable, defined as the average intensity of an area centered on the point of interest. When we obtain the cumulative distributions of concentration through the C-A method, these self-similarities are reflected by linear segments (power-law relationships) in a log-log plot. Different exponents resulting from the plot reveal that there are different sets with self-similar properties in the whole set. Values of the exponents are related to geometrical fractal dimensions and singularity exponents (Cheng et al. 1994), and the slope-change points offer us the possibility to define thresholds in all subsets which have appeared. Therefore, the image is segmented according to all the slope-change points. The initial assumption is not necessarily of multifractality, although, as we will see, monofractality falls short in the majority of the examples since several slope-change points are obtained.

The new segmentation method (C-A segmentation) will be applied in the slice 128 as an example. The implementation will follow these steps:

1. The first step is the transformation of the grayscale image to an intensity concentration map. To do so, singularity maps are calculated, where singularity exponents reflect the concentration strength of each point of the image. Firstly, we define a measure μ over the image. The easiest definition of a measure for a square set $S(x, r)$ centered on pixel x and side r is

$$\mu\left(S\left(x,r\right)\right) = \sum_{k \in S(x,r)} I_k \tag{6.48}$$

where I_k is the intensity of pixel k. In the images we are analyzing, the intensity of each pixel is in the range from 0 to 255.

We will use the window-based method (Cheng 2001) to obtain singularity exponents of the image. Using a sliding window (e.g., a square window) of variable sides $r_{\min} = r_1 < r_2 < \ldots < r_n = r_{\max}$, we apply the expression:

$$\alpha(x) = \lim_{r \to r_{\min}} \frac{\ln \mu\left(S\left(x,r\right)\right)}{\ln r} \tag{6.49}$$

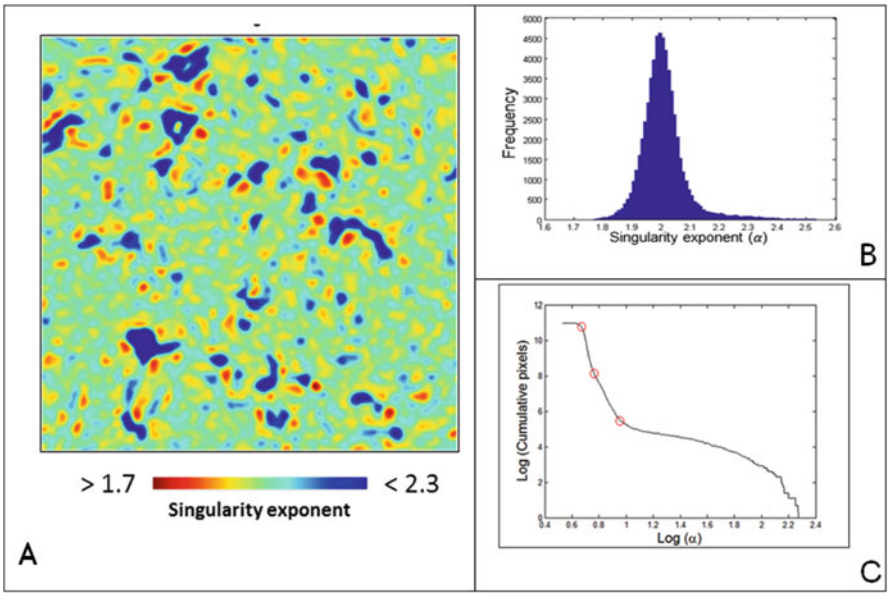


Fig. 6.17 Singularity map of slice 128 presented in Fig. 6.16 section C. (**a**) Hölder exponent (α) histogram from the singularity map (**b**) Log-log cumulative distributions obtained by the "C-A method" (**c**) *Red circles* (in **c**) are slope-change point candidates to be thresholds in the "singularity C-A" method

The side r of the sliding window is always an odd integer number of pixels, and the minimum value is $r_{min} = 1$ *pixel*. The values we have used in the analyzed soil images are $r_{min} = 3$ *pixels* and $r_{max} = 15$ *pixels*. So we have seven values to calculate the slope $\alpha(x)$ by linear regression. The results are shown in Fig. 6.17.

The singularity map gives information about two well-differentiated areas: the area with pixels with α <2 (pixels with convex neighborhood) that we have defined as an enriched concentration area and corresponds to solid bodies in the image and the area with pixels with α >2 (pixels with concave neighborhood) that we have defined as a depleted concentration area corresponding to pore space in the image.

2. The second step consists in applying the C-A method. We calculate cumulative distribution of α-values (singularity exponents) as indicated in Fig. 6.17 section B. We define the area constituted by α-values greater than a given value C as the number of pixels with singularity exponents greater than C:

$$A\,(\alpha \geq C) = N^{0}\,\text{of pixels}\;\;(\alpha \geq C) \tag{6.50}$$

When we obtain log-log plots, linear segments (power-law relationships) appear with their respective slope-change points (see Fig. 6.17 section C). At this point we cannot accurately locate the position of these points. We solve this problem in the following steps:

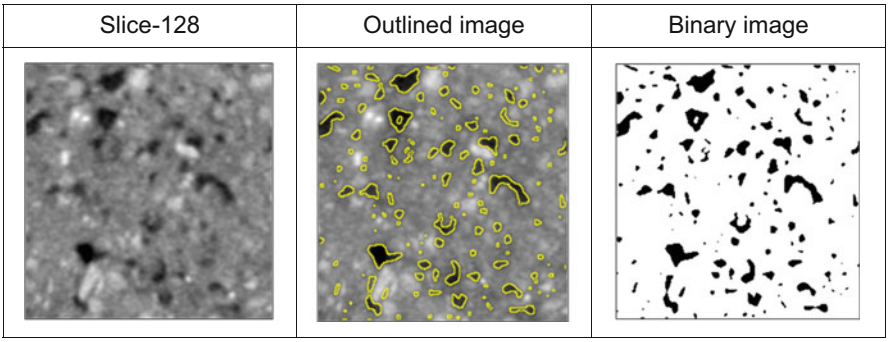


Fig. 6.18 *Left*: grayscale image corresponding to slice 128. *Middle*: overlapping of grayscale and borders of segmented image. *Right*: binarized images after applying the "singularity C-A" method

3. The third step consists in locating slope-change points through the WTMM method; in this case we will use a null second-moment wavelet called the "Mexican hat." This kind of wavelet detects points with local maxima curvature. The WTMM method accurately finds all slope-change points we have obtained through the C-A method (red points drawn in Fig. 6.17 section C).
4. Finally, we choose as the segmentation threshold the first slope-change point with value $\alpha > 2$, i.e., the first value located in the depleted area since this value defines the first set with self-similar properties completely contained in the depleted area, the potential area to be the pore space. In this specific case, the threshold value is $\alpha_c = 1.987$. The binary image obtained is showed in Fig. 6.18. For further details, see Martín-Sotoca et al. (2017).

References

Allain C, Cloitre M (1991) Characterizing the lacunarity of random and deterministic fractal sets. Phys Rev A 44:3552–3558

Anderson AN, McBratney AB (1995) Soil aggregates as mass fractals. Aust J Soil Res 33:757–772

Anderson SH, Gantzer CJ, Boone JM, Tully RJ (1988) Rapid nondestructive bulk density and soil-water content determination by computed tomography. Soil Sci Soc A J 52:35–40

Anderson SH, Wang H, Peyton RL, Gantzer CJ (2003) Estimation of porosity and hydraulic conductivity from x-ray CT-measured solute breakthrough. In: Mees et al (eds) Applications of X-ray computed tomography in the geosciences, Geological Society of London Special Publications 222. Geological Society of London, London, pp 135–149

Baveye P, Boast CW (1998) Fractal geometry, fragmentation processes and the physics of scale-invariance: an introduction. In: Parlange B, Stewart (eds) Fractals in soil science. CRC Press, Boca Raton

Andraud C, Beghdadi A, Lafait J (1994) Entropic analysis of random morphologies. Physica A 207:208–217

Andraud C, Beghdadi A, Haslund E, Hilfer R, Lafait J, Virgin B (1997) Local entropy characterization of correlated random microstructures. Physica A 235:307–318

Barnsley MF, Devaney RL, Mandelbrot BB, Peitgen HO, Saupe D, Voss RF (1988) In: Peitgen HO, Saupe D (eds) The science of fractal images. Springer, New York

Bartoli F, Bird NR, Gomendy V, Vivier H, Niquet S (1999) The relation between silty soil structures and their mercury porosimetry curve counterparts: fractals and percolation. Eur J Soil Sci 50: 9–17

Beghdadi A, Andraud C, Lafait J, Peiro J, Perreau M (1993) Entropic and multifractal analysis of disordered morphologies. Fractals 3:671

Behzad G-A, Millan H, Huang G (2011) A review of fractal, prefractal and pore-solid-fractal models for parameterizing the soil water retention curve. Can J Soil Sci 91:1–14

Bird N, Perrier E, Rieu M (2000) The water retention curve for a model of soil structure with pore and solid fractal distributions. Eur J Soil Sci 55:55–63

Bird N, Díaz MC, Saa A, Tarquis AM (2006) A review of fractal and multifractal analysis of soil pore-scale images. J Hydrol 322:211–219

Blair J, Falconer RE, Milne A, Young IM, Crawford JW (2007) Modeling three-dimensional microstructure in heterogeneous media. Soil Sci Soc Am J 71:1807–1812

Bouabid R, Nater EA, Barak P (1992) Measurement of pore size distribution in a lamellar Bt horizon using epifluorescence microscopy and image analysis. Geoderma 53(3–4):309–328

Brakensiek DL, Rawls WJ, Logsdon SD, Edwards WM (1992) Fractal description of macroporosity. Soil Sci Soc Am J 56:1721–1723

Buczkowski S, Kyriacos S, Nekka F, Cartilier L (1988) The modified box-counting method: Analysis of some characteristic parameters. Pattern Recogn 31(4):411–418

Bullock P, Murphy CP (1980) Towards the quantification of soil structure. J Microsc 120(3): 317–328

Cheng Q (1997a) Discrete multifractals. Math Geol 29(2):245–266

Cheng Q (1997b) Multifractal modeling and lacunarity analysis. Math Geol 29(7):919–932

Cheng Q (1999a) The gliding box method for multifractal modeling. Comput Geosci 25: 1073–1079

Cheng Q (1999b) Multifractality and spatial statistics. Comput Geosci 25(9):949–961

Cheng Q (2001) Singularity analysis for image processing and anomaly enhancement. Annual Conference of the International Association for Mathematical Geology, Cancún

Cheng Q (2007) Multifractal imaging filtering and decomposition methods in space, Fourier frequency, and eigen domains. Nonlin Process Geophys 14:293–303

Cheng Q, Agterberg FP, Ballantyne SB (1994) The separation of geochemical anomalies from background by fractal methods. J Geochem Explor 51:109–130

Chhabra A, Jensen RV (1989) Direct determination of the $f(\alpha)$ singularity spectrum. Phys Rev Lett (12):62, 1327–1330

Chun HC, Giménez D, Yoon SW (2008) Morphology, lacunarity and entropy of intra-aggregate pores: aggregate size and soil management effects. Geoderma 146(1):83–93

Cortina-Januchs MG, Quintanilla-Dominguez J, Vega-Corona A, Tarquis AM, Andina D (2011) Detection of pore space in CT soil images using artificial neural networks. Biogeosciences 8(2):279–288

Crawford JW, Matsui N (1996) Heterogeneity of the pore and solid volume of soil: distinguishing a fractal space from its non-fractal complement. Geoderma 73:183–195

Crawford JW, Matsui N, Young IM (1995) The relation between the moisture-release curve and the structure of soil. Eur J Soil Sci 46:369–375

Cumbrera R, Tarquis AM, Gascó G, Millán H (2012) Fractal scaling of apparent soil moisture estimated from vertical planes of vertisol pit images. J Hydrol 452–453:205–212

Dathe A, Thulner M (2005) The relationship between fractal properties of solid matrix and pore space in porous media. Geoderma 129:279–290

Dathe A, Eins S, Niemeyer J, Gerold G (2001) The surface fractal dimension of the soil-pore interface as measured by image analysis. Geoderma 103:203–229

Dathe A, Tarquis AM, Perrier E (2006) Multifractal analysis of the pore-and solid-phases in binary two-dimensional images of natural porous structures. Geoderma 134(3):318–326

Dullien FAL (1992) Porous media: fluid transport and pore structure, 3rd edn. Academic, San Diego
Dullien FAL (2012) Porous media: fluid transport and pore structure, 2nd edn. Academic Press Inc, pp 574
Elliot TR, Heck RJ (2007a) A comparison of optical and CT techniques for void analysis in soil thin sections. Geoderma 141:60–70
Elliot TR, Heck RJ (2007b) A comparison of 2D and 3D thresholding of CT imagery. Can J Soil Sci 87(4):405–412
Evertsz CJG, Mandelbrot BB (1992) In: Peitgen H, Jurgens H, Saupe D (eds) Chaos and fractals: new frontiers of science. Springer, New York, p 921
Falconer K (2003) Fractal geometry. Mathematical foundations and applications, 2nd edn. Wiley, West Sussex
Fallico C, Tarquis AM, De Bartoloa S, Veltria M (2010) Scaling analysis of water retention curves for unsaturated sandy loam soils by using fractal geometry. EurJ Soil Sci 61:425–436
Favretto S (2008) Applications of X-ray computed microtomography to material science: devices and perspectives. Ph.D. thesis. Universit degli studi di Trieste
Feder J (1989) Fractals. Plenum Press, New York
Fies JC, Bruand A (1990) Textural porosity analysis of a silty clay soil using pore volume balance estimation, mercury porosimetry and quantified backscattered electron scanning image (BESI). Geoderma 47:209–219
Gantzer CJ, Anderson SH (2002) Computed tomographic measurement of macroporosity in chisel-disk and no-tillage seedbeds. Soil Till Res 64:101–111
Gefen Y, Meir Y, Mandelbrot BB, Aharony A (1983) Geometric implementation of hypercubic lattices with noninteger dimensionality by use of low lacunarity fractal lattices. Phys Rev Lett 50(3):145–148
Geyger E, Beckmann W (1967) Apparate und methoden der mikromorphometrishen strukturanalyse des bodens. In: Kubiena WL (ed) Die Mikromorphometrishen Bodenanalyse. Ferdinand Enke, Stuttgart, pp 36–57
Ghanbarian-Alavijeh B, Millán H, Huang G (2011) A review of fractal, prefractal and pore-solid-fractal models for parameterizing the soil water retention curve. Can J Soil Sci 91(1):1–14
Gibson JR, Lin H, Bruns MA (2006) A comparison of fractal analytical methods on 2- and 3-dimensional computed tomographic scans of soil aggregates. Geoderma 134:335–348
Giménez D, Allmaras RR, Nater EA, Huggins DR (1997) Fractal dimensions for volume and surface of interaggregate pores - scale effects. Geoderma 77:19–38
Giménez D, Allmaras RR, Huggins DR, Nater EA (1998) Mass, surface and fragmentation fractal dimensions of soil fragments produced by tillage. Geoderma 86:261–278
Grau J, Médez V, Tarquis AM, Saa A, Díaz MC (2006) Comparison of gliding box and box-counting methods in soil image analysis. Geoderma 134:349–359
Grevers MCJ, de Jong E (1994) Evaluation of soil-pore continuity using geostatistical analysis on macroporosity in serial sections obtained by computed tomography scanning. In: Anderson, Hopmans (eds) Tomography of soil-water-root processes. SSSA, Madison, pp 73–86. SSSA Special Bub. 36
Griffith DA (1988) Advanced spatial statistics. Kluwer Academic Publishers, Boston
Hapca SM, Houston AN, Otten W, Baveye PC (2013) New objective segmentation method based on minimizing locally the intra-class variance of grayscale images. Vadose Zone J 12(3)
Hentschel HGE, Procaccia I (1983) The infinite number of generalized dimensions of fractals and strange attractors. Physica D 8:435–444
Horgan GW (1998) Mathematical morphology for analysing soil structure from images. Eur J Soil Sci 49:161–173
Houston AN, Otten W, Baveye PC, Hapca SM (2013a) Adaptive-window indicator kriging: a thresholding method for computed tomography. Comput Geosci 54:239–248
Houston AN, Schmidt S, Tarquis AM, Otten W, Baveye PC, Hapca SM (2013b) Effect of scanning and image reconstruction settings in X-ray computed microtomography on quality and segmentation of 3D soil images. Geoderma 207:154–165

Ismail SNA (1975) Micromorphometric soil-porosity characterization by means of electro – optical image analysis (Quantimet 720), Soil Survey Papers No. 9. Neth Soil Survey Inst, Wageningen. 104 pp

Jongerius A, Schoonderbeek D, Jager A, Kowalinski S (1972) Electro-optical soil porosity investigation by means of Quantimet-B equipment. Geoderma 7(3–4):177–198

Kapur JN, Sahoo PK, Wong AKC (1985) A new method for gray level picture thresholding using the entropy of the histogram. Comp Vision Graph Image Process 29:273–285

Karperien A (2013) FracLac for ImageJ (1999–2013) http://rsb.info.nih.gov/ij/plugins/fraclac/FLHelp/Introduction.htm

Kerscher M, Mecke K, Schuecker P, Böhringer H, Guzzo L, Collins CA, Schindler S, De Grandi S, Cruddace R (2001) Non-Gaussian morphology on large scales: Minkowski functionals of the REFLEX cluster catalogue. Astron Astrophys 377:1–16

Kravchenko AN, Martin MA, Smucker AJM, Rivers ML (2009) Limitations in determining multifractal spectra from pore–solid soil aggregate images. Vadose Zone J 8(1):220–226

Kravchenko A, Falconer R, Grinev D, Otten W (2010) Fungal colonization in soils of contrasting managements: modelling fungal growth in 3D pore volumes of undisturbed soil samples. Ecol Appl 21:1202–1210

Larkin TJ, Canuto HC, Kettunen MI, Booth TC, Hu D-E, Krishnan AS, Bohndiek SE, Neves AA, Charles ML, Hobson MP, Brindle KM (2014) Analysis of image heterogeneity using 2D Minkowski functionals detects tumor responses to treatment. Magn Reson Med 71(1):402–410

Lehmann P (2005) Pore structures: measurement, characterization and relevance for flow and transport in soils. Proc Appl Math Mech 5:39–42

Lehmann P, Wyss P, Flisch A, Lehmann E, Vontobel P, Krafczyk M, Kaestner A, Beckmann F, Gygi A, Fluhler H (2006) Tomographical imaging and mathematical description of porous media used for the prediction of fluid distribution. Vadose Zone J 5(1):80–97

Li CH, Lee CK (1993) Minimum cross entropy thresholding. Pattern Recogn 26(4):617–625

Liu Y, Xia Q, Cheng Q, Wang X (2013) Application of singularity theory and logistic regression model for tungsten polymetallic potential mapping. Nonlinear Process Geophys 20:445–453

Lovejoy S, Tarquis A, Gaonac'h H, Schertzer D (2008) Single-and multiscale remote sensing techniques, multifractals, and modis-derived vegetation and soil moisture. Vadose Zone J 7:533–546

Luo L, Lin H (2009) Lacunarity and fractal analyses of soil macropores and preferential transport using micro-X-ray computed tomography. Vadose Zone J 8:233–241

Mandelbrot BB (1983) The fractal geometry of nature. Macmillan Ed

Martín-Sotoca JJ, Saa-Requejo A, Grau JB, Tarquis AM (2017) New segmentation method based on fractal properties using singularity maps. Geoderma 287:40–53

McBratney AB, Moran CJ (1990) A rapid method of analysis for macropore structure: 2 Stereological model, statistical analysis, and interpretation. Soil Sci Soc Am J 54:509–522

McBratney AB, Moran CJ, Stewart JB, Cattle SR, Koppi AJ (1992) Modifications to a method of rapid assessment of soil macropore structure by image analysis. Geoderma 53(3–4):255–274

Mecke K (1998) Integral geometry and statistical physics. Int J Mod Phys B 12(9):861–899

Mecke K, Arns CH (2005) Fluids in porous media: a morphometric approach. J Phys Condens Matter 17:S503–S534

Michielsen K, De Raedt H (2001) Integral-geometry morphological image analysis. Phys Rep 347:461–538

Moran CJ (1994) Image processing and soil micromorphology. In: Ringrose-Voase AJ, Humphreys GS (eds) Soil micromorphology: studies in management and genesis, Roc. M Int. Working Meeting onSoil Micromorphology, Townsville, Australia, July 1992. Developments in Soil Science 22. Elsevier, Amsterdam, pp 459–482

Moran CJ, McBratney AB (1992a) Acquisition and analysis of three-component digital images of soil structure: I. Method. J Soil Sci 43:541–549

Moran CJ, McBratney AB (1992b) Acquisition and analysis of three-component digital images of soil structure: II. Application to seed beds in a fallow management trial. J Soil Sci 43:551–566

Moran CJ, Koppi AJ, Murphy BW, McBratney AB (1988) Comparison of the macropore structure of a sandy loam surface soil horizon subjected to two tillage treatments. Soil Use Manag 4(3):96–102

Moran CJ, McBratney AB, Koppi AJ (1989) A rapid method for analysis of soil macropore structure. I. Specimen preparation and digital binary image production. Soil Sci Soc Am J 53:921–928

Murphy WM, Scholl JM, Baretto I (1977) Effects of cutting management on eight subtropical pasture mixtures. Agron J 69(4):662–666

Oh W, Lindquist B (1999) Image thresholding by indicator kriging. IEEE Trans Pattern Anal Mach Intell 21:590–602

Ojeda-Magaña B, Quintanilla-Domínguez J, Ruelas R, Tarquis AM, Gomez-Barba L, Andina D (2014) Identification of pore spaces in 3D CT soil images using PFCM partitional clustering. Geoderma 217:90–101

Oleschko K (1998) Delesse principle and statistical fractal sets: 1. Dimensional equivalents. Soil Tillage Res 49:255

Oleschko K, Brambila F, Aceff F, Mora LP (1998) From fractal analysis along a line to fractals on the plane. Soil Till Res 45:389–406

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9:62–66

Pachepsky YA, Yakovchenko V, Rabenhorst MC, Pooley C, Sikora LJ (1996) Fractal parameters of pore surfaces as derived from micromorphological data: effect of long term management practices. Geoderma 74:305

Pajor R, Falconer R, Hapca S, Otten W (2010) Modelling and quantifying the effect of heterogeneity in soil physical conditions on fungal growth. Biogeosciences 7:3731–3740

Perfect E, Donnelly B (2015) Bi-phase box counting: an improved method for fractal analysis of binary images. Fractals 23(1):1540010. 11 pp

Perfect E, Gentry RW, Sukop MC, Lawson JE (2006) Multifractal Sierpinski carpets: theory and application to upscaling effective saturated hydraulic conductivity. Geoderma 134:240–252

Perret JS, Prasher SO, Kantzas A, Langford C (1997) 3D visualization of soil macroporosity using X-ray CAT scanning. Can Agric Eng J 39:249–261

Perret JS, Prasher SO, Kantzas A, Langford C (1998) Characterization of macropore morphology in a sandy loam soil using X-ray computer assisted tomography and geostatistical analysis. Can Water Resour J 23:143–166

Perret J, Prasher SO, Kantzas A, Langford C (1999) Three-dimensional quantification of macropore networks in undisturbed soil cores. Soil Sci Soc Am J 63:1530–1543

Perret JS, Prasher SO, Kacimov AR (2003) Mass fractal dimension of soil macropores using computed tomography: from box-counting to the cube-counting algorithm. Eur J Soil Sci 54:569–579

Perrier E, Bird N (2002) Modelling soil fragmentation: the PSF approach. Soil Tillage Res 64 (1–2):91–99

Perrier E, Rieu M, Sposito G, de Marsily G (1996) Models of the water retention curve for soils with a fractal pore-size distribution. Water Resour Res 32(10):3025–3031

Perrier E, Bird N, Rieu M (1999) Generalizing the fractal model of soil structure: the PSF approach. Geoderma 88(1999):137–l64

Perrier E, Tarquis AM, Dathe A (2006) A program for fractal and multifractal analysis of two-dimensional binary images: computer algorithms versus mathematical theory. Geoderma 134(3):284–294

Peyton RL, Gantzer CJ, Anderson SH, Haeffner BA, Pfeifer P (1994) Fractal dimension to describe soil macropore structure using X ray computed tomography. Water Resour Res 30:691–700

Pierret A, Capowiez Y, Belzunces L, Moran CJ (2002) 3D reconstruction and quantification of macropores using X-ray computed tomography and image analysis. Geoderma 106:247–271

Plotnick RE, Gardner RH, Hargrove WW, Prestegaard K, Perlmutter M (1996) Lacunarity analysis: a general technique for the analysis of spatial patterns. Phys Rev E 53:5461–5468

Posadas AND, Giménez D, Quiroz R, Protz R (2003) Multifractal characterization of soil pore spatial distributions. Soil Sci Soc Am J 67:1361–1369

Protz R, Van den Bygaart AJ (1998) Towards systematic image analysis in the study of soil micromorphology. Sci Soils 3. Available online at http://link.springer.de/link/service/journals/

Protz R, Sweeney SJ, Fox CA (1992) An application of spectral image analysis to soil micromorphology: I. Methods of analysis. Geoderma 53:275–287

Rachman A, Anderson SH, Gantzer CJ (2005) Computed-tomographic measurement of soil macroporosity parameters as affected by staff-stemmed grass hedges. Soil Sci Soc Am J 69:1609–1616

Rasband W (2006) Rasband, W.S., Image J, U. S. National Institutes of Health, Maryland, USA, http://imagej.nih.gov/ij/, 1997–2016

Rasiah V, Aylmore LAG (1998a) Estimating microscale spatial distribution of conductivity and pore continuity using computed tomography. Soil Sci Soc Am J 62:1197–1202

Rasiah V, Aylmore LAG (1998b) Characterizing the changes in soil porosity by computed tomography and fractal dimension. Soil Sci 163:203–211

Ridler TW, Calvard S (1978) Picture thresholding using an iterative selection method. IEEE Trans Syst Man Cybern 8:630–632

Rieu M, Sposito G (1991) Fractal fragmentation, soil porosity, and soil water properties: I Theory. Soil Sci Soc Am. 1(55):1231–1238. : II Applications. Soil Sci Soc Am. 1(55):1239–1244

Ringrose-Voase AJ (1996) Measurement of soil macropore geometry by image analysis of sections through impregnated soil. Plant Soil 183:27–47

Ringrose-Voase AJ, Bullock P (1984) The automatic recognition and measurement of soil pore types by image analysis and computer programs. J Soil Sci 35:673–684

Ripley BD (1988) Statistical inference for spatial processes. Cambridge University Press, Cambridge

Rockhold ML, Yarwood RR, Selker JS (2004) Coupled microbial and transport processes in soils. Vadose Zone J 3:368–383

Rogasik H, Crawford JW, Wendroth O, Young IM, Joshko M, Ritz K (1999) Discrimination of soil phases by dual energy X-ray tomography. Soil Sci Soc Am J 63:741–751

Roy A, Perfect E (2014) Lacunarity analyses of multifractal and natural grayscale patterns. Fractals 22:1440003. 9 pp

Russ JC (1995) The image processing handbook. CRC Press LLC, Boca Raton

Saa A, Gascó G, Grau JB, Antón JM, Tarquis AM (2007) Comparison of gliding box and box-counting methods in river network analysis. Nonlinear Process Geophys 14(5):603–613

Sahoo P, Wilkins C, Yeager J (1997) Threshold selection using Renyi's entropy. Pattern Recogn 30(1):71–84

Serra J (1982) Image analysis and mathematical morphology by Jean Serra. Academic, Orlando. 0-12-637240-3

Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 13(1):146–165

Stewart JB, Moran CJ, Wood JT (1999) Macropore sheath: quantification of plant root and soil macropore association. Plant Soil 211:59–67

Taina I, Heck R, Elliot T (2008) Application of x-ray computed tomography to soil science: a literature review. Can J Soil Sci 379(88):1–19

Tarquis AM, Losada JC, Benito RM, Borondo F (2001) Multifractal analysis of tori destruction in a molecular Hamiltonian system. Phys Rev E 65:016213

Tarquis AM, Giménez D, Saa A, Díaz MC, Gascó JM (2003) Scaling and multiscaling of soil pore systems determined by image analysis. In: Radcliffe Y, Selim (eds) Scaling methods in soil physics, pachepsky. CRC Press, Boca Raton, 434 pp

Tarquis AM, McInnes KJ, Key JR, Saa A, Garcia MR, Diaz MC (2006) Multiscaling analysis in a structured clay soil using 2D images. J Hydrol 322(1):236–246

Tarquis AM, Mèndez,V, Grau JB, Antòn JM, Andina D (2007). Fractals as pre-processing tool for computational intelligence application. In Computational Intelligence. Springer US, pp 193–212

Tarquis AM, Heck RJ, Grau JB, Fabregat J, Sanchez ME, Antón JM (2008) Influence of thresholding in mass and entropy dimension of 3-D soil images. Nonlinear Process Geophys 15(6):881–891

Tarquis AM, Heck RJ, Andina D, Alvarez A, Antón JM (2009) Pore network complexity and thresholding of 3D soil images. Ecol Complex 6(3):230–239

Tarquis A, Platonov A, Matulka A, Grau J, Sekula E, Diez M, Redondo J (2014) Application of multifractal analysis to the study of sar features and oil spills on the ocean surface. Nonlinear Process Geophys 21:439–450

Terribile F, FitzPatrick EA (1992) The application of multilayer digital image processing techniques to the description of soil thin sections. Geoderma 55(1–2):159–174

Torre IG (2014) Theory and application of multifractal analysis methods in images for the study of soil structure. Master thesis, UPM

Torre Iván G, Losada Juan C, Tarquis, Ana M. 2016 Multiscaling properties of soil images. Biosyst Eng. http://dx.doi.org/10.1016/j.biosystemseng.2016.11.006

Vogel HJ, Kretzschmar A (1996) Topological characterization of pore space in soil – sample preparation and digital image analysis. Geoderma 73:23–38

Vogel HJ, Weller U, Babel U (1993) Estimating orientation and width of channels and cracks at soil polished blocks—a stereological approach. Geoderma 56(1–4):301–316

Vogel HJ (2000) A numerical experiment on pore size, pore connectivity, water retention, permeability, and solute transport using network models. Eur J Soil Sci 51:99–105

Vogel HJ, Hoffmann H, Roth K (2005) Studies of crack dynamics in clay soil. I. Experimental methods, results, and morphological quantification. Geoderma 125:203–211

Voss R (1986) Random fractals: characterization and measurement. In: Pynn R, Skjeltorp A (eds) Scaling phenomena in disordered systems. Plenum, New York

Wang W, Kravchenko AN, Smucker AJM, Rivers ML (2011) Comparison of image segmentation methods in simulated 2D and 3D microtomographic images of soil aggregates. Geoderma 162:231–241

Warner GS, Nieber JL, Moore ID, Geise RA (1989) Characterizing macropores in soil by computed tomography. Soil Sci Soc Am J 53:653–660

Young IM, Crawford JW (1991) The fractal structure of soil aggregates: its measurement and interpretation. Eur J Soil Sci 42(2):187–192

Young IM, Crawford JW, Nunan N, Otten W, Spiers A (2008) Microbial distribution in soils: physics and scaling. Adv Agron 100:81–121

Zamora-Castro SA, Oleschko K, Flores L, Ventura E Jr, Parrot JF (2008) Fractal mapping of pore and solid attributes. Vadose Zone J 7:473–492

Zhou H, Perfect E, Li BG, Lu YZ (2010) Effects of bit depth on the multifractal analysis of grayscale images. Fractals 18(01):127–138

Zhou H, Perfect E, Lu YZ, Li BG, Peng XH (2011) Multifractal analyses of grayscale and binary soil thin section images. Fractals 19(03):299–309

Chapter 7
Pedotransfer Functions and Soil Inference Systems

José Padarian, Jason Morris, Budiman Minasny, and Alex. B. McBratney

"You can't make a silk purse out of a sow's ear".

Jonathan Swift
Anglo-Irish essayist (1667–1745)

7.1 Introduction

The term *pedotransfer function* (PTF) was coined by Bouma (1989) as 'translating data we have into what we need'. Pedotransfer functions are regression functions used to predict soil properties that would be otherwise infeasible to obtain. Typical reasons for this infeasibility include, but are not limited to, the cost, time, difficulty or hazard involved in procuring direct measurements. Each PTF is developed around some insight into a soil's physical, chemical or biological properties that relates a set of input parameters (predictor properties) to an output parameter (a predicted property).

Pedotransfer functions (PTFs) have multiples uses. They are essential, for example, in soil carbon stock assessment (Chap. 23) based on legacy soil data, where bulk density is usually not measured. PTFs can also be used to estimate soil organic carbon pools required in soil carbon evolution models. In digital soil mapping (Chap. 12), the use of pedotransfer functions is to provide more useful information in relation to soil attributes or soil functions. Pedotransfer functions can further be used to estimate the soil's condition or capability (e.g. available water capacity). The predicted properties resulting from PTFs can be used as inputs into process-based simulation models to run scenarios on the effects of different agricultural management on soil functioning, drainage, evapotranspiration and biomass yields.

J. Padarian (✉) • J. Morris • B. Minasny • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: jose.padarian@sydney.edu.au; jason.morris@sydney.edu.au;
budiman.minasny@sydney.edu.au; alex.mcbratney@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_7

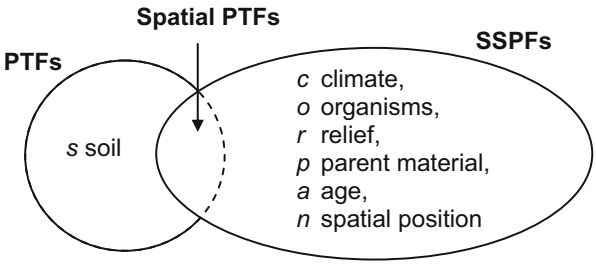


Fig. 7.1 A Venn diagram showing the relationship between pedotransfer functions (PTFs), soil spatial prediction functions (SSPFs) and, their intersection, spatial PTFs

Some consider prediction of soil attributes from environmental variables (e.g. climate and topographic indices) in a spatial context as *pedotransfer functions*, but we would caution against that. We called this particular case the soil spatial prediction functions (SSPFs) (see Chap. 12 for more details). Pedotransfer functions sensu stricto are when we are predicting soil attributes from other soil attributes, or $S = f(s)$. There is a possible intersection or area of overlap, between PTFs and SSPFs (e.g. the spatial component), which results in what we call *spatial* pedotransfer functions. Figure 7.1 illustrates the differences and possible overlap between PTFs and SSPFs. Pachepsky et al. (2001) and Romano and Palladino (2002) illustrate examples of spatial or contextual pedotransfer functions, but they are examples of $S = f(s,r)$.

7.2 A Brief History of Pedotransfer Functions

Reviews on the development and the use of PTFs can be found in Pachepsky et al. (1999, 2015) and Wosten et al. (2001). Most of these reviews, however, are limited to the prediction of soil hydraulic properties, which regulate the retention and movement of water and chemicals in soils.

The concept of using empirical relations to predict soil properties can be traced to Briggs and McLane (1907) and Briggs and Shantz (1912) in their work on determining the wilting coefficient. Furthermore, various 'rule of thumbs' were formulated to estimate various soil properties. Probably because of its particular difficulty and cost of measurement, the most comprehensive research in developing PTFs has been for the estimation of water retention. With the introduction of the concepts of field capacity (FC) and permanent wilting point (PWP) by Veihmeyer and Hendricksen (1927), research during the period of 1950–1980 attempted to correlate particle-size distribution, bulk density and organic matter content with water content at field capacity (FC, θ at -33 kPa), permanent wilting point (PWP, θ at -1500 kPa) and available water content (AWC = FC – PWP). Nielsen and Shaw (1958), for example, presented a parabolic relationship between clay content and PWP from 730 Iowa soils.

In the 1960s various papers dealt with the estimation of FC, PWP and AWC, notably in a series of papers by Salter and Williams (1965a, b, 1966, 1967, 1969).

These explored relationships between texture classes and available water capacity, which are now known as *class PTFs*. Salter and Williams also developed functions relating the particle-size distribution to AWC, now known as *continuous PTFs*.

In the 1970s more comprehensive research using large databases was developed. A particularly good example is the study by Hall et al. (1977) who used soil samples from England and Wales. Hall et al. (1977) established field capacity, permanent wilting point, available water content and air capacity as a function of textural class, as well as deriving continuous functions estimating these soil water properties. In the USA, Gupta and Larson (1979) developed 12 functions relating particle-size distribution and organic matter content to water content at water potentials ranging from −4 to −1500 kPa.

With the flourishing development of hydraulic models (van Genuchten 1980) and computer modelling of soil water and solute transport (de Wit and van Keulen 1972), the need for hydraulic properties as input to these models became more and more evident. Clapp and Hornberger (1978) derived average values for the parameters of a power-function water retention curve, sorptivity and saturated hydraulic conductivity for different texture classes. In probably the first research of its kind, Bloemen (1980) derived the relationships between parameters of the Brooks and Corey hydraulic model and particle-size distribution.

Lamp and Kneib (1981) introduced the term *pedofunction*, while Bouma and van Lanen (1986) used the term *transfer function*. To avoid confusion with the terminology, transfer function which is used in other disciplines with many different meanings, Bouma (1989) later termed the *pedotransfer function*.

From the 1990s to the early 2000s, the development of hydraulic PTFs became a popular topic of research. Results of such research have been reported widely from various countries globally, including the UK (Mayr and Jarvis 1999), Australia (Minasny and McBratney 2000), the Netherlands (Wösten et al. 1995), Germany (Scheinost et al. 1997b) and Iran (Ghorbani and Homaei 2002).

Since the late 2000s, the popularity of developing hydraulic PTFs continued (Santra and Das 2008; Twarakavi et al. 2009; Haghverdi et al. 2012). Here, the development of PTFs for special conditions is worth noting, such as saline and saline-alkali soils of Iran (Abbasi et al. 2011), permafrost soils of China (Yi et al. 2013) and volcanic ash soils of Japan (Nanko et al. 2014), and the use and development of PTFs for continental or global extent such as the work presented by Dai et al. (2013) for China, Hollis et al. (2012) and Tóth et al. (2015) for Europe and Glendining et al. (2011) for the world.

In addition, some PTFs consider adjustments because of the differences in criteria and measurements from existing pedotransfer functions. For example, as outlined in the previous chapter of this book (Fig. 5.3), sand fractions are different according to the IUSS/Australian classification (particle diameter 20–2000 μm) and the FAO/USDA criteria (particle diameter 50–2000 μm). Padarian et al. (2012) give equations for converting between these two classification systems. On the other hand, Henderson and Bui (2002) established relationships between pH measured in water and pH measured in $CaCl_2$.

Although most PTFs have been developed to predict soil hydraulic properties, they are not restricted to hydraulic properties only. PTFs for estimating soil physical, mechanical, chemical and biological properties have also been developed (Table 7.1). In addition, PTFs were developed to also predict processes such as deep

Table 7.1 Some examples of pedotransfer functions

Predicted soil properties	Predictor variables	Authors
Physical properties		
Infiltration rate after a certain period	Initial water content, moisture deficit, total porosity, non-capillary porosity, hydraulic conductivity	Canarache et al. (1968)
Soil thermal conductivity	Texture, organic matter content, water content	De Vries (1966), Hubrechts and Feyen (1996)
Bulk density	Particle-size distribution	Rawls (1983)
		Hollis et al. (2012)
Infiltration parameters	Particle-size distribution, bulk density, organic C content, initial water content, root content	van de Genachte et al. (1996)
Gas diffusivity	Air-filled porosity at −10 kPa	Moldrup et al. (2000)
Mechanical properties		
Soil mechanical resistance	Organic carbon content, clay content, bulk density	Mirreh and Ketcheson (1972), da Silva and Kay (1997)
Soil shrinkage curve	Clay content	Crescimanno and Provenzano (1999)
Volumetric shrinkage, liquid limit, plastic limit, plasticity index	Organic matter content, clay content, CEC	Mbagwu and Abeh (1998)
Degree of overconsolidation	Bulk density, void ratio	McBride and Joose (1996)
Rate of structural change	Organic matter content, clay content	Rasiah and Kay (1994)
Soil erodibility factor	Geometric mean particle-size, clay and organic matter content	Torri et al. (1997)
Chemical properties		
Cation exchange capacity (CEC)	Clay content, organic matter content, pH	Bell and van Keulen (1995), Curtin and Rostad (1997)
Critical P level, P buffer coefficient	Clay content	Cox (1994), Chen et al. (1997)
Soil organic matter	Soil colour	Fernandez et al. (1988)
P sorption	pH in NaF	Gilkes and Hughes (1994)
pH buffering capacity	Organic matter content, clay content	Wong et al. (2013)

(continued)

Table 7.1 (continued)

Predicted soil properties	Predictor variables	Authors
Al saturation	Base saturation, organic carbon content, pH	Jones (1984)
P saturation	Extractable P, Al	Kleinman et al. (1999)
K/Ca exchange	Clay content, extractable K	Scheinost et al. (1997a)
Total nitrogen	Organic carbon content, pH, sand and clay content	Glendining et al. (2011)
As and Cd sorption	Clay content, pH, organic carbon content, dithionite extractable Fe	Schug et al. (1999)
Phosphorous (P) adsorption	Clay content, pH, soil colour	Sheinost and Schwertmann (1995)
Cd sorption coefficient	Clay content, organic carbon content, pH	Springob et al. (1998)
Haematite content	Soil colour	Torrent et al. (1983)
Biological properties		
Microbial phylotype richness and diversity	pH	Fierer and Jackson (2006)
Respiration rate	Water content	Wildung et al. (1975)
Nitrogen mineralisation parameters	CEC, total N, organic carbon content, silt and clay content	Rasiah (1995)
OC pools	Total OC, clay	Weihermueller et al. 2013

percolation. For example, Selle and Huwe (2005) used a regression tree approach to simplify process-based models to identify key soil and environmental variables which govern percolation. Wessolek et al. (2008) called these hydro-pedotransfer functions, as soil and hydrological variables are used to predict other soil processes. Wessolek et al. (2008) developed empirical functions that predict deep percolation and evapotranspiration from soil conditions, vegetation and land uses.

Pachepsky et al. (2015) reviewed more recent developments in PTFs and identified research gaps that require future work:

- The need for sufficient upscaling of PTFs. PTFs were mainly generated on point observations, and many applications require simulations on regional or continental extent. An example is saturated hydraulic conductivity which is highly dependent on the measurement support.
- The need for more regional or specific PTFs for saline soils, calcareous and gypsiferous soils, peat soils, paddy soils, soils with well-expressed shrink-swell behaviour, and soils affected by freeze-thaw cycles.
- The need for parameters governing biogeochemical processes, such as in soil carbon and nitrogen evolution models, where parameters are related to organic matter pools (e.g. Weihermueller et al. 2013). For these cases, soil heat transfer and water availability inputs can be improved.

- The need to expand work on the spatial and temporal structure of PTFs which is not well known.
- The use of PTFs in large-scale projects, where soil information is usually not represented properly. In soil carbon stock assessment studies, bulk density is usually not measured, so PTFs for bulk density are required, which can be a main source of uncertainty (Hollis et al. 2012).

7.3 Developing Pedotransfer Functions

The basic steps for developing PTFs are simple – in theory, $S = f(s)$, and therefore:

1. Collect a sufficient data set of soil properties (S and s) that are suspected to having empirical relationships to each other.
2. Set aside a certain fraction of the data for developing the PTFs, and use the remaining data for testing the performance of the PTFs (e.g. an 80:20% split of the data set).
3. Choose a modelling method f for analysing the data (e.g. linear regression, neural networks or other machine learning algorithms), and subsequently develop the empirical equations.
4. Test the empirical equations on the testing data to show their validity.
5. Calculate the output uncertainty.

7.4 Predictors

There are several sources of information that can be used to predict soil properties and that can be considered as input for pedotransfer functions. Here, we will present the use of PTFs and their potential predictors which are sourced from the laboratory, field description (including soil morphology) as well as the soil electromagnetic spectrum.

7.4.1 Laboratory Data

Laboratory analysis of soil samples is usually conducted to allocate a particular soil profile to an existing soil class. The high cost of laboratory analysis, however, drove the development of empirical relationships by relating more easily or routinely measured soil properties to other attributes that are, for example, more useful for soil management purposes. One of the well-known examples is the estimation of available water capacity from particle-size distribution. The development in

pedotransfer functions is boosted by the availability of large national or regional soil databases, which allows the use of machine learning tools. The most useful variable in predicting soil physical properties is perhaps clay content, as it affects moisture retention, soil strength and many physical and chemical processes. Routine analysis usually lacks of physical data. Research is still mainly focusing on improving the prediction of hydraulic properties, such as water retention and saturated hydraulic conductivity. Some simpler analysis, however, has also been utilised to estimate more difficult-to-measure properties, such as pH in sodium fluoride which is an indication of phosphorous sorption capacity (Gilkes and Hughes 1994).

7.4.2 Field Description and Soil Morphology

Most research has been focused on correlating laboratory-determined soil properties with more difficult-to-measure properties, mainly because of the availability of comprehensive soil survey databases and the presumption that these properties are most appropriate for predictive purposes. However, it has also been recognised for some time that soil morphological description could be used as predictor (O'Neal 1949, 1952; McKeague et al. 1984; McKenzie and McLeod 1989; McKenzie and Jacquier 1997).

Calhoun et al. (2001) contended that soil morphology and field description have been underutilised in the development of pedotransfer functions. They presented the representation of Jenny's state factors through the variables' physiography, parent material, horizon, field texture and structure as collected in soil surveys for predicting bulk density. They demonstrated that morphology and field descriptors account for more variability in predicting bulk density than laboratory measurement of particle size and organic carbon. Physiographic description and soil morphological characterisation (slope gradient, position of the slope and horizon classes) were also found as useful predictors of water retention (Rawls and Pachepsky 2002).

Several studies have been successful in predicting hydraulic conductivity by using soil morphological features (e.g. O'Neal 1952; McKeague et al. 1982). However, the descriptive systems and interpretative guidelines in conventional soil survey have been largely qualitative and only appropriate for a given range of soils. McKenzie et al. (1991) found that several published descriptive systems for inferring hydraulic properties provided poor predictions for a limited range of soils from South Australia. McKenzie and Jacquier (1997) reasoned that good predictive relationships should only be expected when the field criteria used have a logical physical connection with hydraulic properties. They further postulated that predictive systems that develop direct relationships between hydraulic properties and field criteria of physical significance should be superior to systems that rely on classified entities such as horizons or soil series. They devised a simple visual estimate of areal porosity and found that saturated conductivity can be estimated from field texture, grade of structure, areal porosity, bulk density, dispersion index and horizon type. A similar idea was performed by Lin et al. (1999), who

converted morphological properties to scores which are related to water flow. From these studies, it was concluded that additional morphological descriptors to those routinely surveyed may be needed to improve the predictive capacity.

7.4.3 Handheld and On-the-Go Proximal Soil Sensing and Remote Sensing

7.4.3.1 Handheld, Stationary, Proximal Soil Sensing

As outlined in Chap. 5, in traditional soil surveys, soil scientists used the visible light spectrum through the Munsell soil colour chart to determine soil colour and the presence of pedological features like mottles or concretions. Furthermore, it was discussed in the previous chapter that developments in spectroscopy have resulted in an increase in the potential for soil analysis, and we will include a short summary of its capability here (Fig. 7.2). Diffuse reflectance infrared spectroscopy in both the visible-near (400–700–2500 nm) and mid-infrared ranges (2500–25,000 nm) allows rapid acquisition of soil information in the field or in the laboratory. Diffuse reflectance infrared spectroscopy is based on the fact that molecules have specific frequencies at which they rotate or vibrate corresponding to discrete energy levels. Absorption spectra of compounds are a unique reflection of their molecular structure. Spectral signatures of soil materials are characterised by their reflectance to a particular wavelength in the electromagnetic spectrum. Soil spectra in the vis-NIR and MIR ranges can be used to estimate a range of soil physical, chemical and biological properties simultaneously. Good results were reported for measurement of total C, total N, clay and sand content, CEC and microbial activity (Soriano-Disla et al. 2014).

Mid-infrared (MIR) spectroscopy usually produces better predictions than vis-NIR. The use of MIR also enables estimation of various soil organic carbon pools derived from tedious and time-consuming physical fractionation procedures. These pools can be used as inputs in soil carbon evolution models. Vis-NIR spectrometers particularly are used extensively and gained popularity in soil science because they are also available in a portable format and easy and ready to use in the field and require minimal or even no sample preparation. Reviews on the use of vis-NIR for predicting soil properties can be found in Stenberg et al. (2010) and Soriano-Disla et al. (2014).

Because soil is a complex mixture of materials, it is difficult to assign specific features of the spectra to specific chemical components. Ultraspectral data obtained from infrared spectrometers contain thousands of reflectance values as a function of wavelength. Since there are more predictor variables than the observations and predicted soil attributes as outlined in the previous chapter, methods that reduce the dimension of the spectra are required. Principal component regression and partial least squares (PLS) methods are commonly utilised. Principal component regression reduces the dimension of the spectra via principal component analysis and then form

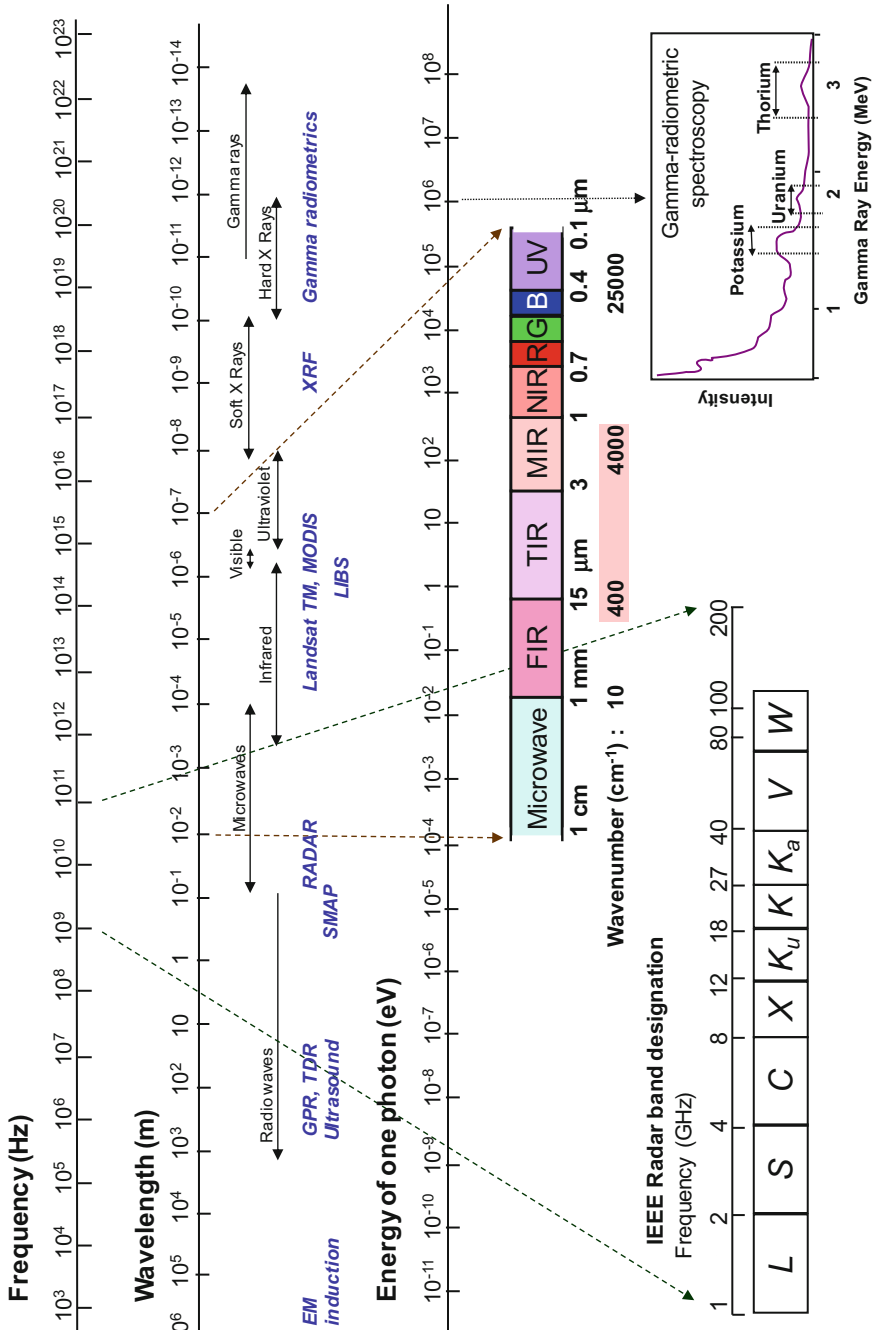


Fig. 7.2 The electromagnetic spectrum and regions useful for soil measurement

linear regression between the principal components and soil attributes (Martens and Naes 1989; Chang et al. 2001). Partial least squares (PLS) (Martens and Naes 1989) extracts successive linear combinations of the spectra, which optimally address the combined goals of explaining response variation and explaining predictor variation. Other machine learning techniques that are capable of variable (wavelength) selection have also been found useful (e.g. Minasny and McBratney 2008; Sarajith et al. 2016).

In addition to vis-NIR spectroscopy, the direct measurement of the elemental concentration of soils in the field also became possible using energy-based portable X-ray fluorescence (XRF) devices (Weindorf et al. 2012). Bulk density can also be estimated utilising photogrammetry via a digital single-lens camera or laser scanning (Bauer et al. 2014; Rossi et al. 2008).

7.4.3.2 On-the-Go Proximal Soil Sensing

While we can collect detailed soil information at limited locations using conventional methods of soil analysis and interpolate resulting values across space and time using geostatistics, in some instances it would be more beneficial if we could directly measure soil information at a fine spatial scale (e.g. measurements every 2–20 m). In this instance, proximal soil sensing offers a cost- and time-effective solution (Viscarra Rossel et al. 2010). Proximal soil sensing acquires information about soil through the use of field-based sensors that are placed in proximity to the soil (within 2 m) or within the soil body, which is in contrast to remote sensing (McBratney et al. 2011a, b). The development and use of on-the-go proximal soil sensing techniques is motivated by the need for high-resolution spatial and temporal soil information. Proximal soil sensors operate on a range of frequencies in the electromagnetic spectrum, from microwaves to gamma rays. These sensing devices either measure soil properties directly or can be used to make inferences via PTFs about specific soil properties. Often sensors are also used simultaneously to overcome the limitations of single-sensor data interpretation (Wong et al. 2010). For example, electromagnetic induction instruments (EMI) are used to measure the soil's electrical conductivity, a highly valuable soil property that is influenced by soil porosity, moisture content, salinity, temperature and the amount and composition of soil colloids.

Ground-penetrating radar, electrical resistivity as well as electrical conductivity sensors are available to monitor the spatial distribution of soil moisture (Adamchuk et al. 2004). In addition, gamma ray spectrometers have been used to measure the amount of potassium, uranium and thorium in the upper soil profile which is most likely directly related to the parent material the surveyed soil originated from (Dickson and Scott 1997). Local PTFs have been developed to estimate soil attributes (such as clay and organic carbon content) from the sensed variables (e.g. bulk electrical conductivity, gamma K).

As outlined in Chap. 5, portable sensors can now be used in the field on profile and core faces for pedological studies, which is termed digital soil morphometrics

(Hartemink and Minasny 2014). Field observation via proximal sensors and PTFs should be fused in an inference system into a powerful approach for estimating a range of soil properties for pedological studies, precision farming or contamination assessment (Horta et al. 2015).

7.4.3.3 Remote Sensing

The value of remote sensing over proximal sensing is that large spatial extents can be covered quickly with many estimates. The inferred value of remotely sensed data either airborne or satellite sourced has been shown to be an efficient means of assessing the condition of natural resources at reasonably broad scales (and this will be discussed further in Chap. 13). The remotely sensed data can include spectral, radar, thermal and radiometric signals. These reflect the environmental and soil condition and are known to be associated with soil properties. Mulder et al. (2011) reviewed the application of optical and microwave remote sensing for soil and terrain mapping. Soil properties that have been measured include mineralogy, texture, soil iron content, soil moisture content, soil organic carbon content, soil salinity and carbonate content. Its use for soil mapping is, however, hampered by vegetation cover. Nevertheless, indicators, such as plant functional groups, NDVI and productivity changes, can be used as indications of soil properties.

The application of remotely sensed infrared data for mapping soil clay content and mineralogy is demonstrated by Mulder et al. (2013) and Gomez et al. (2015). Some studies demonstrated that time series data collected from remotely sensed data can be used to derive soil hydraulic properties. Dimitrov et al. (2014) derived soil hydraulic parameters, surface roughness and soil moisture of a tilled bare soil plot using measured brightness temperatures at 1.4 GHz (L-band), rainfall and potential soil evaporation. This required a radiative transfer model and a soil hydrologic model combined with an optimisation routine.

7.5 Modelling Approaches

Approaches to develop PTFs can be purely empirical or physico-empirical. Empirical approaches attempt to find relationships between the predictor and predicted variables using regression analysis or various machine learning models. In a physico-empirical approach, the soil properties are derived based on some physical principles. For example, in water retention curve prediction, Arya and Paris (1981) translated the particle-size distribution into a water retention curve by converting solid mass fractions to water content and pore-size distribution into hydraulic potential by means of the capillary equation. Zeiliguer et al. (2000) proposed an additive model for soil water retention, which assumed that water retention of a soil can be approximated by the sum of the components of water retention of its textural composition.

Considering the type of data we wish to predict, we can distinguish single point and parametric PTFs. Single point PTFs predict a soil property, while parametric PTFs predict parameters of a model.

Most survey agencies have their own 'rule of thumb' for predicting soil properties. One form is a look-up table, which usually relates field texture class to properties such as clay content, available water capacity, etc. These rules or tables are usually derived from experience and expert knowledge or from means of properties for a particular class in a soil database.

For the continuous predicted variables, a range of machine learning models can be used to derive PTFs, finding relationships between the predictor and predicted variables. Many of the modern regression techniques are described in Hastie et al. (2009). The methods range from linear regression, generalised linear models (GLM) and generalised additive models (GAM) to regression trees, random forests, neural networks, genetic programming and fuzzy systems. Most of these tools are available in commercial and open-source projects. R (https://www.r-project.org) and Python (https://www.python.org/) are commonly used by the scientific community, because they offer many free-of-use advanced mathematical and machine learning tools.

The predictive power and interpretability vary between models depending on their complexity. Tables 7.2 and 7.3 provide a guideline for various models. The more complex the model, the more parameters it will have, so users need to be aware of the principle of parsimony (which is a general principle that for any model, which provides an adequate fit for a set of data, the one with the fewest parameters

Table 7.2 Common machine learning algorithms used for developing PTFs

Multiple regression
The general purpose of multiple regression is to analyse the relationship between several independent or predictor variables and a dependent or predicted variable. Multiple regression analysis fits a straight line (or plane in an *n*-dimensional space, where *n* is the number of independent variables) to the data
Generalised linear models (GLM)
A class of models that arise for a natural generalisation of ordinary linear models. The transformed dependent variable values are predicted from (are linked to) a linear combination of predictor variables; the transformation is referred to as the link function; also different distributions can be assumed for the dependent variable values
Generalised additive models (GAM)
Models that use smoothing techniques, such as splines to identify and represent possible nonlinear relationships between the predictor and predicted variables. GAM is a generalisation of GLM where the linear function of the predictor is replaced by an unspecified (non-parametric) function, obtained by applying a scatterplot smoother to the scatterplot of partial residuals (for the transformed dependent variable values)
Partial least squares (PLS)
This is an alternative to multiple linear regression that can deal with data having more independent variables than observation points. PLS constructs a new set of components as regressor variables which are a linear combination of the original variables. The components in partial least squares are determined by both the response variable(s) and the predictor variables

(continued)

Table 7.2 (continued)

Artificial neural networks
A flexible mathematical structure modelled after the functioning of the nervous system, capable of fitting nonlinear relationships. The essential feature is a network of simple processing elements (neurons) joined together by weights
Regression tree
This is an alternative to multiple regression. Rather than fitting a model to the data, a tree structure is generated by dividing the sample recursively into a number of groups, each division being chosen so as to maximise some measurable difference in the predicted variable in the resulting two groups. The resulting structure provides easy interpretation as variables most important for prediction can be identified quickly
Random forests
Random forest is an extension of regression trees where many trees are generated, varying the number of covariates used and using a bootstrap sampling of the training data. Then, an ensemble model is generated by aggregating the individual trees
Genetic programming
Machine learning method for evolving computer programs, following the concepts of natural selection and genetics, to solve complex problems
Support vector machine (SVM)
A method that looks for an optimal separating hyperplane between two classes by maximising the margin between the closest observations of each class. In a regression case, the observations lie in between the two borders of the margin (supporting vectors), which are separated from the hyperplane by $\pm\ \varepsilon$ (maximum error)

Based on Hastie et al. (2009)

is to be preferred) (Lark 2001). There is a limit for predictive models; here, users should choose the simplest model that can adequately account for the variation in the prediction. Models with high complexity will appear to fit the data very well; however, these may also cause overfitting or include too many parameters in the model; thus the model will fit the noise of the data. It is recommended to split the data into a calibration and validation set, using the calibration data for fitting and then testing or validating the model with a validation set (see Hastie et al. (2009) for more detail). Wosten et al. (2001) compared the performance of three models to predict water content at −33 kPa from basic soil properties using the same data set. They reported that the accuracy of all three methods was similar and suggested that the improvement of fit may not be expected from the use of different models, but from a better set of data.

7.5.1 Ensemble Models

An alternative to selecting a single predictive model is model ensembles. This consists of creating multiple models and combining them to obtain a single final model. The advantage of this method is that, most of the time, the combined model performs better than any of the individual models in terms of lower error and

Table 7.3 Comparison of different mathematical predictive models

Feature	Linear models	GLM	GAM	Regression tree	Random forests	Neural network	Genetic programming	SVM
Parsimony	☺	😐	☹	☺	☹	☹	☺	😐
Interpretability	☺	😐	☹	☺	😐	☹	☺	☹
Nonlinearity	☹	☹	☺	☺	☺	☺	☺	☺
Handling of mixed data type (qualitative and quantitative)	😐	☺	☺	☺	☺	☹	☺	☺
Computational efficiency (large data)	☺	☺	😐	☺	☹	☹	😐	😐
Predictive power	☹	😐	😐	😐	☺	☺	😐	☺

☺ good, ☹ poor 😐 fair
Adapted from Hastie et al. (2009)

unaltered bias. This method has been used for almost 200 years as pointed out in an interesting review by Clemen (1989). Baker and Ellison (2008a) discussed various aspects of implementation of ensemble methods for soil studies. In soil science, examples of its use are Baker and Ellison (2008b) who used ensemble ANN for PTFs. Kim et al. (2015) combined two microwave satellite soil moisture products, Malone et al. (2014) combined estimates of soil properties from soil maps and regression kriging prediction, and Padarian et al. (2014) generated an ensemble map of soil available water capacity in Australia.

Guber et al. (2009) suggested the use of all available PTFs in a multimodel prediction technique. They used 19 published PTFs as inputs in Richards' soil water flow equation; the output of the 19 simulations was then combined to obtain a more optimal soil water prediction. The challenge in this type of ensemble method is how to calibrate and to use appropriate weighting for each of the PTF to obtain an optimal prediction.

7.6 Characterising PTF's Performance

As with all numerical methods, there are questions concerning how well any prediction agrees with real observational data. In the literature, PTFs can be characterised by their accuracy, reliability, uncertainty and validity, as well as their ultimate utility. A brief survey of these concepts follows.

7.6.1 Accuracy

Accuracy refers to how well a PTF predicts its target property based on inputs taken from the training data. It measures the performance of a PTF on its training data (a PTF has 'seen' the data). Usually accuracy is expressed in terms of error, the difference between observed and predicted values. Weynants et al. (2009) amongst others used several common statistic measures for evaluating the accuracy of PTFs: the root-mean-square error (RMSE), mean absolute error (MAE), mean error (ME) or bias, coefficient of determination (R^2) and the model efficiency. Accuracy in PTFs can also be computed with other statistics, e.g. the concordance correlation coefficient which measures how close the model predictions fall along a 45-degree line from the origin with the measured data (or a slope of exactly 1) (Lawrence and Lin 1989).

7.6.2 Reliability

Reliability in PTFs refers to a PTF's performance in making predictions on data outside its original training data (data a PTF has not 'seen') (Pachepsky and Rawls 1999). A reliable PTF should produce accurate predictions for seen data (data used in the model development process), as well as unseen data (data that had not been used in the model development process) (Baker and Ellison 2008a). Pachepsky and Rawls (1999) state that the reliability of PTFs can be estimated by cross validation, or using an independent data set. In the cross validation method, the training data set is split into two subsets – a calibration set and a validation set; two-thirds of the data for calibration and one-third for testing are a common practice. However, the results from such cross validation can be biassed against the data set used. If a PTF is intended for prediction over a region, the independent test data set should contain observations that are unbiased (in statistics, collected based on a random sampling approach). PTFs that lack independent validation result in potentially optimistic assumptions about the functions' predictive performance.

7.6.3 Validity

Validity has to do with how appropriate a particular PTF is in predicting a soil property from a given soil sample. The greater the similarity of a soil sample to the soil used to develop a PTF, the greater the assumed validity of that PTF. Validity can be in terms of the geographical and pedological region over which a PTF's original training data were collected. If a PTF is used to predict soil properties outside its original data boundaries, its validity is doubtful (Wösten et al. 1999). Not surprisingly, PTFs perform best on soils having similar parent material and pedogenesis to the soils used to develop them (Bruand et al. 2003). Acutis and Donatelli (2003) stated that validity in PTFs is strictly related to the data set used to develop them. They add that when many PTFs are available to predict the same property, knowing which one to choose is a difficult task.

An important mechanism for establishing validity is stratification or the custom creation of PTFs strictly on soil-type or classification scheme basis. Stratification has been conducted according to soil horizons (Hall et al. 1977); soil classes (Batjes 1996); textural classes (Tietje and Hennings 1996); hydraulic-functional horizons (Wösten et al. 1986); great soil groups, temperature regime and moisture regime (Pachepsky and Rawls 1999); parent material and horizon morphology (Franzmeier 1991); numerical soil class (Williams et al. 1983); and management units (Droogers and Bouma 1997).

Validity can also refer to the congruence between some input data set and the original training set. Despite knowledge that the validity of a given PTF should not be interpolated or extrapolated beyond the pedological origin or soil type on which it is developed, there is still a lack of appropriate information that adequately describes

the calibration data, and, thus, we know very little about where a published PTF may be applied. There is still a lack of a mechanism that can automatically check its validity. Tranter et al. (2009) give a method for determining the valid domain of a PTF based on Mahalanobis distance of the predictor space, cautioning that it is unwise to extrapolate PTFs beyond these bounds. The uncertainty estimates of a PTF can also be a measure of its validity.

7.6.4 Uncertainty

Uncertainty refers to the variability in a prediction from its mean value. This occurs because PTF inputs and outputs are random variables. Therefore, they have a mean value and a variance. PTF uncertainty is typically reported as the prediction variance. Uncertainty in PTF prediction can be quantified in terms of structural uncertainty due to flaws in the PTF model, uncertainty due to sampling and measurement errors and parameter uncertainty of the PTF.

Vereecken and Herbst (2004) suggest three approaches to handling uncertainty in PTFs: (1) Compute the RMSE at 90% confidence; (2) Quantify parameter uncertainty using a covariance in PTFs during the calibration process; and (3) Use a Monte Carlo analysis to quantify parameter uncertainty associated with sampling effects in the calibration database, e.g. the bootstrap method (Efron and Tibshirani 1993).

PTF uncertainty can be computed empirically based on the calibration error using the fuzzy k-means with extragrade (FkME) method given by Tranter et al. (2010). It does not seek to disseminate sources of error but rather expresses uncertainty in the form of a prediction interval determined empirically from the calibration data. The method partitions the predictor space into classes of similar model errors, with each class represented by a prediction interval determined from the empirical distribution of the error. In addition, it also identifies those observations that exist outside the convex hull of the calibration data, thus ascertaining validity of the PTF. Those observations outside the convex hull are considered outliers of the calibration data and subsequently have their uncertainty penalised by a simple multiplier.

7.6.5 Utility

Wösten et al. (2001) stated that the utility of PTFs in modelling is defined as the correspondence between measured and simulated functional soil behaviour. This can be interpreted to mean that the authors advocate validating the final use (utility) of the PTFs, not just the PTF predictions. An example would be to develop some PTFs that predict water retention and conductivity and then to use those predictions in crop simulations to predict seasonal water storage. Thus, the validation occurs

at the seasonal water storage level, not at the level of the individual predictions of water retention and conductivity.

7.7 Spatial Pedotransfer Functions

Most PTFs have been calibrated from point source data and assume spatial independence. In digital soil mapping, we are interested in estimating the spatial distribution of soil properties. Pringle et al. (2007) recommended that an investigator who wishes to apply a PTF in a spatially distributed manner first has to establish the spatial scales relevant to their particular study site. Following this, the investigator must ascertain whether these spatial scales correspond to those that are adequately predicted by the available PTFs. Pringle et al. (2007) proposed three aspects of performance in the evaluation of a spatially distributed PTF: (i) the correlation of observed and predicted quantities across different spatial scales, (ii) the reproduction of observed variance across different spatial scales and (iii) the spatial pattern of the model error. For an example of predicting water retention across a 5 km transect, they showed that the tested PTFs performed quite well in reproducing a general spatial pattern of soil water retention; however, the magnitude of observed variance was underestimated. Springer and Cundy (1987) compared the parameters of the Green-Ampt infiltration equation from field measurements and those calculated from PTFs. They showed that the mean and variance of the parameters when estimated by PTFs were not preserved; the variances are always lower. The spatial trends and cross-correlations amongst the parameters were also reduced. They further used the PTFs to simulate overland flow and found that the results were significantly different when using field-measured parameters.

When measured properties are spatially limited, spatial prediction is required to generate a continuous map. Combination of spatial interpolation methods such as kriging and PTFs can generate a continuous map, and there are two possibilities to combine them. The first approach is to first interpolate related soil properties at unvisited locations using kriging and then to apply PTFs to the interpolated variables. The second approach applies PTFs to point measurements and then interpolates the predicted results. Bocneau (1998) compared these approaches to estimate CEC in West Flanders province, Belgium, and found that the performance of both methods is almost equal. Sinowski et al. (1997) compared these approaches in estimating the water retention curve and found that the first approach yields better prediction.

Heuvelink and Pebesma (1999) discussed the role of support or scale. As most PTFs were derived from point sources, they are not valid at the block support. This means that in the situation where the PTF input is available at point support and where output is required at block support, spatial aggregation should take place after the functions are calculated. It is essential to separate spatial aggregation from spatial interpolation. Interpolation should better take place before a function or model is executed because this enables a more efficient use of the spatial distribution

characteristics of individual inputs. When a model is executed with interpolated inputs, it is important to note the uncertainty of the interpolation.

7.8 Soil Inference Systems

While there are many similar pedotransfer functions generated using new or existing data sets, there seems to be much less effort in gathering and using the available PTFs. McBratney et al. (2002) proposed a soil inference system that would match the available input with the most appropriate PTF to predict properties with the lowest uncertainty. The soil inference system was proposed as a way of collecting and making better use of pedotransfer functions that have been abundantly generated. McBratney et al. (2002) demonstrated the first approach towards building a soil inference system is to create a very rudimentary system in the form of a specially adapted spreadsheet. Such a rudimentary inference system has two essentially new features. Firstly, it contains a suite of published pedotransfer functions, and the output of one PTF can act as the input to other functions (if no measured data are available). Secondly, the uncertainties in estimates are inputs, and the uncertainties of subsequent calculations are performed. The input consists of the essential soil properties.

The inference engine will work in the following manner:

1. Predict all the soil properties using all possible combinations of inputs and PTFs.
2. Select the combination that leads to a prediction with the minimum variance.

There have been some attempts at pattern matching of PTFs using a distance metric (Tranter et al. 2009) or nearest-neighbour algorithm (Nemes et al. 2006). However, there have been no research applications that do what soil inference systems (SINFERS) aim to do, to build a system that would chain the PTF predictions together while accounting for uncertainty.

Morris et al. (2016) built an expert system software, which uses rules to select appropriate PTFs and predicts new property values and error estimates. SINFERS can use the estimated property values as new inputs, which can trigger more matching patterns and more PTFs to 'fire' cyclically until the knowledge base is exhausted and SINFERS has inferred everything it can about what it was originally given.

7.9 Soil Spectral Inference Systems

As discussed in Sect. 7.4.3, soil spectroscopy and proximal soil sensing research have mainly focused on spectral calibration and prediction of a range of soil properties using multivariate statistics. PTF research, on the other hand, is mainly focusing on predicting soil model parameters from other soil properties. There is no

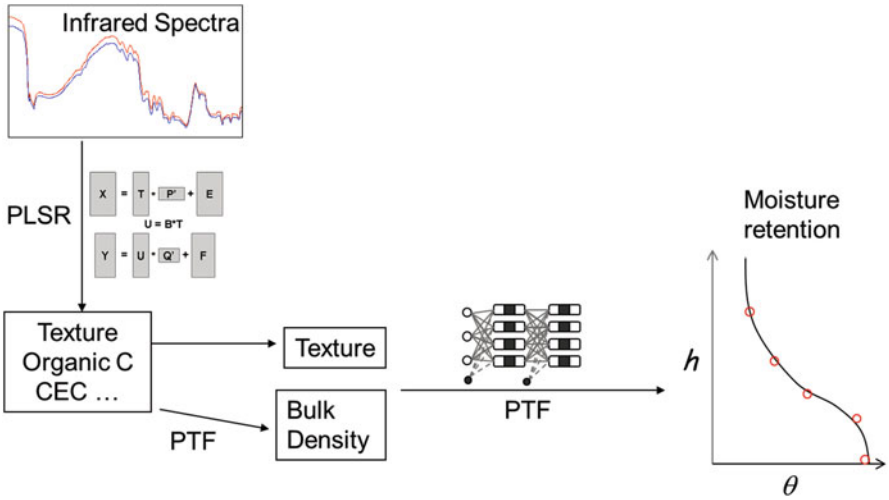


Fig. 7.3 An example of a spectral soil inference system. Soil spectra were used to predict important soil properties, and these properties were, in turn, used to predict other properties, applying well established PTFs

real connection between these research areas which have the same aim, to predict one soil property from other soil properties $S = f(s)$.

It is desired to develop soil spectral calibrations for a complete suite of soil physical, chemical and biological properties. However, this might not be possible, mainly for two reasons: (i) Not all soil properties show a spectral response and (ii) the development of a comprehensive soil spectral library is quite challenging. McBratney et al. (2006) proposed a spectral soil inference system (SPEC-SINFERS), where soil diffuse reflectance spectroscopy is linked with PTFs. SPEC-SINFERS uses soil spectra to estimate various basic soil properties which are then used to infer other important and functional soil properties via pedotransfer functions (Fig. 7.3). An important feature to be considered is the propagation of both input and model uncertainties. Tranter et al. (2008) demonstrated the use of the SPEC-SINFERS approach in predicting volumetric soil water retention. This is for sure a research area that requires future investigations.

References

Abbasi Y, Ghanbarian-Alavijeh B, Liaghat AM, Shorafa M (2011) Evaluation of pedotransfer functions for estimating soil water retention curve of saline and saline-alkali soils of Iran. Pedosphere 21(2):230–237

Acutis M, Donatelli M (2003) SOILPAR 2.00: software to estimate soil hydrological parameters and functions. Eur J Agron 18(3):373–377

Adamchuk VI, Hummel JW, Morgan MT, Upadhyaya SK (2004) On-the-go soil sensors for precision agriculture. Comput Electron Agric 44(1):71–91
Arya LM, Paris JF (1981) A physicoempirical model to predict soil moisture characteristics from particle size distribution and bulk density data. Soil Sci Soc Am J 45(1023):1030
Baker L, Ellison D (2008a) The wisdom of crowds – ensembles and modules in environmental modelling. Geoderma 147(1):1–7
Baker L, Ellison D (2008b) Optimisation of pedotransfer functions using an artificial neural network ensemble method. Geoderma 144(1):212–224
Batjes NH (1996) Development of a world data set of soil water retention properties using pedotransfer rules. Geoderma 71:31–52
Bauer T, Strauss P, Murer E (2014) A photogrammetric method for calculating soil bulk density. J Plant Nutr Soil Sci 177(4):496–499
Bell MA, Van Keulen H (1995) Soil pedotransfer functions for four Mexican soils. Soil Sci Soc Am J 59(3):865–871
Bloemen GW (1980) Calculation of hydraulic conductivities of soils from texture and organic matter content. J Plant Nutr Soil Sci 143(5):581–605
Boucneau G (1998) Geographical information science applied to soils of West-Flanders. PhD thesis. Faculty of Agricultural and Applied Biological Sciences. University of Ghent
Bouma J (1989) Using soil survey data for quantitative land evaluation. Adv Soil Sci 9:177–213
Bouma J, Van Lanen HAJ (1987) Transfer functions and threshold values: from soil characteristics to land qualities
Briggs LJ, McLane JW (1907) The moisture equivalent of soils. USDA Bur Soils Bull 45:1–23
Briggs LJ, Shantz HL (1912) The wilting coefficient and its indirect determination. Bot Gaz 53(1):20–37
Bruand A, Fernandez PP, Duval O (2003) Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. Soil Use Manag 19(3):232–242
Calhoun FG, Smeck NE, Slater BL, Bigham JM, Hall GF (2001) Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. Soil Sci Soc Am J 65:811–819
Canarache A, Motoc E, Dumitriu R, Rijtema P (1968) Infiltration rate as related to hydraulic conductivity, moisture deficit and other soil properties. In: Water in the unsaturated zone. Proceedings of the Wageningen symposium, vol 1, pp 392–401
Chang C-W, Laird DW, Mausbach MJ, Hurburgh CR (2001) Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci Soc Am J 65:480–490
Chen G, Yost RS, Li ZC, Wang X, Cox FR (1997) Uncertainty analysis for knowledge-based decision aids: application to PDSS (Phosphorous Decision Support System). Agric Syst 55:461–471
Clapp RB, Hornberger GM (1978) Empirical equations for some soil hydraulic properties. Water Resour Res 14(4):601–604
Clemen RT (1989) Combining forecasts: a review and annotated bibliography. Int J Forecast 5(4):559–583
Cox FR (1994) Current phosphorus availability indices: characteristics and shortcomings. In: Soil testing: prospects for improving nutrient recommendations, (soiltestingpros), pp 101–113
Crescimanno G, Provenzano G (1999) Soil shrinkage characteristic curve in clay soils. Soil Sci Soc Am J 63:25–32
Curtin D, Rostad HPW (1997) Cation exchange and buffer potential of Saskatchewan soils estimated from texture, organic matter and pH. Can J Soil Sci 77(4):621–626
Da Silva A, Kay BD (1997) Estimating the least limiting water range of soils from properties and management. Soil Sci Soc Am J 61:877–883
Dai Y, Shangguan W, Duan Q, Liu B, Fu S, Niu G (2013) Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling. J Hydrometeorol 14(3):869–887

DeVries DA (1966) Thermal properties of soils. In: van Wijk WR (ed) Physics of plant environment, 2nd edn, pp 231–234
de Wit CT, van Keulen H (1972) Simulation of transport processes in soils (No. 2, p 100) Pudoc
Dickson BL, Scott KM (1997) Interpretation of aerial gamma-ray surveys: adding the geochemical factors. AGSO J Aust Geol Geophys 17(2):187–200
Dimitrov M, Vanderborght J, Kostov KG, Jadoon KZ, Weihermüller L, Jackson TJ, Bindlish R, Pachepsky Y, Schwank M, Vereecken H (2014) Soil hydraulic parameters and surface soil moisture of a tilled bare soil plot inversely derived from L-band brightness temperatures. Vadose Zone J 13(1)
Droogers P, Bouma J (1997) Soil survey input in exploratory modelling of sustainable soil management practices. Soil Sci Soc Am J 61:1704–1710
Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. In: Monographs on statistics and applied probability 57. Chapman & Hall, London
Fernandez RN, Schulze DG, Coffin DL, Van Scoyoc GE (1988) Color, organic matter, and pesticide adsorption relationships in a soil landscape. Soil Sci Soc Am J 52(4):1023–1026
Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 103(3):626–631
Franzmeier DP (1991) Estimation of hydraulic conductivity from effective porosity data for some Indiana soils. Soil Sci Soc Am J 55:1801–1803
Ghorbani DS, Homaei M (2002) Derivation of the retention curve parameters using pedotransfer functions. J Agric Eng Res 3(12):1–16
Gilkes RJ, Hughes JC (1994) Sodium fluoride pH of south-western Australian soils as an indicator of P-sorption. Aust J Soil Res 32:755–766
Glendining MJ, Dailey AG, Powlson DS, Richter GM, Catt JA, Whitmore AP (2011) Pedotransfer functions for estimating total soil nitrogen up to the global scale. Eur J Soil Sci 62(1):13–22
Gomez C, Drost APA, Roger JM (2015) Analysis of the uncertainties affecting predictions of clay contents from VNIR/SWIR hyperspectral data. Remote Sens Environ 156:58–70
Guber AK, Pachepsky YA, van Genuchten MT, Simunek J, Jacques D, Nemes A, Nicholson TJ, Cady RE (2009) Multimodel simulation of water flow in a field soil using pedotransfer functions. Vadose Zone J 8(1):1–10
Gupta S, Larson WE (1979) Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. Water Resour Res 15(6):1633–1635
Haghverdi A, Cornelis WM, Ghahraman B (2012) A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. J Hydrol 442:46–54
Hall DG, Reeve MJ, Thomasson AJ, Wright VF (1977) Water retention, porosity and density of field soils, Technical Monograph No. 9. Soil Survey of England and Wales, Harpenden
Hartemink AE, Minasny B (2014) Towards digital soil morphometrics. Geoderma 230:305–317
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference and prediction, Springer Series in Statistics, 2nd edn. Springer-Verlag, New York
Henderson BL, Bui EN (2002) An improved calibration curve between soil pH measured in waterand CaCl2. Soil Res 40(8):1399–1405
Heuvelink GBM, Pebesma EJ (1999) Spatial aggregation and soil process modeling. Geoderma 89:47–65
Hollis JM, Hannam J, Bellamy PH (2012) Empirically derived pedotransfer functions for predicting bulk density in European soils. Eur J Soil Sci 63(1):96–109
Horta A, Malone B, Stockmann U, Minasny B, Bishop TFA, McBratney AB, Pallasser R, Pozza L (2015) Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: a prospective review. Geoderma 241:180–209
Hubrechts L, Feyen J (1996) Pedotransfer functions for thermal soil properties. Institute for Land and Water Management. Katholieke University, Leuven
Jones CA (1984) Estimation of percent aluminum saturation from soil chemical data 1. Commun Soil Sci Plant Anal 15(3):327–335

Kim S, Parinussa RM, Liu YY, Johnson FM, Sharma A (2015) A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. Geophys Res Lett 42(16):6662–6670

Kleinman PJ, Bryant RB, Reid WS (1999) Development of pedotransfer functions to quantify phosphorus saturation of agricultural soils. J Environ Qual 28(6):2026–2030

Lamp J, Kneib W (1981) Zur quantitativen Erfassung und Bewertung von Pedofunktionen. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 32:695–711

Lark RM (2001) Some tools for parsimonious modelling and interpretation of within-field variation of soil and crop systems. Soil Tillage Res 58:99–111

Lawrence I, Lin K (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics:255–268

Lin HS, McInnes KJ, Wilding LP, Hallmark CT (1999) Effects of soil morphology on hydraulic properties: I. Quantification of soil morphology. Soil Sci Soc Am J 63:948–954

Malone BP, Minasny B, Odgers NP, McBratney AB (2014) Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232:34–44

Martens G, Naes T (1989) Multivariate calibration. Wiley, New York

Mayr T, Jarvis NJ (1999) Pedotransfer functions to estimate soil water retention parameters for a modified Brooks–Corey type model. Geoderma 91(1):1–9

McBratney AB, Minasny B, Cattle SR, Vervoort RW (2002) From pedotransfer function to soil inference system. Geoderma 109:41–73

Mbagwu JSC, Abeh OG (1998) Prediction of engineering properties of tropical soils using intrinsic pedological parameters. Soil Sci 163(2):93–102

McBratney AB, Minasny B, Viscarra Rossel R (2006) Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. Geoderma 136(1–2):272–278

McBratney AB, Minasny B, Tranter G (2011a) Necessary meta-data for pedotransfer functions. Geoderma 160(3):627–629

McBratney AB, Minasny B, Whelan BM (2011b) Defining proximal soil sensing. 2nd Global workshop – proximal soil sensing. Montreal, Canada

McBride RA, Joosse PJ (1996) Overconsolidation in agricultural soils: II. Pedotransfer functions for estimating preconsolidation stress. Soil Sci Soc Am J 60(2):373–380

McKeague JA, Wang C, Topp GC (1982) Estimating saturated hydraulic conductivity from soil morphology. Soil Sci Soc Am J 46:1239–1244

McKeague JA, Eilers RG, Thomasson AJ, Reeve MJ, Bouma J, Grossman RB, Favrot JC, Renger M, Strebel O (1984) Tentative assessment of soil survey approaches to the characterization and interpretation of air-water properties of soils. Geoderma 34(1):69–100

McKenzie NJ, Jacquier DW (1997) Improving the field estimation of saturated hydraulic conductivity in soil survey. Aust J Soil Res 35:803–825

McKenzie NJ, MacLeod DA (1989) Relationships between soil morphology and soil properties relevant to irrigated and dryland agriculture. Aust J Soil Res 27:235–258

McKenzie NJ, Smettem KRJ, Ringrose-Voase AJ (1991) Evaluation of methods for inferring air and water properties of soils from field morphology. Aust J Soil Res 29:587–602

Minasny B, McBratney AB (2000) Evaluation and development of hydraulic conductivity pedotransfer functions for Australian soil. Aust J Soil Res 38:905–926

Minasny B, McBratney AB (2008) Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom Intell Lab Syst 94(1):72–79

Mirreh HF, Ketcheson JW (1972) Influence of soil bulk density and matric pressure on soil resistance to penetration. Can J Soil Sci 52(3):477–483

Moldrup P, Olesen T, Gamst J, Schjønning P, Yamaguchi T, Rolston DE (2000) Predicting the gas diffusion coefficient in repacked soil water-induced linear reduction model. Soil Sci Soc Am J 64(5):1588–1594

Morris J (2016) General method for predicting soil data via pattern matching on pedotransfer functions. University of Sydney, Faculty of Agriculture and Environment. http://hdl.handle.net/2123/14575

Mulder VL, De Bruin S, Schaepman ME, Mayr TR (2011) The use of remote sensing in soil and terrain mapping – a review. Geoderma 162(1):1–19

Mulder VL, Plötze M, de Bruin S, Schaepman ME, Mavris C, Kokaly RF, Egli M (2013) Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1–2.4 μm) and regression tree analysis. Geoderma 207:279–290

Nanko K, Ugawa S, Hashimoto S, Imaya A, Kobayashi M, Sakai H, Ishizuka S, Miura S, Tanaka N, Takahashi M, Kaneko S (2014) A pedotransfer function for estimating bulk density of forest soil in Japan affected by volcanic ash. Geoderma 213:36–45

Nemes A, Rawls WJ, Pachepsky YA (2006) Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. Soil Sci Soc Am J 70(2):327–336

Nielsen DR, Shaw RH (1958) Estimation of the 15-atmosphere moisture percentage from hydrometer data. Soil Sci 86(2):103–105

O'Neal AM (1949) Some characteristics significant in evaluating permeability. Soil Sci 67:403–409

O'Neal AM (1952) A key for evaluating soil permeability by means of certain field clues. Soil Sci Soc Am Proc 16:312–315

Pachepsky YA, Rawls WJ (1999) Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci Soc Am J 63:1748–1756

Pachepsky YA, Timlin DJ, Ahuja LR (1999) The current status of pedotransfer functions: their accuracy, reliability and utility in field- and regional-scale modeling. In: Corwin DL, Loage K, Ellsworth TR (eds) Assessment of non-point source pollution in vadose zone, Geophysical Monograph, vol 108. American Geophysical Union, Washington, DC, pp 223–234

Pachepsky YA, Timlin DJ, Rawls WJ (2001) Soil water retention as related to topographic variables. Soil Sci Soc Am J 65:1787–1795

Pachepsky YA, Rajkai K, Tóth B (2015) Pedotransfer in soil physics: trends and outlook—A review—. Agrokém Talajt 64(2):339–360

Padarian J, Minasny B, McBratney A (2012) Using genetic programming to transform from Australian to USDA/FAO soil particle-size classification system. Soil Res 50(6):443–446

Padarian J, Minasny B, McBratney AB, Dalgliesh N (2014) Predicting and mapping the soil available water capacity of Australian wheatbelt. Geoderma Reg 2:110–118

Pringle MJ, Romano N, Minasny B, Chirico GB, Lark RM (2007) Spatial evaluation of pedotransfer functions using wavelet analysis. J Hydrol 333(2):182–198

Rasiah V (1995) Comparison of pedotransfer functions to predict nitrogen mineralization parameters of one and two pool models. Commun Soil Sci Plant Anal 26(11–12):1873–1884

Rasiah V, Kay BD (1994) Characterizing changes in aggregate stability subsequent to introduction of forages. Soil Sci Soc Am J 58(3):935–942

Rawls WJ, Pachepsky YA (2002) Using field topographic descriptors to estimate soil water retention. Soil Sci 167:423–435

Romano N, Palladino M (2002) Prediction of soil water retention using soil physical data and terrain attributes. J Hydrol 265:56–75

Rossi AM, Hirmas DR, Graham RC, Sternberg PD (2008) Bulk density determination by automated three-dimensional laser scanning. Soil Sci Soc Am J 72(6):1591–1593

Salter PJ, Williams JB (1965a) The influence of texture on the moisture characteristics of soils: I. A critical comparison for determining the available water capacity and moisture characteristics curve of a soil. J Soil Sci 16:1–15

Salter PJ, Williams JB (1965b) The influence of texture on the moisture characteristics of soils: II. Available water capacity and moisture release characteristics. J Soil Sci 16:310–317

Salter PJ, Berry G, Williams JB (1966) The influence of texture on the moisture characteristics of soils: III. Quantitative relationships between particle size, composition and available water capacity. J Soil Sci 17:93–98

Salter PJ, Williams JB (1967) The influence of texture on the moisture characteristics of soils: IV. A method of estimating available water capacities of profiles in the field. J Soil Sci 18:174–181

Salter PJ, Williams JB (1969) The influence of texture on the moisture characteristics of soils: V. Relationships between particle-size composition and moisture contents at the upper and lower limits of available water. J Soil Sci 20:126–131
Santra P, Das BS (2008) Pedotransfer functions for soil hydraulic properties developed from a hilly watershed of Eastern India. Geoderma 146(3):439–448
Sarathjith MC, Das BS, Wani SP, Sahrawat KL (2016) Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. Geoderma 267:1–9
Scheinost AC, Sinowski W, Auerswald K (1997a) Regionalization of soil buffering functions: a new concept applied to K/Ca exchange curves. Adv GeoEcol 30:23–38
Scheinost AC, Sinowski W, Auerswald K (1997b) Regionalization of soil water retention curves in a highly variable soilscape, I. Developing a new pedotransfer function. Geoderma 78:129–143
Schug B, Hoß T, Düring RA, Gäth S (1999) Regionalization of sorption capacities for arsenic and cadmium. Plant Soil 213(1-2):181–187
Selle B, Huwe B (2005) Optimising soil-hydrological predictions using effective CART models. Adv Geosci 5:37–41
Sinowski W, Scheinost AC, Auerswald K (1997) Regionalization of soil water retention curves in a highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function. Geoderma 78:145–159
Soriano-Disla JM, Janik LJ, Viscarra Rossel RA, Macdonald LM, McLaughlin MJ (2014) The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl Spectrosc Rev 49(2):139–186
Springer EP, Cundy TW (1987) Field-scale evaluation of infiltration parameters from soil texture for hydrologic analysis. Water Resour Res 23:325–334
Springob G, Böttcher J (1998) Parameterization and regionalization of Cd sorption characteristics of sandy soils.II. Regionalization: Freundlich k estimates by pedotransfer functions. J Plant Nutr Soil Sci 161(6):689–696
Stenberg B, Rossel RAV, Mouazen AM, Wetterlind J (2010) Chapter five-visible and near infrared spectroscopy in soil science. Adv Agron 107:163–215
Tietje O, Hennings V (1996) Accuracy of the saturated hydraulic conductivity prediction by pedotransfer functions compared to the variability within FAO textural classes. Geoderma 69:71–84
Torrent J, Schwertmann U, Fechter H, Alferez F (1983) Quantitative relationships between soil color and hematite content. Soil Sci 136(6):354–358
Torri D, Poesen J, Borselli L (1997) Predictability and uncertainty of the soil erodibility factor using a global dataset. Catena 31(1):1–22
Tóth B, Weynants M, Nemes A, Makó A, Bilas G, Tóth G (2015) New generation of hydraulic pedotransfer functions for Europe. Eur J Soil Sci 66(1):226–238
Tranter G, Minasny B, McBratney AB, Viscarra Rossel R, Murphy B (2008) Comparing spectral soil inference systems and mid-infrared spectroscopy predictions of soil volumetric moisture retention. Soil Sci Soc Am J 72:1394–1400
Tranter G, McBratney AB, Minasny B (2009) Using distance metrics to determine the appropriate domain of pedotransfer function predictions. Geoderma 149(3):421–425
Tranter G, Minasny B, McBratney AB (2010) Estimating pedotransfer function prediction limits using fuzzy-means with extragrades. Soil Sci Soc Am J 74(6):1967–1975
Twarakavi NK, Šimůnek J, Schaap MG (2009) Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. Soil Sci Soc Am J 73(5):1443–1452
Van de Genachte G, Mallants D, Ramos J, Deckers JA, Feyen J (1996) Estimating infiltration parameters from basic soil properties. Hydrol Process 10(5):687–701
Van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44(5):892–898
Veihmeyer FJ, Hendrickson AH (1927) Soil-moisture conditions in relation to plant growth. Plant Physiol 2(1):71
Vereecken H, Herbst M (2004) Statistical regression. Dev Soil Sci 30:3–19

Viscarra Rossel RA, McBratney AB, Minasny B (2010) Proximal soil sensing. Progress in soil science. Springer, Dordrecht

Weihermueller L, Graf A, Herbst M, Vereecken H (2013) Simple pedotransfer functions to initialize reactive carbon pools of the RothC model. Eur J Soil Sci 64(5):567–575

Weindorf DC, Zhu Y, McDaniel P, Valerio M, Lynn L, Michaelson G, Clark M, Ping CL (2012) Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and albic horizons. Geoderma 189–190:268–277

Wessolek G, Duijnisveld WHM, Trinks S (2008) Hydro-pedotransfer functions (HPTFs) for predicting annual percolation rate on a regional scale. J Hydrol 356(1):17–27

Weynants M, Vereecken H, Javaux M (2009) Revisiting Vereecken pedotransfer functions: introducing a closed-form hydraulic model. Vadose Zone J 8(1):86–95

Wildung RE, Garland TR, Buschbom RL (1975) The interdependent effects of soil temperature and water content on soil respiration rate and plant root decomposition in arid grassland soils. Soil Biol Biochem 7(6):373–378

Williams J, Prebble JE, Williams WT, Hignett CT (1983) The influence of texture, structure and clay mineralogy on the soil moisture characteristic. Aust J Soil Res 21:15–32

Wong MTF, Wittwer K, Oliver YM, Robertson MJ (2010) Use of EM38 and gamma ray spectrometry as complementary sensors for high-resolution soil property mapping. In: Viscarra Rossel RA, McBratney AB, Minasny B (eds) Proximal soil sensing. Springer, Dordrecht, pp 343–349

Wong MTF, Webb MJ, Wittwer K (2013) Development of buffer methods and evaluation of pedotransfer functions to estimate pH buffer capacity of highly weathered soils. Soil Use Manag 29(1):30–38

Wösten JHM, Bannink MH, De Gruijter JJ, Bouma J (1986) A procedure to identify different groups of hydraulic conductivity and moisture retention curves for soil horizons. J Hydrol 86:133–145

Wösten JHM, Lilly A, Nemes A, Le Bas C (1999) Development and use of a database of hydraulic properties of European soils. Geoderma 90(3):169–185

Wösten JHM, Finke PA, Jansen MJW (1995) Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66:227–237

Wösten JHM, Pachepsky YA, Rawls WJ (2001) Pedotransfer functions: bridging gap betwen available basic soil data and missing soil hydraulic characteristics. J Hydrol 251:123–150

Yi X, Li G, Yin Y (2013) Comparison of three methods to develop pedotransfer functions for the saturated water content and field water capacity in permafrost region. Cold Reg Sci Technol 88:10–16

Zeiliguer AM, Pachepsky YA, Rawls WJ (2000) Estimating water retention of sandy soils using the additivity hypothesis. Soil Sci 165:373–383

Part IV
Soil Materials, Horizons and Profiles

"Nature has endowed the earth with glorious wonders and vast resources that we may use for our own ends. Regardless of our tastes or our way of living, there are none that present more variations to tax our imagination than the soil, and certainly none so important to our ancestors, to ourselves, and to our children."

Charles Kellogg

In this book section, we consider some basic soil entities from a pedometric point of view. All the entities are notionally three dimensional. We consider soil materials, the soil within a small volume within the horizon, soil horizons and sequences of soil horizons or otherwise known as profiles.

We shall deal with the classification of soil objects. The terminology of soil classification is a little confused. First, we define some words in the way we use them here.

Classification – the setting up or creation of classes
Taxonomy – the theory of classification
Classification system – the list of classes and their definitions
Allocation, identification – putting new unknowns into pre-existing classes

The word 'soil' can be tacked on the front to give the appropriate phrases and meanings. So, using these definitions, Isbell et al. (1997) classifies, the rest of us allocate or identify, and hardly anyone bothers with soil taxonomy.

Reference

Isbell RF, McDonald WS, Ashton LJ (1997) Concepts and rationale of the Australian soil classification. ACLEP, CSIRO Land and Water, Canberra

Chapter 8
Soil Material Classes

Nathan P. Odgers and Alex. B. McBratney

> *"When you get close to the raw materials and taste them at the moment they let go of the soil, you learn to respect them".*
>
> Rene Redzepi

Soil classification is really about answering the question *what makes a soil?* Or, perhaps, *what makes one soil different from another?* To answer questions like these, soil classifiers create taxonomic rules to separate one kind of soil from another and categorise and make sense of the diverse pattern of the soil continuum. Traditionally a great deal of consideration has been given to characterising and classifying the whole soil profile in a top-down fashion. Pedometric methods allow us to answer the same questions in a bottom-up trajectory. Thus, the starting point is not the whole soil profile or even its major constituents, the soil horizons. Rather we start by classifying the actual, tangible, skeleton of soil itself: the *soil material*.

8.1 Soil Material Classes

We classify soil because we think the soil varies sufficiently in its properties from place to place for distinct kinds to be recognised. These various kinds of soil ought to be identifiable in the physical space (the natural landscape) and, hopefully, in the

N.P. Odgers (✉)
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia

Soils and Landscapes Team, Manaaki Whenua – Landcare Research, PO Box 69040,
Lincoln 7640, New Zealand
e-mail: OdgersN@landcareresearch.co.nz

A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: alex.mcbratney@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_8

attribute space—the *n*-dimensional space defined by *n* measurable soil attributes (MacVicar 1969). In classifying soils, we aim to use soil information gathered from the physical space to define soil classes that are as coherent as possible in the attribute space.

Taxonomists classify objects we may call *individuals*. An individual is an object that is "complete in itself" (Cline 1949). Individuals can be allocated to (can become members of) classes, and a *universe* is a superclass that contains all individuals. A *particulate* universe is one in which discrete objects can readily be counted. Biological universes are frequently particulate universes: for example, in the universe of birds, individual birds are readily discernible. On the other hand, a universe of soils does not seem to fit the description of a particulate universe since its characteristics frequently change from place to place more gradually than sharply (Simonson and Gardiner 1960). Discrete, definitive bodies of soil are not as discernible as individual organisms often are in a biological universe. Soil is, in other words, a continuum and therefore occupies a *continuous* universe in which individuals cannot readily be counted; individuals in a continuous universe must be created arbitrarily (Knox 1965).

Because soils occupy a continuous universe, the identification of soil individuals has been the source of some contention (Buol 2003). Yet as a practical matter, some notion of a soil individual has traditionally been necessary in order to discretise the soil continuum and simplify its classification.

8.1.1 Point Representation of Soil

The pedon is probably the most well-known soil individual. The pedon and the polypedon were devised as three-dimensional soil individuals for the purpose of soil sampling and classification, respectively (Soil Survey Staff 1993) and are fundamental concepts in the Soil Taxonomy classification system (Soil Survey Staff 1999). A polypedon is a collection of contiguous pedons of like taxa and is considered internally homogenous for taxonomic purposes.

Delineation of pedons in the field is complicated by the fact that soil occupies a continuous universe. For example, there is effectively an infinite number of pedons since their boundaries in the landscape must be established arbitrarily (Knox 1965). This means that their dimensions are also arbitrary, although guidelines dictate their minimum and maximum limits (Soil Survey Staff 1993). Several methods of setting the dimensions of a pedon have been proposed that are based on, for example, the volume of soil occupied by plant roots or the minimum volume that can be sampled by a particular instrument or the examination of lateral soil horizon variability (Simonson and Gardiner 1960; Soil Survey Staff 1993).

In reality what we tend to examine and classify in the field are two-dimensional soil profiles rather than three-dimensional soil bodies (Webster 1968). We often assume, implicitly, that the variation in the third dimension is irrelevant for taxonomic purposes because it is impractical in most circumstances to excavate and describe a three-dimensional body of sufficient size and in sufficient detail.

Since soil is a continuum, it seems appropriate that in our attempts to classify it we should not ignore the lateral variation when sampling it; however, it appears that little impetus is traditionally provided to collect such information. Since sampling and classification are linked, it suggests that a revised philosophy of the soil individual is necessary.

There is precedent in the literature. According to Holmgren (1986), the conceptualisation of a soil individual as an arbitrary volume is unsuitable because a soil property measurement is actually made on a volume of material originating *from a specific location on Earth*. Location is therefore a definitive "point of origin" with respect to which a soil can be characterised and classified.

It makes sense that the observations themselves should consist of a small volume representation, which we shall call a *soil material*. The soil material would logically fit inside recognised soil horizons and as such is fairly congruent with Holmgren's (1988) sampling *locule*. In geostatistical terminology, we are saying that a soil sample must have a defined geometric support. A soil material may or may not also be the same as a representative elementary volume (REV), which Bouma (1985) considered is the smallest volume that can represent a given soil horizon and lead to a consistent population of data. In aggregated soils, REVs are peds and in sandy soils they are individual sand grains.

The substance of Holmgren's revised pedon seems to be more related to the practice in soil physics, chemistry and biology. It may be aligned with the geostatistical and REV concept above although the space-time geometric support of this operationally defined object is not clear. A set of nine cores of fixed radius sampled on a grid 20 m apart and centred on some location (x, y) on which a set of measurements are made and attributes recorded according to some schema could be considered an *observational pedon*. The set of data thus obtained could be called a *pedon description*.

8.1.2 Soil Material as a Collection of Soil Properties

A large number of soil properties, for example, physical, geochemical and biological, can be attributed to or measured on a particular soil material. We shall call a collection of soil properties describing the soil material a *soil material description* (Table 8.1). The infrared spectrum offers another multivariate description (Fidêncio et al. 2001; Leone and Sommer 2000; Valeriano et al. 1995). The soil material is a real entity, whereas a soil material description is a virtual one. If different sets of properties are used to describe a soil material, then the soil material descriptions are different entities.

Table 8.1 Descriptions of soil material from three profiles observed at Pokolbin in the Lower Hunter Valley, New South Wales, Australia

Australian Soil Classification suborder	Brown Dermosol	Red Chromosol	Brown Kandosol
Horizon designation	A1 horizon	B21 horizon	B21 horizon
Texture grade	Clay loam	Light-medium clay	Light clay
Moist colour hue	10YR	5YR	7.5YR
Moist colour value	4	3	3
Moist colour chroma	3	4	4
Dry colour hue	10YR	5YR	7.5YR
Dry colour value	5	3	5
Dry colour chroma	4	4	4
Structure	Moderately pedal	Strongly pedal	Apedal massive
pH (1:5 H_2O)	5.67	6.74	5.53
pH (1:5 $CaCl_2$)	4.67	5.43	4.25
Electrical conductivity ($\mu S\ cm^{-1}$)	145.6	54.9	87.3

8.2 Creation of Classes

Scientific classifications of soil have been made since at least the late nineteenth century. These were often induced from supposed genesis. Later, more objective classification schemes were devised. The most widely known is probably today's Soil Taxonomy and its antecedent, the so-called Seventh Approximation. Most soil classifications in use today are hierarchical systems in which the soil universe is segregated into progressively more detailed, mutually exclusive classes as one proceeds down the hierarchy. This is the same model that has traditionally been applied in biology, but researchers have occasionally questioned its application to the classification of soils, in part because of the genetic assumptions implicit in the hierarchical structure (e.g. Leeper 1956).

Hughes and Lindley (1955) were some of the first researchers to classify soils using a statistical procedure although research into numerical soil classification of soils really began in earnest in the 1960s (Bidwell and Hole 1964a, b; Campbell et al. 1970; Grigal and Arneman 1969; Rayner 1966) based on work done in the biological sciences (Sneath and Sokal 1962; Sokal 1963).

Cluster analysis is the application of numerical methods to the classification of multivariate data. The goal of cluster analysis is to partition a population of individuals into classes by finding groups of similar individuals in the multivariate attribute space. The aim is that individuals allocated to the same class should be similar to each other and dissimilar to individuals allocated to the other classes (Fisher and van Ness 1971).

An individual being classified is known as a pattern, $\mathbf{x}$, and is denoted as a vector of d features or attributes x_i:

$$\mathbf{x} = (x_1, x_2, x_3 \ldots, x_d)$$

The vector $\mathbf{x}$ is usually assumed to be a row vector. A *pattern set* is a set of individuals, denoted $\boldsymbol{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ \ldots, \mathbf{x}_n\}$. In our case the attributes x_i are soil attributes. They may be measured on a continuous scale or coded on an ordinal or nominal scale. In practice, most researchers avoid the inclusion of attributes measured on ordinal and nominal scales because they can complicate the computation of pairwise similarities between individuals.

It is important to note that the clustering process is agnostic about the true partition structure amongst the population. Cluster algorithms generally partition the population of individuals into a set of classes even if natural classes are not present (Jain 2010).

8.2.1 Similarity Measures

Numerical classification relies on being able to assess the degree of similarity or resemblance between a pair of individuals. Measures that can do so are generically known as coefficients of similarity. Sneath and Sokal (1973) identified four classes of similarity coefficients: *distance coefficients*, *association coefficients*, *correlation coefficients* and *probabilistic similarity coefficients*. Association coefficients typically measure similarity on the basis of agreement in the state of qualitative attributes and include the Jaccard coefficient (Jaccard 1908) and the simple matching coefficient (Sokal and Michener 1958). Correlation coefficients measure similarity on the basis of proportionality and independence between a pair of attribute vectors and include the Pearson product-moment coefficient. Probabilistic similarity coefficients involve information statistics which measure the homogeneity of a given partition of individuals; an example can be found in Estabrook (1967).

Distance coefficients measure the distance between individuals in various ways. According to Sneath and Sokal (1973), they have the greatest intellectual appeal to taxonomists compared to other kinds of similarity coefficients since they are the easiest to visualise. They are the similarity coefficients used most frequently by pedometricians over the last few decades. Strictly speaking distance coefficients are measures of *dissimilarity* since their values increase with decreasing similarity (Clifford and Williams 1976). In the remainder of this section, we briefly describe several popular distance coefficients.

8.2.1.1 Euclidean Distance

The Euclidean distance is one of the most familiar measures of distance and was introduced to numerical taxonomy by Sokal (1961). In the two-dimensional case, it is equivalent to Pythagoras' theorem but can easily be extended to compute dissimilarities in higher-dimensional spaces. The Euclidean distance D_E^2 between individuals i and j is computed as

$$D_E^2\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sum\nolimits_{p=1}^{d} \left(\mathbf{x}_{ip} - \mathbf{x}_{jp}\right)^2 = \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \left(\mathbf{x}_i - \mathbf{x}_j\right) \tag{8.1}$$

Several pre- or post-treatments can be applied. For example, since D_E^2 grows larger as the number of attributes increases, it is a common practice to divide it by d (Clifford and Williams 1976; Webster and Oliver 1990). In addition, measurements of soil attributes tend to have different natural numerical ranges, irrespective of units of measurement. For example, soil particle-size fractions expressed gravimetrically have a range of 0–100 g kg^{-1}, whereas soil organic carbon measurements in mineral soils are frequently in the range of 0–5%. This means that attributes with larger ranges tend to have greater influence in the calculation of the Euclidean distance (Clifford and Williams 1976; Jain et al. 1999; Kantardzic 2011; Moore and Russell 1967), which may be undesirable. To avoid this we can standardise each attribute by range prior to computation of the distance (Arkley 1976):

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}) \tag{8.2}$$

or by variance:

$$x' = (x - \bar{x}) / SD_x \tag{8.3}$$

where x' is the standardised value and $\bar{x}$, x_{min} and x_{max} are the mean, minimum and maximum values of the observed range of x.

Finally, distributions of soil attributes are often skewed. Moore and Russell (1967) noted that the Euclidean distance appears to be sensitive to the shape of attribute distributions. Skew may be reduced either by transformation of the values, typically using a logarithmic or square root transformation, or by truncating extreme but rarely occurring values (Arkley 1976).

8.2.1.2 Mahalanobis Distance

Unlike the Euclidean distance, the Mahalanobis distance (Mahalanobis 1936) takes into account the differences in variance between the soil attributes and the correlations between them. As a result the attributes do not need to be standardised prior to calculation of the distance. The Mahalanobis distance is calculated as follows (McBratney and de Gruijter 1992):

$$D_M^2\left(\mathrm{x}_i, \mathrm{x}_j\right) = \left(\mathrm{x}_i - \mathrm{x}_j\right)^{\mathrm{T}} \mathbf{E}^{-1} \left(\mathrm{x}_i - \mathrm{x}_j\right) \tag{8.4}$$

where $\mathbf{E}$ is the sample covariance matrix of $\mathbf{X}$.

8.2.1.3 Gower Distance

Both the Euclidean distance and Mahalanobis distance require that the attributes are quantitative continuous variables like clay content or electrical conductivity;

Table 8.2 Values assign to s_{ijp} and δ_{ijp} for dichotomous nominal variables (Gower 1971)

	Values of attribute p			
i	+	+	−	−
J	+	−	+	−
s_{ijp}	1	0	0	0
δ_{ijp}	1	1	1	0

however, many soil attributes are measured on a qualitative scale. There are several types of qualitative variables (Williams 1976a). The first kind are called nominal variables and represent an attribute that may take many states, such as type of aggregate coatings (clay, manganese, iron, organic, etc.). A given individual can be in one state only. Although we may encode the states on an integer scale for convenience, the ordering of the states implies no special meaning. A special case of nominal variable is what Gower (1971) refers to as a *dichotomous variable*. Dichotomous variables record the presence or absence of a feature, such as mottling or stones. Two individuals have a higher degree of similarity to each other if both are in possession of the same feature than if it is absent in one individual. Absence of a feature in both individuals does not infer the same degree of similarity as its presence in both individuals since it may not be known whether the feature can occur in the populations to which the individuals belong.

Ordinal variables are similar to nominal variables in that they represent an attribute that may take many states. In this case the order of the states when encoded into an integer scale is important although the distance between states may be unknown (Williams 1976a). An example of an ordinal-valued attribute is stone size classes (e.g. 2–6 mm, 6–20 mm, 20–50 mm, >50 mm).

Gower (1971) described a general coefficient of similarity that is able to handle qualitative and nominal quantitative attributes. It is calculated as follows:

$$S_{ij} = \sum_{p=1}^{d} s_{ijp}\delta_{ijp} / \sum_{p=1}^{d} \delta_{ijp} \tag{8.5}$$

where s_{ijp} is the similarity score between individuals i and j for attribute p. The quantity δ_{ijp} represents the possibility of making a comparison between individuals i and j for attribute p; it takes a value of 1 when attribute p can be compared and 0 otherwise. When all d comparisons are possible, $\sum_{p=1}^{d} \delta_{ijp} = d$.

The method of computing the similarity score varies depending on the kind of attribute. In the case of dichotomous nominal variables, s_{ijp} and δ_{ijp} are assigned the appropriate values in Table 8.2.

In the case of multistate nominal variables, $s_{ijp} = 1$ if individuals i and j have the same state for attribute p and $s_{ijp} = 0$ otherwise.

For quantitative variables, s_{ijp} is computed as follows:

$$s_{ijp} = 1 - \left| x_i - x_j \right| / R_p \tag{8.6}$$

where R_p is the numerical range of attribute p either in the population or the sample.

Despite its flexibility, the Gower distance has not been widely used in pedometric studies although examples exist in spatial prediction research (e.g. Mallavan et al. 2010; McKenzie and Austin 1993; Zhang et al. 2013) and soil classification research (e.g. Beaudette et al. 2013; Oliver and Webster 1989; Roudier et al. 2016).

8.2.2 Preprocessing

Before we undertake a cluster analysis, it is frequently necessary to perform some preprocessing operations on the attributes of the individuals we wish to cluster. We may wish to do so for a number of reasons, including the following: we would like to choose an optimal subset of attributes or we may want to modify the influence that certain attributes have on the cluster analysis. Preprocessing objectives may be realised by undertaking several common tasks including examination of the correlations between attributes, standardisation to a common range of values, transformation to reduce skew and determination of appropriate weights (Arkley 1976). The steps we need to take are frequently determined by the assumptions underlying a particular clustering algorithm. In this section we briefly describe a couple of the more common tasks.

8.2.2.1 Optimal Subset of Attributes

We may be presented with a large number of attributes and wish to select an optimal subset. It is tempting to merely choose the several that we think are the most important, or the most interpretable or that we have the most experience with, but such a choice is likely to be suboptimal with respect to the information about the individuals' partition structure that a given set of attributes carry (Arkley 1976). In any dataset of soil material attributes, it is likely that many will be correlated with each other, and some highly so. For the sake of cluster analysis, an optimal set of attributes is the set in which the correlations between the attributes are minimised. On the other hand, for maximum pedological interpretability, they should also be as correlated as possible with other attributes not used in the cluster analysis (Norris 1971). For quantitative attributes this can be achieved by examining the Pearson correlation matrix arising from a pairwise comparison of all available attributes and selecting a subset according to some heuristic. An example of such a matrix for attributes of some undisturbed soils in Israel is presented in Table 8.3 (after Banin and Amiel 1970). For example, Sarkar et al. (1966), after assembling a dataset of 61 soil attributes, examined pairs having a correlation greater than or equal to 0.90 and eliminated the attribute that was most highly correlated with other attributes. By doing so they reduced the number of attributes to 51. Through trial and error, they were able to make bigger reductions by lowering the pairwise comparison threshold.

A similar but more rigorous outcome can be achieved by subjecting a dataset containing n soil attributes to the well-known method of principal components

Table 8.3 Correlation coefficients for attributes of some undisturbed soils in Israel, after Banin and Amiel (1970)

	S	CEC	$CaCO_3$	OM	Clay	Silt	Sand	W_H	W_{15}	$W_{1/3}$	W_S	W_A
S												
CEC	0.930											
$CaCO_3$	−0.195	−0.160										
OM	0.313	0.067	−0.063									
Clay	0.950	0.869	−0.085	0.393								
Silt	0.206	0.213	0.755	0.266	0.310							
Sand	−0.881	−0.828	−0.156	−0.423	−0.950	−0.589						
W_H	0.972	0.900	−0.187	0.419	0.962	0.282	−0.918					
W_{15}	0.922	0.897	0.033	0.374	0.900	0.379	−0.895	0.940				
$W_{1/3}$	0.886	0.885	0.183	0.290	0.881	0.540	−0.929	0.892	0.951			
W_S	0.865	0.882	−0.194	0.286	0.906	0.224	−0.847	0.899	0.848	0.834		
W_A	0.702	0.733	0.354	0.130	0.719	0.671	−0.829	0.710	0.738	0.911	0.686	

Cell values are correlation coefficients, *r*, between relevant pair of attributes
S specific surface area (measured in $m^2\ g^{-1}$), *CEC* cation exchange capacity (meq $100\ g^{-1}$), $CaCO_3$ calcium carbonate content (%), *OM* organic matter (%), *Clay* clay content(%), *Sand* sand content (%), W_H hygroscopic moisture content (%), W_{15} moisture content at 15 atm. (%), $W_{1/3}$ moisture content at 1/3 atm. (%), W_S moisture content of saturated paste (%), W_A available moisture (%)

analysis (Hotelling 1933). In simple terms, PCA involves the rotation of the *n* orthogonal axes of the feature space formed by the *n* attributes in such a way that the rotated first axis accounts for the most variance in the dataset, the rotated second axis accounts for the second largest component of variance in the dataset and so on. The rotated axes are called principal components, and the values of the individuals on each principal component axis are called scores. The principal component axes are orthogonal with each other, which guarantees that the scores measured on them are uncorrelated with each other. The full set of principal component axes are needed to describe the rotated feature space entirely, but it is probable that most of the variance is described by the first few components (Norris 1971). Thus, with respect to reducing the number of attributes required for cluster analysis, the original attributes can be replaced with the individuals' scores on the first few principal component axes with the certainty that the scores on each axis are independent of each other and that minimal information has been lost compared to that contained in the original dataset. The disadvantage is that the principal component scores are more difficult to interpret than the original attribute values and the original attributes cannot be recovered unless the scores of all the principal components are known. Kyuma and Kawaguchi (1976) used the approach when classifying Japanese paddy soils to reduce a set of 12 soil material attributes to the scores on two principal components prior to classification using a dendrogram.

8.2.2.2 Weighting of Attributes

We may also wish to weight certain attributes prior to the cluster analysis according to their perceived importance or by some other rule. Whether or not attributes are weighted often depends on whether the resulting classification is intended to reflect general-purpose or special-purpose use. If a general-purpose classification is desired, then it is generally accepted that all attributes should remain unweighted so that each has equal value and importance (Sneath and Sokal 1962). Arkley (1976) noted that this is in conflict with traditional general-purpose hierarchical soil classifications where attributes that are used as partitioning criteria at higher nodes of the hierarchy have greater effective weight in the classification than attributes appearing lower in the hierarchy.

Weighting specific attributes means that they can exert greater influence in the cluster analysis compared to the attributes that remain unweighted, and this can be reflected in the class definitions (Gibbons 1968). This has application in the construction of special-purpose classification systems if certain soil uses can be shown to depend on specific soil attributes.

8.2.3 Kinds of Clustering Algorithms

The field of cluster analysis is broad, and researchers have devised many clustering algorithms. Not all of them have been applied to soils, and, of those that have, some have proved more popular than others. The algorithms themselves can be classified in various ways. For example, Williams (1976b) made a classification with respect to application in agricultural science that reflects a time when fuzzy classification was still in its infancy. Clustering algorithms are often described in terms of (i) whether or not the classes are known a priori (supervised versus unsupervised algorithms), (ii) the trajectory of class formation (agglomerative versus partitional classification), (iii) the structure of the resulting classification (hierarchical versus non-hierarchical classification) and (iv) the exclusivity of their classes (exclusive versus nonexclusive classification). Criteria (i)–(iii) relate to the structure of the classification systems, whereas criterion (iv) relates to the nature of the classes. In this section we briefly explore these concepts.

8.2.3.1 Supervised Versus Unsupervised Classification

Supervised classification is carried out when the classes are known a priori. Supervised techniques were some of the first numerical classification techniques applied to soils. The ordination techniques that appeared in the late 1950s enabled researchers to group existing soil taxa on the basis of quantifiable similarity (Bidwell and Hole 1964b). Hole and Hironaka (1960) published one of the first studies to do so. They applied ordination to the grouping of soils in the Miami taxonomic family from Ohio and to representative profiles of great soil groups collected from around the world. Results of ordinations were often summarised using dendrograms (Bidwell and Hole 1964a, b; Rayner 1966).

Contemporary methods of classification tree analysis and artificial neural networks are also supervised classification techniques although they are more frequently used to calibrate environmental and landscape characteristics with soil observations for the purpose of spatial prediction. A recent exception was the study by Ribeiro et al. (2014) who used fuzzy classification trees to examine the relationships between soil properties and classes in the Brazilian soil classification system.

Unsupervised classification, on the other hand, attempts to discover natural classes amongst a group of individuals. It does not assume any pre-existing soil classification. For this reason it is often useful for exploring if natural groups are present in an unclassified collection of soil material samples. In the soil literature, popular unsupervised classification algorithms have included the *k*-means and, later, fuzzy *k*-means algorithms and their derivatives, described later in this chapter.

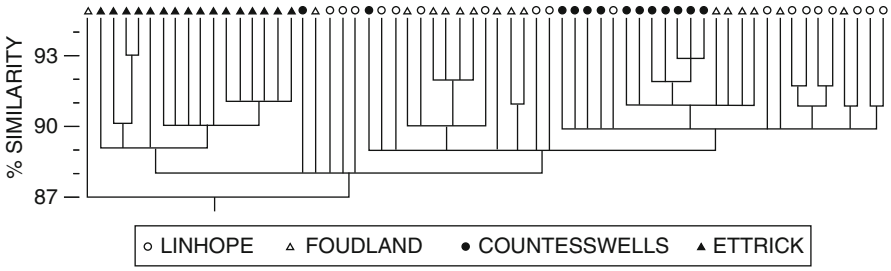


Fig. 8.1 Dendrogram of some Scottish soils produced via a single-linkage agglomerative hierarchical classification procedure (Muir et al. 1970, reproduced with permission)

8.2.3.2 Agglomerative Versus Partitional Classification

Agglomerative clustering works by grouping individuals together into larger and larger groups on the basis of the similarity between them. Given a population of n individuals, n initial clusters are formed that each contain one individual. Clusters are successively joined in n–1 steps until all have been joined together into one cluster (Fig. 8.1). At any stage, the two clusters that are chosen to be joined are those that are the closest according to some distance or similarity metric, typically in the Euclidean space. There are many versions of agglomerative clustering, and they differ primarily in how the intercluster distance is calculated. They are reviewed in detail by Anderson (1971) and Webster (1977).

The simplest agglomerative clustering method is known as nearest neighbour clustering. In this method, the intercluster distance is defined as the smallest of the distances between the members of each cluster. A more complex method known as centroid clustering (Gower 1967) defines the intercluster distance as the squared distance between cluster centroids.

Partitional, or *divisive*, clustering algorithms divide a population of individuals into a set of k classes. A key problem with this kind of clustering is choosing the number of clusters (Dubes 1987; Jain et al. 1999). Patterns corresponding to the initial class centroids may be chosen at random or by using expert knowledge. Using a similarity measure, individual patterns are allocated to the cluster whose centroid is the closest. After all patterns have been allocated, the cluster centroids are often recalculated and patterns reallocated. Iteration continues in this fashion until some measure of goodness is acceptable. One of the most well-known partitional clustering algorithms, k-means clustering, is frequently attributed to Lloyd (1982), who devised it in the 1950s. Hartigan published a more efficient version in the late 1970s (Hartigan 1975; Hartigan and Wong 1979).

8.2.3.3 Hierarchical vs. Non-hierarchical Classification

In a taxonomic hierarchy, individuals belong to groups at the lowest or most specific level of the hierarchy. Figure 8.1 depicts a hierarchical classification of some Scottish soils, although the lowest level of the hierarchy is at the top of the graph. These groups belong to more general groups at the next hierarchic level, and so on until all groups are united in a single, general group (Webster and Oliver 1990). Soil classes are frequently organised in this way due to their simplicity and relative ease of use for allocation. Hierarchies may be generated agglomeratively or divisively (Anderson 1971; Webster 1977).

Many contemporary soil classification systems employ hierarchical structures. Despite this, there are circumstances in which a hierarchical structure is neither applicable nor ideal (Dale et al. 1989; Webster and Oliver 1990). For example, hierarchies were commonly formed divisively by choosing one or a few so-called *diagnostic properties* as the subdividing criterion (or criteria) at each hierarchic level. The subdivision is typically mutually exclusive so, for instance, class A may have a topsoil organic carbon content of less than 4% and class B greater than or equal to 4%. The classes at the lowest level of the hierarchy are thus defined by a unique set of attribute values because their class limits do not overlap with those of other classes. Such classification has been called *monothetic* (Sokal and Sneath 1963). However, as Webster (1968) pointed out, soil is *polythetic* in the sense that a set of classes may possess many shared attributes that cannot be subdivided mutually exclusively.

8.2.3.4 Exclusive Versus Nonexclusive Algorithms

Clustering algorithms may also be classified according to the character of the classes that they produce. Once we allocate an individual to a class, we say that the individual has a degree of *membership*, m, in the class. This membership quantifies the degree to which the individual possesses the characteristics of the central concept of a class and can be expressed on a numerical scale from 0 to 1 (Burrough et al. 1997).

Exclusive algorithms produce classes with boundaries that are discontinuous, or hard, or crisp. Crisp classes are exclusive because an individual can belong to, or have *complete* membership in, one and only one crisp class. Numerically, in an exclusive system, an individual can have $m = 1$ to one and only one class and $m = 0$ to all others. A hard partition of n individuals into k classes can be represented by an $n \times k$ matrix of memberships $\mathbf{M} = (m_{ic})$. The following conditions apply in order to ensure that the classes are mutually exclusive, jointly exhaustive and non-empty (McBratney and de Gruijter 1992):

$$\sum_{c=1}^{k} m_{ic} = 1, \quad i = 1, \ldots, n \tag{8.7}$$

$$\sum_{i=1}^{n} m_{ic} > 0,\ c = 1, \ldots, k \tag{8.8}$$

$$m_{ic} \in \{0, 1\},\ i = 1, \ldots, n; c = 1, \ldots, k \tag{8.9}$$

The first condition ensures that, for a given individual, the memberships across all classes sum to unity; the second ensures that, for a given class, the sum of memberships of all individuals is greater than zero; the third ensures that membership can only take values of 0 or 1.

Class limits are often defined as a set of discriminating criteria using statements such as the following:

> members of class Alpha possess an A-horizon clay content of 10–20% clay.

In the preceding example, although some within-class variation in A-horizon clay is permitted, the limits themselves are crisp because an individual with an A-horizon clay content less than 10% or greater than 20% is excluded from membership in class Alpha. It is implicitly assumed that all change between classes occurs at the class boundaries and that the within-class variation is irrelevant at least for interpretive purposes (Burrough 1989). An advantage of crisp classes is that allocation is relatively simple and can often be achieved through the construction of a device such as a dichotomous key.

Crisp classes are not without drawbacks, however. Fundamentally, that their boundaries are hard means that they disregard the natural continuity in the soil attribute space (McBratney et al. 1992). As such they are incapable of representing vague concepts (Metternicht 2003). Furthermore, as variation in the soil attribute space is continuous, any placement of crisp boundaries is arbitrary. Other concerns are more practical. For example, although within-class variation in the diagnostic properties may be ignored in practice, it must be adequately known for the class limits to be established. A system of crisp classes is prone to misclassification in certain circumstances. For example, measurement error may lead to misclassification of an individual (Webster 1968) and a group of otherwise similar profiles may be allocated to different classes because they vary with respect to a single diagnostic property.

Nonexclusive algorithms produce classes with boundaries that are continuous, or fuzzy, or overlapping. Continuous classes are nonexclusive because an individual can belong to, or have *partial* membership in, more than one class simultaneously. Thus, they allow for vagueness in the class definitions that crisp classes cannot accommodate. In a nonexclusive system, the membership requirements are relaxed so that m can vary continuously between 0 and 1. In other words, the third assumption, above (Eq. 8.9), is replaced with the following:

$$m_{ic} \in [0, 1],\ i = 1, \ldots, n; c = 1, \ldots, k \tag{8.10}$$

Continuous classes are usually defined by a central concept or centroid which is essentially a soil material description consisting of the modal values for a suite

of soil attributes. Memberships are based on the degree of similarity between an individual and the class centroid. Such a definition is compatible with what Cline (1949) envisioned.

The vagueness in class allocation that is enabled by partial membership confers several advantages to continuous classes. For example, the placement of arbitrary class limits is avoided. Intergrade soils—soils with characteristics intermediate to those of two or more class definitions—are better accommodated. Finally, some of the risk of misallocation is avoided since an individual can still have partial membership in an alternate class.

8.2.4 k-Means Clustering and Its Derivatives

In this section we discuss the *k*-means clustering algorithm and several of its derivatives, which are amongst the most popular clustering algorithms used to classify soil material.

8.2.4.1 *k*-Means Algorithm

The *k*-means clustering algorithm is an unsupervised, partitional, non-hierarchical algorithm that partitions a population of individuals into crisp classes (Hartigan 1975).

The *k*-means clustering algorithm works as follows:

1. Choose a value of k with the restriction that $1 < k < n$.
2. Initialise k cluster centroids. Cluster centroid patterns may be initialised randomly or by using expert knowledge.
3. Compute distance between cluster centroids and patterns of individuals being classified.
4. Recompute cluster centroids once all individuals have been allocated to a class.
5. Repeat steps 3–4 until some convergence criterion is met (e.g. the allocation has not changed).

The similarity between individual patterns is typically computed using the Euclidean distance, although the Mahalanobis distance may also be used (Mao and Jain 1996). In step 3, an individual is allocated to the class of the cluster to which it is closest in Euclidean space. In step 4, recomputation of the cluster centroids is achieved by averaging the values of the attributes of the individuals assigned to it:

$$c_{j,p} = \frac{\sum_{\mathbf{x}_i \in \mathbf{X}} x_{ip}}{|C_j|}, 1 \leq p \leq d \tag{8.11}$$

where $|C_j|$ is the cardinality of the cluster C_j.

The algorithm reaches convergence when some criterion is met, such as when individuals cease to move from one cluster to another or when the value of an objective function ceases to decrease significantly from iteration to iteration (Jain et al. 1999). A common objective function, which computes the total Euclidean distance of the patterns to their cluster centres, is computed as follows:

$$J(\mathbf{M}, \mathbf{C}) = \sum_{i=1}^{n} \sum_{c=1}^{k} m_{ic} d^2 (\mathbf{x}_i, \mathbf{c}_c) \tag{8.12}$$

The *k*-means algorithm is easy to implement, which has made it a popular choice in cluster analyses in many fields. Despite this, users need to be aware of several factors. First, the algorithm is sensitive to the choice of the initial cluster centroids. A random selection of initial cluster centroids is likely to lead to a partition of individuals that is only *locally* optimal since an exhaustive search for the combination of individuals that yields the *global* minimum value of the objective function is computationally prohibitive. Indeed it may be impossible to prove that a given partition is globally optimal (Steinley 2006). Some researchers (Falkenauer and Marchand 2001; Hartigan 1975; Jain 2010) suggested selecting the best partition, in terms of the minimal objective value, from a pool of partitions created by running the *k*-means algorithm several times, but the best partition in this case is still unlikely to be the global optimum.

Second, because the Euclidean distance is typically used to quantify similarity between individuals, the algorithm produces convex, approximately spherical, clusters in the attribute space (Jain 2010). Since soil classes are rarely this shape (Odeh et al. 1992), this distance is likely to be inappropriate.

In soil science the *k*-means algorithm has been applied not only to soil classification (e.g. Bormann 2010; Minasny and McBratney 2006) but also digital soil mapping (Bui and Moran 2001) and sampling design (Brus et al. 2006).

8.2.4.2 Fuzzy *k*-Means

The fuzzy *k*-means algorithm is an unsupervised, partitional, non-hierarchical classification algorithm that produces continuous classes. The fuzzy *k*-means algorithm extends the notion of fuzzy logic (e.g. Zadeh 1965) to cluster analysis to allow for classes to overlap in the attribute space. Introduced in the early 1980s (Bezdek 1981), the algorithm began to be used for soil classification in the early 1990s (Odeh et al. 1990).

The fuzzy *k*-means algorithm functions in much the same way as the *k*-means algorithm except in the computation of membership functions. In the fuzzy *k*-means algorithm, the computation of cluster centroids is modified to account for the partial memberships of all the individuals associated with each cluster.

The degree of fuzziness in the fuzzy clustering can be modified and is controlled by the parameter φ, the so-called *fuzziness exponent*. The degree of fuzziness relates to the degree of overlap of the resulting classes in the attribute space. The minimum

value of φ is 1.0, which is equivalent to no fuzziness, and yields a hard partition into non-overlapping classes, much like the *k*-means algorithm. As φ increases, the degree of overlap of classes in the attribute space increases.

The membership *m* to class *c* of individual *i* is computed using the following:

$$m_{ic} = \frac{d_{ic}^{-2/(\varphi-1)}}{\sum_j^k d_{ij}^{-2/(\varphi-1)}}, i = 1, \ldots, n; c = 1, \ldots, k \tag{8.13}$$

The cluster centroids are computed using the following:

$$\mathbf{c}_c = \frac{\sum_{i=1}^n m_{ic}^{\varphi} \mathbf{x}_i}{\sum_{i=1}^n m_{ic}^{\varphi}}, c = 1, \ldots, k \tag{8.14}$$

Optimisation attempts to minimise the following objective function, which is a weighted sum of the distances between every pattern and every cluster centroid:

$$J_B(\mathbf{M}, \mathbf{C}) = \sum_{i=1}^n \sum_{c=1}^k m_{ic}^{\varphi} d_{ic}^2 \tag{8.15}$$

Compared to classes arising from the *k*-means algorithm, fuzzy classes are more robust because they have been shown to contain more information (Lagacherie et al. 1997) and be less sensitive to errors in the attribute data (Heuvelink and Burrough 1993).

Users need to be aware of some of the same factors relating to *k*-means applications. For example, continuous classes still tend to form spherical or hyperspherical classes in attribute space (Rousseeuw et al. 1996), which is inappropriate if we do not expect our classes to take such a shape. Second, the choice of *k* is still somewhat subjective although cluster validity measures can help to choose an appropriate number. Even so, some (e.g. McBratney and Moore 1985; Odeh et al. 1990) have cautioned that it may not be possible to know how many classes exist in our data because we frequently do not know how representative our soil observations are.

We may also ponder what is the appropriate value of φ. As φ determines the fuzziness of the fuzzy classification, a good value should reflect the fuzziness in the attribute space. This is usually not known in advance (Lagacherie et al. 1997). Odeh et al. (1992) suggested that the optimal φ should represent a balance between preserving natural partitional structures in the dataset and continuity of the classes. They reasoned that *k* should be established first by examining the partition entropies associated with the different *k*, and then φ could be set to reflect the appropriate level of fuzziness. The final choice remains somewhat arbitrary but should be guided by solid expert knowledge of the data.

Researchers typically use values of φ in the range of 1.1–1.5 (e.g. Burrough et al. 2000; Cockx et al. 2007; Dobermann et al. 2003; Triantafilis et al. 2001). Although researchers often do not describe the manner in which they determine φ in their fuzzy *k*-means cluster analyses, some have presented detailed studies in which they attempted to determine an appropriate φ empirically. For example, McBratney

and Moore (1985) used the derivative of J_B (**M**, **C**) with respect to φ to determine optimal values of φ and k. de Bruin and Stein (1998), in a fuzzy cluster analysis of landform components across a valley hillslope in the Netherlands, determined the optimal k and φ in terms of how well the fuzzy memberships could predict the variation in topsoil clay content across the hillslope. They found that the optimal fuzzy partition, with $k = 4$ and $\varphi = 2.1$, was able to account for about 70% variation in the soil property.

We illustrate the use of the fuzzy k-means algorithm by cluster analysis of 81 soil material samples from profiles observed at Pokolbin in the Lower Hunter Valley in New South Wales, Australia. The soil material samples were taken from within a pedogenetic horizon of their respective profile; while specific horizons were not favoured, B2 horizons were the most frequent sources of the soil material because profiles were sampled by auger and other horizons were often not thick enough to contain the volume of material required for laboratory analysis. The samples were attributed with a range of soil properties including clay content, pH and moist *Lab* colour (converted from Munsell colour notation) and effective cation exchange capacity (eCEC; Table 8.4). We ran the fuzzy k-means algorithm on the soil material samples several times in order to produce a set of partitions from $k = 2$ to $k = 25$ classes. A locally optimal partition was obtained when $k = 9$ and $\varphi = 1.2$ (see also Fig. 8.6 and associated discussion). The centroids of the resulting classes are presented in Table 8.4. Relationships between attributes appear to be pedologically sensible. For example, higher clay content is generally associated with lower sand content and vice versa, and darker-coloured soils (those with smaller moist L) tend to be associated with higher total carbon.

A biplot of the first two principal components of the observations' attributes is presented in Fig. 8.2. The 95% density ellipses were computed based on individuals with membership of 0.5 or higher in each class (after Triantafilis et al. 2001). The first two principal components account for about 74% of the variation in the data; the first four principal components account for nearly 95% of the variation. While some pairs of classes, such as E and G, are well separated, others, like A and C, transcend the distribution of several other classes in the two-dimensional representation. The loadings of the attributes enable an examination of the contribution of each attribute to the principal components (Fig. 8.3). Thus, clay content, sand content and electrical conductivity contribute most to the first principal component, whereas total carbon and $CaCO_3$ content contribute the most to the second principal component. Relationships between the soil attributes as expressed by the directions of their loadings vectors also affirm the relationships that are discernible in Table 8.4.

8.2.4.3 Fuzzy k-Means with Extragrades

Consider the synthetic dataset in Fig. 8.4a in which there are five natural classes and several outlying individuals. A weakness of the fuzzy k-means algorithm is that individuals with approximately equal memberships to all cluster centroids may

Table 8.4 Fuzzy *k*-means centroids of soil material classes derived from a collection of soil material samples from profiles observed at Pokolbin in the Lower Hunter Valley in New South Wales, Australia

Class	pH (1:5 H_2O)	Electrical conductivity (dS m^{-1})	Total C (%)	$CaCO_3$ (%)	Clay (%)	Sand (%)	Moist *Lab* colour			eCEC (meq 100 g^{-1})
							L	*a*	*b*	
A	5.6	0.348	0.27	0.41	40.2	47.7	63.5	2.0	15.5	14.98
B	4.8	0.865	0.86	0.41	8.2	15.3	21.1	4.5	12.1	28.27
C	6.8	0.399	0.39	0.40	39.0	44.2	38.9	4.5	18.6	16.89
D	5.5	0.267	0.51	0.41	50.0	39.5	39.6	20.9	29.1	16.45
E	5.2	0.254	0.63	0.40	62.3	23.0	40.5	5.9	19.2	22.62
F	5.3	0.877	0.47	0.52	52.0	32.3	41.3	4.7	17.6	21.79
G	5.9	0.130	0.81	0.41	23.7	67.5	42.1	4.9	17.9	7.83
H	6.2	0.245	0.42	0.86	37.4	52.1	47.7	10.4	42.9	16.23
I	7.1	0.131	1.61	0.61	53.6	29.2	29.0	6.7	17.0	30.97

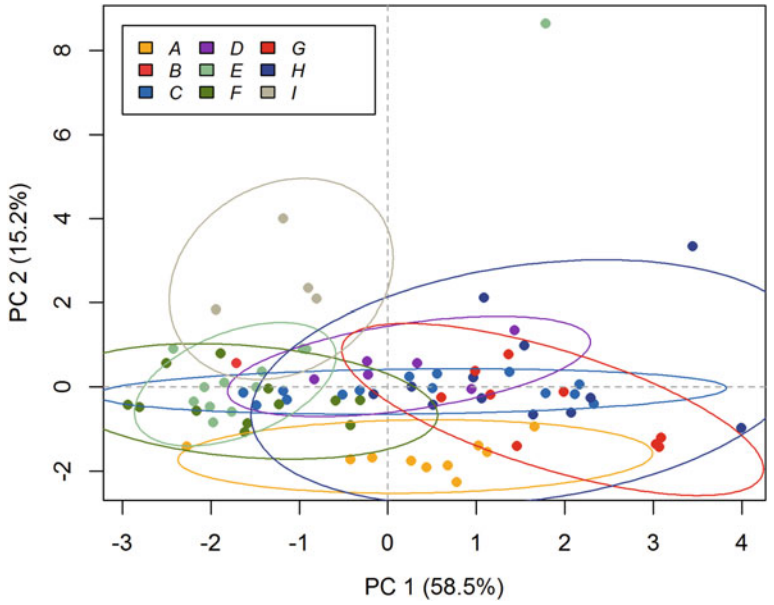


Fig. 8.2 Distribution of the soil material samples along the first two principal components of their attributes

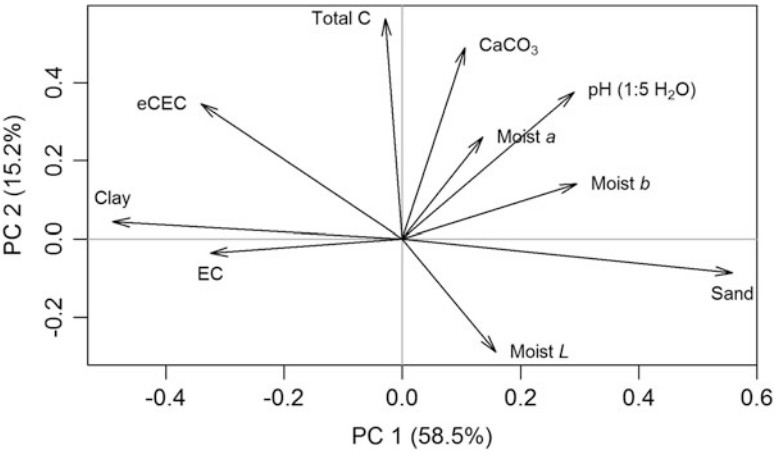


Fig. 8.3 Vectors of the 10 attributes in the space formed by the first two principal components

receive about equal memberships to their classes, whether or not the individuals lie in the centre or the outlying region of the attribute space (de Gruijter et al. 1997). This may lead to a distortion of the locations of the centroids in the attribute space (Fig. 8.4b).

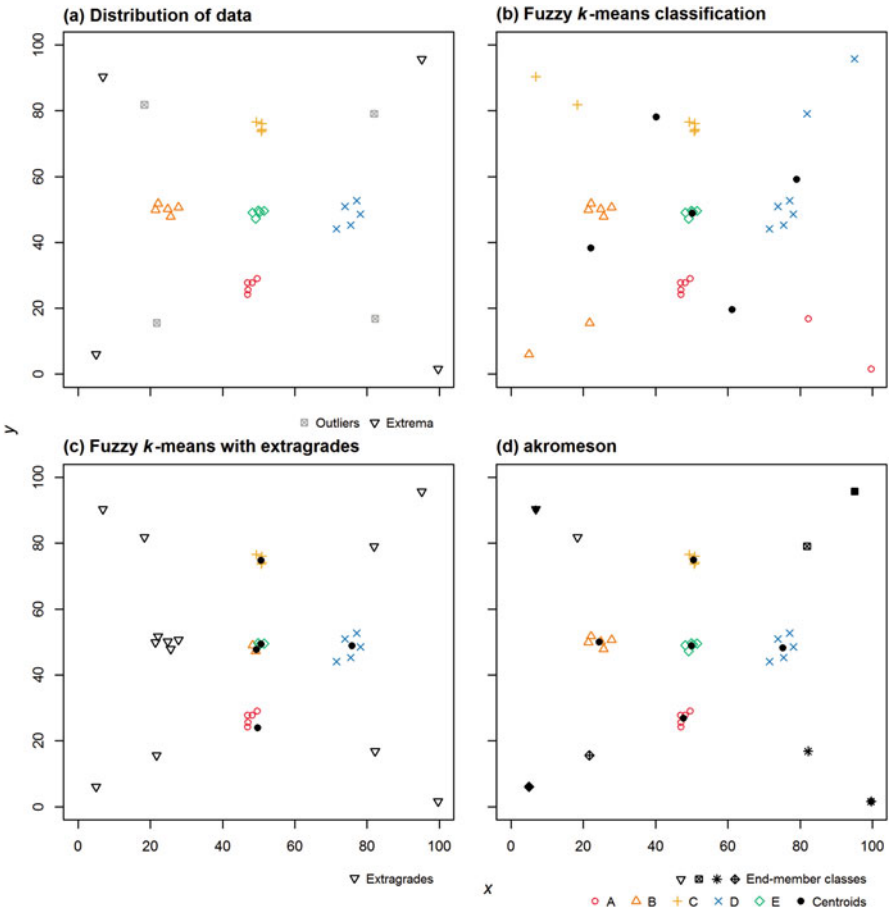


Fig. 8.4 Classification of a synthetic dataset (**a**) by several fuzzy variants of the *k*-means algorithm: (**b**) the fuzzy *k*-means algorithm, (**c**) fuzzy *k*-means with extragrades and (**d**) akromeson, after Hughes et al. (2014)

Fuzzy *k*-means with extragrades (de Gruijter and McBratney 1988; McBratney and de Gruijter 1992) is a modification to the standard fuzzy *k*-means algorithm that allows the modelling of an extragrade class of individuals. It attempts to overcome the weakness of the fuzzy *k*-means algorithm caused by outlying individuals. Although all individuals that do not possess a high degree of membership to any soil class may be termed intergrades, de Gruijter and McBratney (1988) distinguished intergrades located in the space *between* classes from those located in the outlying space. They are termed *intragrades* and *extragrades*, respectively. As failure to distinguish true outliers from mere intragrades may be misleading, de Gruijter and McBratney (1988) defined an extragrade class in which the memberships m_{i*} were made directly dependent on the distances to the class centroids. The membership and centroid-update equations are modified as follows (de Gruijter and McBratney 1988):

$$m_{ic} = \frac{d_{ic}^{-2/(\varphi-1)}}{\sum_{j=1}^{k} d_{ij}^{-2/(\varphi-1)} + \left(\frac{1-a}{a}\sum_{j=1}^{k} d_{ij}^{-2}\right)^{-1/(\varphi-1)}}, \; i = 1, \ldots, n; c = 1, \ldots, k \tag{8.16}$$

$$m_{i*} = \frac{\left(\frac{1-a}{a}\sum_{j=1}^{k} d_{ij}^{-2}\right)^{-1/(\varphi-1)}}{\sum_{j=1}^{k} d_{ij}^{-2/(\varphi-1)} + \left(\frac{1-a}{a}\sum_{j=1}^{k} d_{ij}^{-2}\right)^{-1/(\varphi-1)}}, \; i = 1, \ldots, n \tag{8.17}$$

$$\mathbf{c}_c = \frac{\sum_{i=1}^{n}\left\{m_{ic}^{\varphi} - \frac{1-a}{a} d_{ic}^{-4} m_{i*}^{\varphi}\right\}\mathbf{x}_i}{\sum_{i=1}^{n}\left\{m_{ic}^{\varphi} - \frac{1-a}{a} d_{ic}^{-4} m_{i*}^{\varphi}\right\}}, \; c = 1, \ldots, k. \tag{8.18}$$

The fuzzy objective function is modified accordingly:

$$J_{\mathrm{MG}}(\mathbf{M}, \mathbf{C}) = a\sum_{i=1}^{n}\sum_{c=1}^{k} m_{ic}^{\varphi} d_{ic}^{2} + (1-a)\sum_{i=1}^{n} m_{i*}^{\varphi}\sum_{c=1}^{k} d_{ic}^{-2} \tag{8.19}$$

The memberships to the extragrade class, m_{i*}, spread across regions at larger distances from the class centroids, unlike the memberships of the regular classes which tend to occupy fuzzy hyperspheres around the class centroids (McBratney and de Gruijter 1992).

The parameter α determines the mean extragrade membership; however, the function relating both quantities is generally unknown. Because of this, de Gruijter and McBratney (1988) estimated α empirically using a Regula-Falsi procedure. Lagacherie et al. (1997) estimated α by examination of a two-dimensional representation of the multivariate attribute data in conjunction with their expert knowledge of their study area.

8.2.4.4 Akromeson

The fuzzy k-means with extragrades algorithm enabled individuals lying in the outer parts of the attribute space to be recognised and placed into their own extragrade class. Doing so reduced the leverage the outlying points had on the formation of regular fuzzy classes in the more densely populated parts of the attribute space. Notwithstanding these advantages, estimation of the α parameter which determines the mean extragrade membership is not straightforward, and no clear procedure for doing so exists. Incorrect estimation of the extragrade class may lead to some extragrade individuals being treated as if they were individuals in the centre of the data and vice versa (Fig. 8.4c). In addition extragrades are placed into a single class regardless of their distribution in the attribute space. In reality there may be clusters of individuals in the outlying space, but the fuzzy k-means with extragrades algorithm is unable to resolve them if they exist.

Hughes et al. (2014) introduced the idea of end points, which are the individuals lying at the extremes of the attribute space (the points classified *extrema* in Fig. 8.4a). They can be detected as the vertices of a convex hull around the individuals in attribute space. Hughes et al. (2014) developed an algorithm they called *akromeson* that identifies end points and treats them as fixed centroids in a semi-supervised fuzzy *k*-means cluster analysis (Bensaid et al. 1996). A heuristic process is used to refine the number of end points, k_e, that are used as fixed centroids. The aim of the semi-supervised fuzzy *k*-means clustering algorithm, then, is to find additional clusters, k_d, in the attribute space so that $k = k_e + k_d$. The fuzziness exponent, φ, can be determined using the usual methods (Odeh et al. 1992).

The end result is a set of k classes that includes k_e end point classes (Fig. 8.4d). In Fig. 8.4d, $k_e = 4$ and $k_d = 5$ so $k = 4 + 5 = 9$. The end point classes supersede the single extragrade class completely and as such provide more information about the distribution of individuals in outlying parts of the attribute space. This extra information may be related to real environmental differences between akrograde classes that in turn may be important for managing land.

8.2.4.5 Fuzzy *c*-Numbers

The fuzzy *c*-numbers algorithm (Yang and Ko 1996) is an extension of fuzzy *k*-means clustering that allows the input attributes to be fuzzy numbers. As with the fuzzy *k*-means algorithm, the basic steps are similar to the standard *k*-means algorithm; however, unlike these two approaches, which cluster crisp attribute data, the fuzzy *c*-numbers algorithm clusters attribute data that are fuzzy.

Fuzzy attribute data are actually somewhat commonplace in soil material descriptions. For example, the pH value of a modal soil material representative might be described by a modal value and perceived lower and upper limits. Fuzzy attribute data can be represented using fuzzy numbers, which are fuzzy sets over the set of real numbers and are specified via membership functions. There are many different kinds of fuzzy numbers, but three of the most common are triangular, trapezoidal and Gaussian fuzzy numbers (Liu and Samal 2002a; Yang and Ko 1996), so-called due to the shape of their membership functions. Triangular fuzzy numbers may be symmetric or asymmetric, and we may estimate their parameters from a collection of soil profile measurements (Fig. 8.5). For example, say the distributions of pH (1:5 H_2O) in the 40–50 cm depth interval of a collection of Chromosols and Kurosols from the Lower Hunter Valley in New South Wales, Australia, take the values in Table 8.5. Kurosols are strongly acidic (pH (1:5 H_2O) less than 5.5) in the top 20 cm of the B_2 horizon, whereas Chromosols are not. A symmetric triangular fuzzy number $\tilde{x}_{ip}$ representing the pH may have a maximum membership at the mean pH—the so-called *apex*, x_{ip}, and minima of membership above and below this value at a distance defined by the *spread* S_{ip}, which we may choose to represent as two standard deviations from the mean pH value. Thus, a symmetric triangular fuzzy number $\tilde{x}_{ip}$ can be represented as (x_{ip}, S_{ip}).

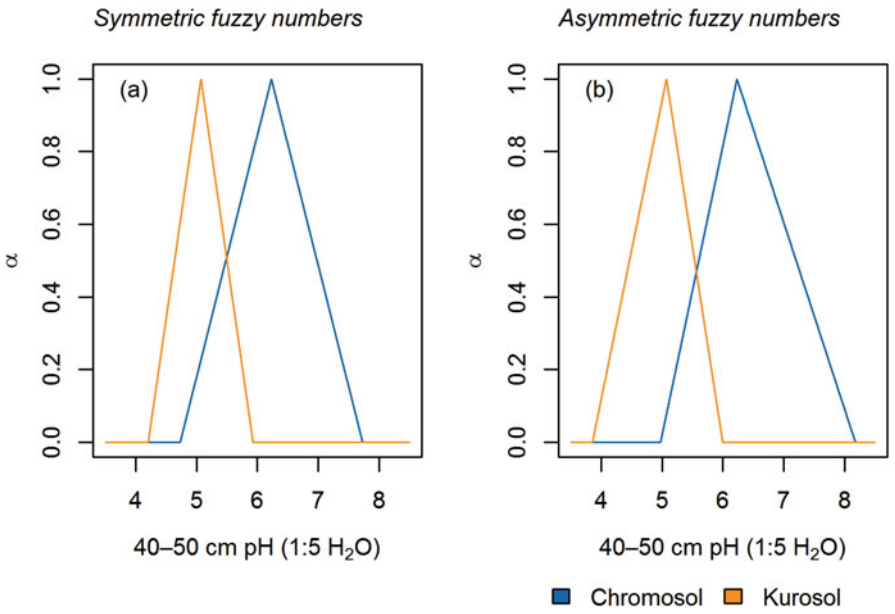


Fig. 8.5 Triangular numbers representing 40–50 cm pH (1:5 H_2O) for Chromosols and Kurosols in the Lower Hunter Valley, New South Wales, Australia

Table 8.5 Descriptive statistics of the distribution of pH (1:5 H_2O) for Chromosols and Kurosols observed at Pokolbin in the Lower Hunter Valley in New South Wales, Australia

	Chromosol	Kurosol
Mean	6.23	5.07
Standard deviation	0.75	0.43
Minimum	4.97	3.85
Maximum	8.18	6.00
n	57	28

On the other hand, the membership minima of an asymmetric triangular fuzzy number are not the same distance from the apex. We may choose to represent them as the observed maximum and minimum pH. An asymmetric triangular fuzzy number $\tilde{x}_{ip}$ can be represented as (x_{ip}, a_{ip}, b_{ip}) where a and b are the lower and upper minima of membership, respectively. Figure 8.5 demonstrates that the difference between symmetric and asymmetric triangular fuzzy numbers is relatively subtle if they are based on a distribution of values that is roughly normally distributed. For mathematical simplicity the symmetric fuzzy numbers are used hereforth.

A soil material description is usually represented as a vector of several attributes rather than a single attribute. A fuzzy vector of symmetric triangular fuzzy numbers can be defined as $\tilde{x}_i = (x_i, \mathbf{P}_{x_i})$ where x_i is the standard pattern vector $\mathbf{x}_i = \{x_{i1}, x_{i2}, \ldots, x_{id}\}$ and $\mathbf{P}_{xi}$ is the *panderance matrix* that contains the spread information (Celmiņš 1987):

$$\mathbf{P}_{x_i} = \begin{bmatrix} s_{i1}^2 & 0 & \cdots & 0 \\ 0 & s_{i2}^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & s_{id}^2 \end{bmatrix} \tag{8.20}$$

Liu and Samal (2002a) computed the distance between two fuzzy vectors $\tilde{a} = (a, \mathbf{P}_a)$ and $\tilde{b} = (b, \mathbf{P}_b)$ as

$$d^2\left(\tilde{a}, \tilde{b}\right) = \|a - b\|^2 + tr\left((\mathbf{P}_a - \mathbf{P}_b)^T (\mathbf{P}_a - \mathbf{P}_b)\right) \tag{8.21}$$

where *tr* is the trace of the matrix product of $(\mathbf{P}_a - \mathbf{P}_b)^T(\mathbf{P}_a - \mathbf{P}_b)$.

The cluster centres are given by $\{\tilde{c}_1, \tilde{c}_2, \tilde{c}_3, \ldots, \tilde{c}_k\}$ where $\tilde{c}_c = (c_c, \mathbf{P}_c)$. The membership of a fuzzy pattern $\tilde{x}_i$ in class c is computed as

$$m_c\left(\tilde{x}_i\right) = \frac{1}{\sum_{j=1}^{k} d^2\left(\tilde{x}_i, \tilde{c}_c\right) / d^2(\tilde{x}_i, \tilde{c}_k)^{1/(\varphi-1)}} \tag{8.22}$$

Update of the centroid patterns $\tilde{c}_j$ is a two-step process because both the apex vector $\mathbf{c}_c$ and panderance matrix $\mathbf{P}_{cj}$ need to be recomputed for each centroid. The apex vector is calculated as

$$\mathbf{c}_c = \frac{\sum_{i=1}^{n} m_c(\tilde{x}_i)^{\varphi} \mathbf{x}_i}{\sum_{i=1}^{n} m_c(\tilde{x}_i)^{\varphi}} \tag{8.23}$$

and the panderance matrix is calculated as

$$\mathbf{P}_c = \frac{\sum_{i=1}^{n} m_c(\tilde{x}_i)^{\varphi} \mathbf{P}_i}{\sum_{i=1}^{n} m_c(\tilde{x}_i)^{\varphi}} \tag{8.24}$$

Liu and Samal (2002a, b) used the fuzzy *c*-numbers algorithm to identify agroecozones in the state of Nebraska in the United States. Although we could not find evidence of the application of the fuzzy *c*-numbers algorithm to soil material classification, and only very infrequent application of fuzzy numbers in the soil classification literature more generally (e.g. Bhattacharya and Solomatine 2006), fuzzy numbers have found application in other aspects of soil science and related fields, including soil sampling (Lark 2000), engineering (Dodagoudar and Venkatachalam 2000; Saboya Jr. et al. 2006), hydrology (Dou et al. 1999; Schulz and Huwe 1999; Verma et al. 2009) and forestry (Kaya and Kahraman 2011). Considering soil attributes are often recorded and presented as uncertain quantities, it appears that the fuzzy *c*-numbers algorithm and others capable of handling fuzzy attributes (e.g. d'Urso and Giordani 2006) could have natural application in soil classification.

8.3 Cluster Validation

Cluster validation usually involves the selection of the partition corresponding to the optimal k from amongst several alternatives. Many cluster validity metrics have been devised that assist in making this decision. Several authors have performed extensive comparisons between a range of these metrics (e.g. Arbelaitz et al. 2013; Milligan and Cooper 1985) although only few have been used in the numerical classification of soil material.

Cluster validity metrics should ideally be independent of the essential parameters of the cluster analysis, such as n, k and φ (Roubens 1982). They can be classified into two categories: *membership-based measures* and *geometry-based measures* (Liu and Samal 2002b). Membership-based validity measures attack the problem of cluster validity by examining the fuzziness of a partition in the membership space. On the other hand, geometry-based cluster validity measures attempt to solve the problem of cluster validity by examining the separation of a partition in the attribute space. This is achieved by quantifying the shape and distribution of clusters with respect to their compactness and their separation from each other, which can be measured via the intra-cluster distance and the intercluster distance, respectively.

8.3.1 Membership-Based Measures

A range of membership-based measures are commonly used. Two of the simplest are the partition coefficient, F, and the partition entropy, H (Bezdek 1981):

$$F = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} m_{ij}^2 \tag{8.25}$$

$$H = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} m_{ij}\log\left(m_{ij}\right) \tag{8.26}$$

The partition coefficient measures the fuzziness of the fuzzy classes and ranges from 0.5, corresponding to the most fuzzy partition, to 1.0, corresponding to the least fuzzy partition (Bezdek 1981). Thus, F is maximised when the partition is hard. The partition entropy is inversely proportional to the goodness of the fuzzy classes, and as such a better partition is indicated when H is minimised (Liu and Samal 2002b).

Roubens (1982) stated that the search of an optimal value of k is complicated by the fact that F and H tend to increase and decrease with k, respectively. He proposed that the optimal k could be found more easily using the fuzziness performance index, FPI, and the normalised classification entropy, NCE.

The FPI is computed as (Odeh et al. 1992)

$$\mathrm{FPI} = 1 - (kF - 1) / (k - 1) \tag{8.27}$$

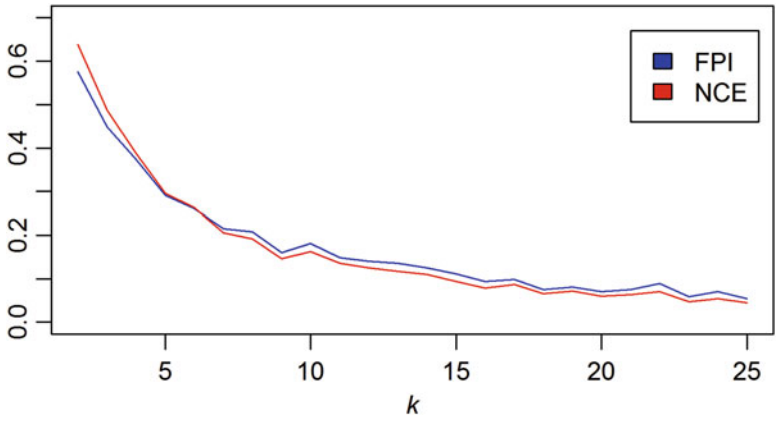


Fig. 8.6 Curves of fuzziness partition index and normalised classification entropy derived from runs of a fuzzy k-means clustering of soil material from the Lower Hunter Valley in New South Wales, Australia, using integer values of k from 2 to 25

The FPI varies between 0 and 1. Like the partition coefficient, it describes the fuzziness of a fuzzy partition. A low value of FPI implies that the continuous classes are relatively hard and that there is little sharing of membership between any pair of them. This suggests that there is a distinct natural partition structure amongst the individuals in the dataset (Odeh et al. 1992). The converse applies when FPI is high.

The *normalised classification entropy* (NCE) describes the uncertainty of the fuzzy partitioning of the individuals (Odeh et al. 1992):

$$\text{NCE} = \frac{H}{\log\ k} \tag{8.28}$$

where H is Bezdek's partition entropy (Bolliger and Mladenoff, 2005). The optimal value of k is usually found by determining a local minimum of both FPI and NCE (Odeh et al. 1992; Triantafilis et al. 2001). The set of classes at $k = 9$ satisfies this criterion in Fig. 8.6 for fuzzy clustering of soil material in Sect. 8.2.4.2.

Membership-based measures of cluster validity have been criticised on account of their lack of a direct connection to the geometry of the clusters (Xie and Beni 1991). Nevertheless, they remain relatively popular in pedometric research. While some soil researchers have used F and/or H (Burrough et al. 2001), most have used FPI and/or NCE (e.g. Bragato 2004; Cockx et al. 2007; Triantafilis et al. 2001; van Alphen and Stoorvogel 2000; Verheyen et al. 2001).

8.3.2 Geometry-Based Measures

Geometry-based measures can be classified as ratio-type and summation-type measures based on how the intra- and intercluster distances are combined (Kim and Ramakrishna 2005). Many of the more common geometry-based cluster validity metrics are of the ratio type, and several are based on the measure that Dunn originally devised (after Halkidi et al. 2001):

$$D = \min_{1\le i\le k}\left\{\min_{i+1\le j\le k-1}\left\{\frac{d\left(C_i, C_j\right)}{\max_{1\le c\le k}\left(\operatorname{diam}\left(C_c\right)\right)}\right\}\right\} \tag{8.29}$$

where $d(C_i, C_j)$ is the dissimilarity function between two clusters C_i and C_j, calculated as

$$d\left(C_i, C_j\right) = \min_{x\in C_i, y\in C_j}\left(d\left(\mathbf{x}, \mathbf{y}\right)\right) \tag{8.30}$$

and diam(Cc) is the diameter of cluster c, calculated as

$$\operatorname{diam}\left(C_c\right) = \max_{x,y\in C}\left(d\left(\mathbf{x}, \mathbf{y}\right)\right) \tag{8.31}$$

Dunn's index is primarily used to validate crisp classes. The numerator in the central term of Dunn's index pertains to the intercluster separation, whereas the denominator pertains to the intra-cluster dispersion. Large values of Dunn's index ought to indicate the presence of compact and well-separated clusters (Halkidi et al. 2001). The Dunn's index has only rarely been used in the pedometric literature (Ließ 2015).

The *compactness and separation validity function*, S (Xie and Beni 1991), is a ratio-type measure of cluster validity that has been used to validate fuzzy clusters. It has been used from time to time in the pedometric and related literature (e.g. Odgers et al. 2011; Sun et al. 2012; Vrindts et al. 2005). It is calculated as

$$S = \frac{\sum_{j=1}^{k}\sigma_j}{nd_{\min}^2} \tag{8.32}$$

where σ_j is the sum of squares of the fuzzy deviation, $m_{ij}\|\mathbf{x}_i - \mathbf{c}_j\|^2$, of individual i from centroid j:

$$\sigma_j = \sum_{i=1}^{n} m_{ij}\left\|\mathbf{x}_i - \mathbf{c}_j\right\|^2 \tag{8.33}$$

The term $\|\mathbf{x}_i - \mathbf{c}_j\|$ is simply the Euclidean distance between individual i and centroid j. σ_j is a measure of non-compactness: that is, the higher the value of σ_j, the further from the centroid are the members of class j (Liu and Samal 2002b).

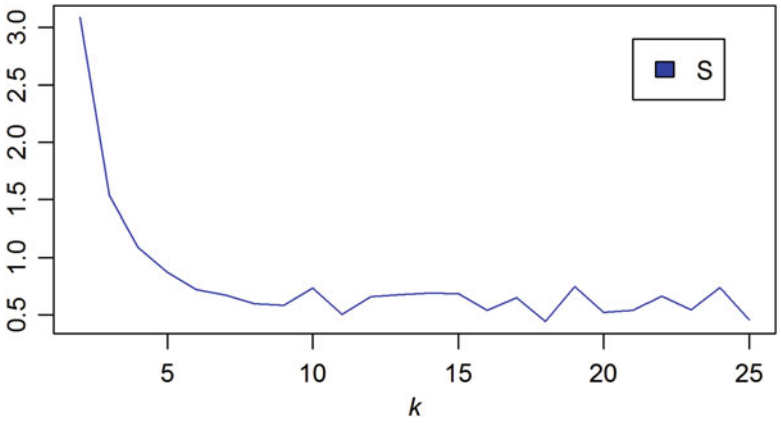


Fig. 8.7 Curve of the compactness and separation validity function, S, derived from runs of a fuzzy k-means clustering of soil material from the Lower Hunter Valley in New South Wales, Australia, using integer values of k between 2 and 25

Finally, $d^2_{\min}$ is the *separation* of the fuzzy partition and is calculated as

$$d^2_{\min} = \min \left\| \mathbf{c}_i - \mathbf{c}_j \right\|^2 \tag{8.34}$$

which is effectively the square of the minimum Euclidean distance between cluster centroids. A larger value of $d^2_{\min}$ indicates that all the clusters are well separated in the feature space (Xie and Beni 1991).

As with other cluster validity criteria, local minima in a curve of S across several values of k may indicate a locally optimal partition. Thus, Fig. 8.7 indicates that an optimal partition of the Lower Hunter Valley soil material in Sect. 8.2.4.2 may occur when $k = 11$.

Xie and Beni (1991) note that S is meaningless when c is very large and close to n because of its tendency to monotonically decrease in these circumstances. In practice this is not often a problem because the values of c that we are interested in are usually much lower than n.

8.4 Allocation to Pre-existing Classes

The placement of new individuals into a class of some classification system is known as *allocation*, *identification* or *diagnosis* (McBratney 1994) although the term *classification* has often been misappropriated to refer to the same process. The act of allocating an individual to a class implies that a classification system exists a priori. In soil science, considerably less has been written about allocation than classification although research has been performed for decades (Norris and

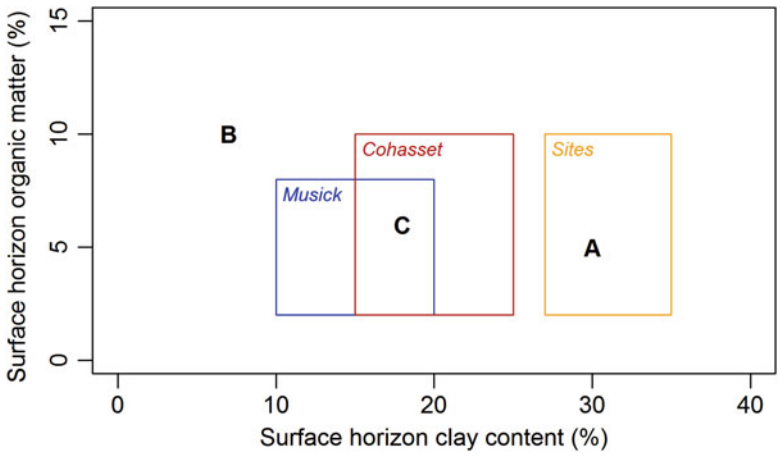


Fig. 8.8 Two-dimensional example of parallelepiped allocation. Parallelepipeds established using class limits for surface horizon clay content (%) and organic matter (%) of several soil series from El Dorado County in California, United States. Data retrieved from SSURGO database (Soil Survey Staff 2016)

Loveday 1971). Like classification itself, allocation can be crisp or fuzzy. Several methods are available, including parallelepiped methods, discriminant analysis and classification trees. Not all have been commonly applied in pedometric research.

8.4.1 Parallelepiped Method

Parallelepiped allocation is one of the most basic methods of allocation. It is sometimes also known as the *box decision rule* or the *level slice procedure* (Campbell 1996). For each of the k classes in a classification system, a parallelepiped is set up in the d-dimensional feature space using the class limits of the d attributes to set its boundaries. Alternatively the standard deviations of the attributes may also be used to set their boundaries. We illustrate with a simple two-dimensional example in Fig. 8.8 using class limits. It should be easy to see that classes are implicitly hypercubic or hypertrapezoidal in many dimensions. Allocation of an individual to class j is simply a matter of identifying which parallelepiped the individual is located inside of. Parallelepiped allocation is therefore crisp. For example, in Fig. 8.8, soil material A with surface horizon clay content of 30% and surface horizon organic matter content of 5% is clearly a member of the Sites series.

Mather and Koch (2011) identify difficulties that occur in two extreme cases. First, an individual may not lie inside any of the k parallelepipeds, making allocation impossible (B in Fig. 8.8). Second, an individual may lie inside the overlapping parallelepipeds of more than one class, in which case a protocol must be in place to determine which class the individual should be allocated to (C in Fig. 8.8). In

both these cases, more information (i.e. more attributes) is necessary to resolve the appropriate class, if any.

Parallelepiped classification has been commonly used to classify land cover from remotely sensed imagery (e.g. Goodenough and Shlien 1974; Jensen 1978; Robinove 1979) but appears not to have been used in pedometric studies.

8.4.2 Minimum Distance Method

The minimum distance method allocates an individual to the class of the centroid to which it has the shortest distance (i.e. is most similar to) in the d-dimensional attribute space. Like the parallelepiped method, allocation is a fairly simple process. Since the Euclidean or Mahalanobis distances are usually used to relate individuals to centroids, classes are implicitly hyperspherical or hyperellipsoidal in the attribute space, respectively. Minimum distance allocation may not be possible if classes do not have centroids.

8.4.3 Discriminant Analysis

According to the principle of discriminant analysis (Fisher 1936), a linear function of the attributes of a population of individuals belonging to two classes can be found that best discriminates between the two classes. Such a function can be represented as

$$X = a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_d x_d \tag{8.35}$$

The coefficients of the linear function are chosen so as to maximise the distance of separation between the class means (i.e. the class centroids) relative to the within-class variability (Healy 1965). In matrix terms, they can be found by multiplying the vector of separation distances $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ by $\mathbf{S}^{-1}$, the inverse of the variance-covariance matrix of the attributes of the sample of individuals (Blackith and Reyment 1971):

$$\mathbf{a} = \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \tag{8.36}$$

A linear function so found is known as a *discriminant function*. The discriminant function is ideally calibrated on representative individuals chosen randomly from members of each class (Webster 1977). A larger value of the function, when evaluated, implies a clearer separation of a pair of classes than does a smaller value. This brings us to an important point: as Blackith and Reyment (1971) point out, without including S^{-1} in the computation of the discriminant coefficients, continued weighted addition of attributes would cause X to increase indefinitely and thus give a false impression of the degree of separation between the pair of classes.

In other words, each attribute would contribute information about the separation of the classes without respect to the information provided by the other attributes. The incorporation of the correlations between the attributes in the computation of **a** constrains the size of X and ensures that each attribute only contributes the amount of information that is unique to it.

In discriminant analysis the attribute distributions of each class are assumed to be multivariate normally distributed and that the standard deviations are equal. Oftentimes these assumptions are unrealistic in practice, and the technique appears to be sufficiently robust against mild departures from them (Blackith and Reyment 1971).

Rao (1948) adapted the method for $k > 2$ classes, in which case there are always $k - 1$ discriminant functions. Healy (1965), Blackith and Reyment (1971) and Webster and Burrough (1974) are amongst those who have described discriminant analysis in geometric terms. Consider two bivariate classes whose distributions are plotted as ellipses in the *discriminant space* defined by their attribute axes. The classes are best separated by a line that passes through the intersection of the two ellipses. The axis drawn orthogonal to such a line has been called the *discriminant axis* and is the best axis for discriminating between the two classes (Webster and Burrough 1974). The concept can of course be extended into as many attribute dimensions d as are necessary, in which case the ellipses become hyperplanes (Webster and Oliver 1990).

In terms of allocation, distances in the discriminant space are Mahalanobis distances. Thus, new individuals may be allocated to the class to whose centroid it is closest in the Mahalanobis sense (Webster and Burrough 1974).

Discriminant analysis was first employed in soil science by Cox and Martin (1937), who used it to quantify the significance of several soil properties for predicting the presence of *Azotobacter*. It has been used subsequently by Hughes and Lindley (1955), Oertel (1961), Norris and Loveday (1971) and Webster and Burrough (1974), amongst others, for the purpose of allocation. Triantafilis et al. (2003) generalised the theory to a fuzzy linear discriminant analysis.

8.4.3.1 Classification Trees

Classification trees are a hierarchical non-parametric example of supervised classification as they require the set of classes to be known a priori. Because the classes are known a priori, they are readily useful for class allocation. Classification trees can be formulated algorithmically ab initio or by manual extraction of the classification rules in existing classification systems (see Fig. 8.9 for an example). Classification trees recursively subdivide a population of individuals into ever more specific subgroups. Points of subdivision are called nodes, and subdivision is conducted by performing a logical test on the threshold of some attribute. Attributes may be continuous or discrete valued. The logical test is formalised in a *decision rule* that also determines the subgroup implied by passing or failing the logical test. A terminal subgroup is called a leaf and contains a single class rather than a pointer

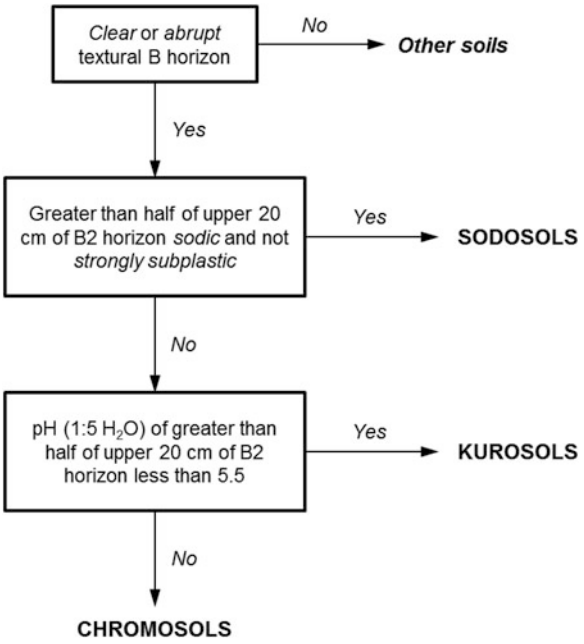


Fig. 8.9 Classification tree for orders of the Australian Soil Classification that possess a clear or abrupt textural B horizon

to a subsequent decision rule. Decision rules are analogous to the if-then rules that are ubiquitous in software programming and may be expressed accordingly. For example, a decision rule may be as follows:

If clay content $\geq$35%: subgroup A
Else subgroup B

When formulated algorithmically, decision rules are usually chosen to optimise some measure of goodness in the subgroups formed by the split, such as within-subgroup purity. According to Friedl and Brodley (1997), a range of metrics can be used to quantify how well this is done; for example, Lagacherie and Holmes (1997) used the Gini index. Pruning may be conducted in order to reduce the size of the tree and to avoid overfitting to the calibration data. Pruning involves merging pairs of leaves and may be accomplished by a cross-validation procedure (Scull et al. 2003).

Classification trees have several advantages over other approaches (Friedl and Brodley 1997). For example, because they are non-parametric, they are insensitive to the distributions of attribute values. They can handle nonlinear relationships between classes and attribute values, and they are able to handle both continuous and discrete-valued attributes. Finally, the tree structure is readily interpretable.

Lagacherie and Holmes (1997) were one of the first to apply classification trees in soil survey, and since then they have become a popular supervised classification tool. It appears, though, that their most frequent use has been in soil survey and mapping rather than in classification of soil material.

8.4.4 Fuzzy k-Means

McBratney (1994) demonstrated how individuals can be allocated to soil classes created using the fuzzy k-means with extragrades algorithm. Allocation is a matter of calculating the degree of membership of the unknown individual to each of the fuzzy classes using Eq. 8.16 and to the extragrade class using Eq. 8.17 described earlier. In order to do this, the parameters k, φ and α must be known, along with the matrix of class centroids and the variance-covariance matrix of the samples used to produce the classification. Furthermore, the Mahalanobis distances between the new individual and the class centroids must also be calculated.

Such an allocation scheme is not exclusively applicable to fuzzy soil classes produced by a numerical soil classification procedure. For example, Mazaheri et al. (1995) employed the technique in order to allocate new profiles to the classes of the Australian Great Soil Groups classification scheme (Stace et al. 1968) Although the Great Soil Groups classification scheme is not a numerical soil classification system, the classes are considered by some (Moore et al. 1983) to be fuzzy in the sense that they are described by a central concept, and, since no taxonomic key has been devised for the purpose of allocation, the class boundaries are somewhat vague.

Despite the fact that the classification scheme is likely to be obsolete now, as the universe of soils it caters for is too small for all Australian conditions (Moore et al. 1983), Mazaheri et al. (1995) reported positive results when they used the fuzzy allocation scheme to allocate six individuals to a class. They determined that the numerical allocation system was not only useful for allocating individuals to classes with partial membership but that it also enables a critical review of the existing classification system in at least two situations: (i) if profiles are allocated with more-or-less equal membership to several classes and (ii) if profiles possess a large extragrade membership.

8.4.5 Soil Horizon Classes

Soil scientists have long recognised soil horizons as the fundamental building blocks of the soil profile. For example, of the eight criteria Marbut (1920) proposed as grounds for the differentiation of soil profiles, seven related to various characteristics of soil horizons.

Attempts have been made to classify soil horizons since the late nineteenth century (FitzPatrick 1967). Dokuchaev was the first to use the A-B-C-horizon nomenclature in his description of Chernozem soils. By itself, though, the A-B-C notation does not convey a lot of information about the character of soil material belonging to a given horizon—or to put it more optimistically, the diversity of soil horizons is greater than can be readily captured by the A-B-C notation (Nikiforoff 1931). For example, a Vertosol B horizon has a very different character to a Podosol B horizon. Horizon subscripts were developed to provide more information in

this respect although the precise meaning of them can vary from jurisdiction to jurisdiction. Bridges (1990) describes a range of other difficulties with the A-B-C system.

The USDA devised a system of diagnostic horizons during its development of the Soil Taxonomy classification system (Soil Survey Staff 1999). Other soil classifications also use their own versions of these horizons (e.g. FAO 2014; Hewitt 2010). Diagnostic horizons are classes of horizons that possess specific soil material characteristics. Their descriptions are often lengthy and complex. They were developed as aids to soil profile classification because specific soil profile classes often require the presence of specific diagnostic horizons. Despite their utility, diagnostic horizons have not been without criticism. For example, FitzPatrick (1976) points out that uncertainty arises in situations where a soil profile contains more than one diagnostic horizon: because the profile is possibly eligible for allocation to more than one soil profile class, a judgement must be made on which diagnostic horizon is more important (or more diagnostic!).

Researchers have since developed more quantitative systems. The best-known and most comprehensive is that of FitzPatrick (e.g. 1993, 1988, 1976, 1967), who devised a system of about 81 classes of soil horizons, which he later called *segments*. Each segment was given a name ending in *-on* and a two-letter code, much like the chemical elements. The codes for a profile's horizons could be assembled to produce a code for the entire profile much like a chemical formula.

Researchers have also used fuzzy numerical classification to create soil layer classes. For example Powell et al. (1992) used the fuzzy *k*-means with extragrades algorithm (de Gruijter and McBratney 1988) to create a set of fuzzy soil layer classes for their study area in the Lockyer Valley in Queensland, Australia. Triantafilis et al. (2001) carried out a similar analysis for soils from the Edgeroi district in the Namoi Valley of New South Wales, Australia. In both cases the researchers found that the numerical soil layer classes were well able to explain pedological and landscape features in their respective study areas.

References

Anderson AJB (1971) Numeric examination of multivariate soil samples. Math Geol 3:1–14

Arbelaitz O, Gurrutxaga I, Muguerza J, Pérez JM, Perona I (2013) An extensive comparative study of cluster validity indices. Pattern Recogn 46:243–256. doi:10.1016/j.patcog.2012.07.021

Arkley RJ (1976) Statistical methods in soil classification research. Adv Agron 28:37–70. doi:10.1016/S0065-2113(08)60552-0

Banin A, Amiel A (1970) A correlative study of the chemical and physical properties of a group of natural soils of Israel. Geoderma 3:185–198. doi:10.1016/0016-7061(70)90018-2

Beaudette DE, Roudier P, O'Geen AT (2013) Algorithms for quantitative pedology: a toolkit for soil scientists. Comput Geosci 52:258–268. doi:10.1016/j.cageo.2012.10.020

Bensaid AM, Hall LO, Bezdek JC, Clarke LP (1996) Partially supervised clustering for image segmentation. Pattern Recogn 29:859–871. doi:10.1016/0031-3203(95)00120-4

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms, advanced applications in pattern recognition. Plenum Press, New York

Bhattacharya B, Solomatine DP (2006) Machine learning in soil classification. Neural Netw 19:186–195. doi:10.1016/j.neunet.2006.01.005
Bidwell OW, Hole FD (1964a) Numerical taxonomy and soil classification. Soil Sci 97:58–62
Bidwell OW, Hole FD (1964b) An experiment in the numerical classification of some Kansas soils. Soil Sci Soc Am Proc 28:263–268
Blackith RE, Reyment RA (1971) Multivariate morphometrics. Academic Press, London
Bolliger J, Mladenoff DJ (2005) Quantifying spatial classification uncertainties of the historical Wisconsin landscape (USA). Ecography 28:141–156
Bormann H (2010) Towards a hydrologically motivated soil texture classification. Geoderma 157:142–153. doi:10.1016/j.geoderma.2010.04.005
Bouma J (1985) Soil variability and soil survey. In: Nielsen DR, Bouma J (eds) Soil spatial variability: proceedings of a Workshop of the ISSS and SSSA, Las Vegas, USA, 30 November–1 December 1984. Pudoc, Wageningen, pp 130–149
Bragato G (2004) Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. Geoderma 118:1–16. doi:10.1016/S0016-7061(03)00166-6
Bridges EM (1990) Soil horizon designations (No. Technical Report 19). International Soil Reference and Information Centre, Wageningen
Brus, D.J., de Gruijter, J.J., van Groenigen, J.W., 2006. Designing spatial coverage samples using the k-means clustering algorithm, in: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital soil mapping—an introductory perspective, developments in soil science. Elsevier, pp. 183–192
Bui EN, Moran CJ (2001) Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103:79–94. doi:10.1016/S0016-7061(01)00070-2
Buol, S.W., 2003. Philosophies of soil classifications: from is to does, in: Eswaran, H., Rice, T., Ahrens, R., Stewart, B.A. (Eds.), Soil classification: a global desk reference. CRC Press LLC, pp. 3–10
Burrough PA (1989) Fuzzy mathematical methods for soil survey and land evaluation. J Soil Sci 40:477–492. doi:10.1111/j.1365-2389.1989.tb01290.x
Burrough PA, van Gaans PFM, Hootsmans R (1997) Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77:115–135. doi:10.1016/S0016-7061(97)00018-9
Burrough PA, van Gaans PFM, MacMillan RA (2000) High-resolution landform classification using fuzzy k-means. Fuzzy Sets Syst 113:37–52. doi:10.1016/S0165-0114(99)00011-1
Burrough PA, Wilson JP, van Gaans PFM, Hansen AJ (2001) Fuzzy k-means classification of topoclimatic data as an aid to forest mapping in the greater Yellowstone area, USA. Landsc Ecol 16:523–546. doi:10.1023/A:1013167712622
Campbell JB (1996) Introduction to remote sensing, 2nd edn. The Guilford Press, New York
Campbell NA, Mulcahy MJ, McArthur WM (1970) Numerical classification of soil profiles on the basis of field morphological properties. Aust J Soil Res 8:43–58. doi:10.1071/SR9700043
Celmiņš A (1987) Least squares model fitting to fuzzy vector data. Fuzzy Sets Syst 22:245–269. doi:10.1016/0165-0114(87)90070-4
Clifford HT, Williams WT (1976) Similarity measures. In: Williams WT (ed) Pattern analysis in agricultural science. CSIRO, East Melbourne, pp 37–46
Cline MG (1949) Basic principles of soil classification. Soil Sci 67:81–92
Cockx L, van Meirvenne M, de Vos B (2007) Using the EM38DD soil sensor to delineate clay lenses in a sandy forest soil. Soil Sci Soc Am J 71:1314–1322. doi:10.2136/sssaj2006.0323
Cox GM, Martin WP (1937) Use of a discriminant function for differentiating soils with different Azotobacter populations. Iowa State Coll J Sci 11:323–331
D'Urso P, Giordani P (2006) A weighted fuzzy c-means clustering model for fuzzy data. Computational Statistics & Data Analysis 50:1496–1523. doi:10.1016/j.csda.2004.12.002
Dale MB, McBratney AB, Russell JS (1989) On the role of expert systems and numerical taxonomy in soil classification. J Soil Sci 40:223–234. doi:10.1111/j.1365-2389.1989.tb01268.x

de Bruin S, Stein A (1998) Soil-landscape modelling using fuzzy c-means clustering of attribute data derived from a digital elevation model (DEM). Geoderma 83:17–33. doi:10.1016/S0016-7061(97)00143-2

de Gruijter JJ, McBratney AB (1988) A modified fuzzy k-means method for predictive classification. In: Bock HH (ed) Classification and related methods of data analysis. Elsevier Science Publishers B.V, Amsterdam, pp 97–104

de Gruijter JJ, Walvoort DJJ, van Gaans PFM (1997) Continuous soil maps—a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77:169–195. doi:10.1016/S0016-7061(97)00021-9

Dobermann A, Witt C, Abdulrachman S, Gines HC, Nagarajan R, Son TT, Tan PS, Wang GH, Chien NV, Thoa VTK, Phung CV, Stalin P, Muthukrishnan P, Ravi V, Babu M, Simbahan GC, Adviento MAA (2003) Soil fertility and indigenous nutrient supply in irrigated rice domains of Asia. Agron J 95:913–923. doi:10.2134/agronj2003.9130

Dodagoudar GR, Venkatachalam G (2000) Reliability analysis of slopes using fuzzy sets theory. Comput Geotech 27:101–115. doi:10.1016/S0266-352X(00)00009-4

Dou C, Woldt W, Bogardi I (1999) Fuzzy rule-based approach to describe solute transport in the unsaturated zone. J Hydrol 220:74–85. doi:10.1016/S0022-1694(99)00065-7

Dubes RC (1987) How many clusters are best?—an experiment. Pattern Recogn 20:645–663. doi:10.1016/0031-3203(87)90034-3

Estabrook GF (1967) An information theory model for character analysis. Taxon 16:86–97. doi:10.2307/1216888

Falkenauer, E., Marchand, A., 2001. Using k-means? Consider ArrayMiner. Presented at the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, Las Vegas

FAO (2014) World reference base for soil resources (No. 106), World soil resources reports. Food and Agriculture Organisation of the United Nations, Rome

Fidêncio PH, Ruisánchez I, Poppi RJ (2001) Application of artificial neural networks to the classification of soils from São Paulo state using near-infrared spectroscopy. Analyst 126:2194–2200. doi:10.1039/B107533K

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7:179–188

Fisher L, van Ness JW (1971) Admissible clustering procedures. Biometrika 58:91–104. doi:10.1093/biomet/58.1.91

FitzPatrick EA (1967) Soil nomenclature and classification. Geoderma 1:91–105

FitzPatrick EA (1976) Soil horizons and homology. Classif Soc Bull 3:68–89

FitzPatrick EA (1988) Soil horizon designation and classification (Technical Paper No. 17), Technical paper. International Soil Reference and Information Centre, Wageningen

FitzPatrick EA (1993) Principles of soil horizon definition and classification. Catena 20:395–402. doi:10.1016/S0341-8162(05)80005-0

Friedl MA, Brodley CE (1997) Decision tree classification of land cover from remotely sensed data. Remote Sens Environ 61:399–409. doi:10.1016/S0034-4257(97)00049-7

Grigal DF, Arneman HF (1969) Numerical classification of some forested Minnesota soils. Soil Sci Soc Am Proc 33:433–438. doi:10.2136/sssaj1969.03615995003300030029x

Gibbons, F.R., 1968. Limitations to the usefulness of soil classification. Presented at the 9th International Congress of Soil Science, Adelaide, South Australia, pp. 159–167

Goodenough, D., Shlien, S., 1974. Results of cover-type classification by maximum likelihood and parallelepiped methods, in: Proceedings of the Second Canadian Symposium on Remote Sensing. Presented at the Second Canadian Symposium on Remote Sensing, Canadian Remote Sensing Society, University of Guelph, pp. 136–164

Gower JC (1967) A comparison of some methods of cluster analysis. Biometrics 23:623–637. doi:10.2307/2528417

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857–874

Halkidi M, Batistakis Y, Vazirgiannis M (2001) On clustering validation techniques. J Intell Inf Syst 17:107–145. doi:10.1023/A:1012801612483
Hartigan JA (1975) Clustering algorithms, probability and mathematical statistics. Wiley, New York
Hartigan JA, Wong MA (1979) Algorithm AS 136: a k-means clustering algorithm. J R Stat Soc Ser C (Appl Stat) 28:100–108
Healy, M.J.R., 1965. Descriptive uses of discriminant functions. In: Mathematics and computer science in biology and medicine: Proceedings of a Conference Held by Medical Research Council in Association with the Health Departments, Oxford, July 1964. Her Majesty's Stationery Office, London, 93–102
Heuvelink GBM, Burrough PA (1993) Error propagation in cartographic modelling using Boolean logic and continuous classification. Int J Geogr Inf Syst 7:231–246. doi:10.1080/02693799308901954
Hewitt AE (2010) New Zealand soil classification, Landcare Research Science Series, vol 1, 3rd edn. Manaaki Whenua Press, Lincoln. doi:10.7931/DL1-LRSS-1-2010
Hole FD, Hironaka M (1960) An experiment in ordination of some soil profiles. Soil Sci Soc Am Proc 24:309–312
Holmgren, G.G.S., 1986. The soil individual. Presented at the Proceedings of the 13th World Congress of Soil Science, Hamburg, pp. 1146–1147
Holmgren GGS (1988) The point representation of soil. Soil Sci Soc Am J 52:712–716
Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441. doi:10.1037/h0071325
Hughes RE, Lindley DV (1955) Application of biometric methods to problems of classification in ecology. Nature 175:806–807. doi:10.1038/175806a0
Hughes PA, McBratney AB, Minasny B, Campbell S (2014) End members, end points and extragrades in numerical soil classification. Geoderma 226–227:365–375. doi:10.1016/j.geoderma.2014.03.010
Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270. doi:10.5169/seals-268384
Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recogn Lett 31:651–666. doi:10.1016/j.patrec.2009.09.011
Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv 31:264–323. doi:10.1145/331499.331504
Jensen JR (1978) Digital land cover mapping using layered classification logic and physical composition attributes. Am Cartographer 5:121–132. doi:10.1559/152304078784022917
Kantardzic, M., 2011. Preparing the data, data mining: concepts, models, methods, and algorithms. Wiley, Hoboken, 26–52
Kaya T, Kahraman C (2011) Fuzzy multiple criteria forestry decision making based on an integrated VIKOR and AHP approach. Expert Syst Appl 38:7326–7333. doi:10.1016/j.eswa.2010.12.003
Kim M, Ramakrishna RS (2005) New indices for cluster validity assessment. Pattern Recogn Lett 26:2353–2363. doi:10.1016/j.patrec.2005.04.007
Knox EG (1965) Soil individuals and soil classification. Soil Sci Soc Am Proc 29:79–84
Kyuma K, Kawaguchi K (1976) Soil material classification for paddy soils in Japan. Soil Sci Plant Nutr 22:111–124. doi:10.1080/00380768.1976.10432973
Lagacherie P, Holmes S (1997) Addressing geographical data errors in a classification tree for soil unit prediction. Int J Geogr Inf Sci 11:183–198. doi:10.1080/136588197242455
Lagacherie P, Cazemier DR, van Gaans PFM, Burrough PA (1997) Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area. Geoderma 77:197–216. doi:10.1016/S0016-7061(97)00022-0
Lark RM (2000) Designing sampling grids from imprecise information on soil variability, an approach based on the fuzzy kriging variance. Geoderma 98:35–59. doi:10.1016/S0016-7061(00)00051-3

Leeper GW (1956) The classification of soils. J Soil Sci 7:59–64. doi:10.1111/j.1365-2389.1956.tb00861.x

Leone AP, Sommer S (2000) Multivariate analysis of laboratory spectra for the assessment of soil development and soil degradation in the southern Apennines (Italy). Remote Sens Environ 72:346–359. doi:10.1016/S0034-4257(99)00110-8

Ließ M (2015) Sampling for regression-based digital soil mapping: closing the gap between statistical desires and operational applicability. Spat Stat 13:106–122. doi:10.1016/j.spasta.2015.06.002

Liu M, Samal A (2002a) A fuzzy clustering approach to define agroecozones. Ecol Model 149:215–228. doi:10.1016/S0304-3800(01)00446-X

Liu M, Samal A (2002b) Cluster validation using legacy delineations. Image Vis Comput 20:459–467

Lloyd SP (1982) Least squares optimization in PCM. IEEE Trans Inf Theory 28:129–137. doi:10.1109/TIT.1982.1056489

MacVicar CN (1969) A basis for the classification of soil. J Soil Sci 20:141–152. doi:10.1111/j.1365-2389.1969.tb01563.x

Mahalanobis PC (1936) On the generalized distance in statistics. Proc Natl Inst Sci India 2:49–55

Mallavan, B.P., Minasny, B., McBratney, A.B., 2010. Homosoil, a methodology for quantitative extrapolation of soil information across the globe, In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S., McBratney, A.B. (Eds.), Digital soil mapping: bridging research, environmental application, and operation, progress in soil science. Springer Science+Business Media B.V., pp. 137–149. doi:10.1007/978-90-481-8863-5_12

Mao J, Jain AK (1996) A self-organizing network for hyperellipsoidal clustering (HEC). IEEE Trans Neural Netw 7:16–29. doi:10.1109/72.478389

Marbut, C.F., 1920. The contribution of soil surveys to soil science, In: Proceedings of the Forty-First Annual Meeting of the Society for the Promotion of Agricultural Science. Presented at the Forty-First Annual Meeting of the Society for the Promotion of Agricultural Science, Springfield, Massachusetts, pp. 116–142

Mather PM, Koch M (2011) Computer processing of remotely-sensed images: an introduction, 4th edn. Wiley, Chichester

Mazaheri SA, Koppi AJ, McBratney AB (1995) A fuzzy allocation scheme for the Australian great soil groups classification system. Eur J Soil Sci 46:601–612

McBratney AB (1994) Allocation of new individuals to continuous soil classes. Aust J Soil Res 32:623–633. doi:10.1071/SR9940623

McBratney AB, de Gruijter JJ (1992) A continuum approach to soil classification by modified fuzzy k-means with extragrades. J Soil Sci 43:159–175. doi:10.1111/j.1365-2389.1992.tb00127.x

McBratney AB, Moore AW (1985) Application of fuzzy sets to climatic classification. Agric For Meteorol 35:165–185. doi:10.1016/0168-1923(85)90082-6

McBratney AB, de Gruijter JJ, Brus DJ (1992) Spacial prediction and mapping of continuous soil classes. Geoderma 54:39–64. doi:10.1016/0016-7061(92)90097-Q

McKenzie NJ, Austin MP (1993) A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. Geoderma 57:329–355. doi:10.1016/0016-7061(93)90049-Q

Metternicht GI (2003) Categorical fuzziness: a comparison between crisp and fuzzy class boundary modelling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios. Ecol Model 168:371–389. doi:10.1016/S0304-3800(03)00147-9

Milligan GW, Cooper MC (1985) An examination of the procedures for determining the number of clusters in a data set. Psychometrika 50:159–179. doi:10.1007/BF02294245

Minasny B, McBratney AB (2006) Mechanistic soil-landscape modelling as an approach to developing pedogenetic classifications. Geoderma 133:138–149. doi:10.1016/j.geoderma.2006.03.042

Moore AW, Russell JS (1967) Comparison of coefficients and grouping procedures in numerical analysis of soil trace element data. Geoderma 1:139–158. doi:10.1016/0016-7061(67)90006-7

Moore, A.W., Isbell, R.F., Northcote, K.H., 1983. Classification of Australian soils, In: Soils: an Australian viewpoint. CSIRO, Melbourne, pp. 253–266

Muir JW, Hardie HGM, Inkson RHE, Anderson AJB (1970) The classification of soil profiles by traditional and numerical methods. Geoderma 4:81–90. doi:10.1016/0016-7061(70)90035-2

Nikiforoff CC (1931) History of A, B, C. Bull Am Soil Surv Assoc 12:67–70

Norris JM (1971) The application of multivariate analysis to soil studies. I Grouping of soils using different properties. J Soil Sci 22:69–80. doi:10.1111/j.1365-2389.1971.tb01594.x

Norris JM, Loveday J (1971) The application of multivariate analysis to soil studies. II The allocation of soil profiles to established groups: a comparison of soil survey and computer method. J Soil Sci 22:395–400. doi:10.1111/j.1365-2389.1971.tb01625.x

Odeh IOA, McBratney AB, Chittleborough DJ (1990) Design of optimal sample spacings for mapping soil using fuzzy-k-means and regionalized variable theory. Geoderma 47:93–122. doi:10.1016/0016-7061(90)90049-F

Odeh IOA, McBratney AB, Chittleborough DJ (1992) Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. Soil Sci Soc Am J 56:505–516. doi:10.2136/sssaj1992.03615995005600020027x

Odgers NP, McBratney AB, Minasny B (2011) Bottom-up digital soil mapping. I Soil layer classes. Geoderma 163:38–44. doi:10.1016/j.geoderma.2011.03.014

Oertel AC (1961) Chemical discrimination of terra rossas and rendzinas. J Soil Sci 12:111–118. doi:10.1111/j.1365-2389.1961.tb00901.x

Oliver MA, Webster R (1989) A geostatistical basis for spatial weighting in multivariate classification. Math Geol 21:15–35. doi:10.1007/BF00897238

Powell B, McBratney AB, MacLeod DA (1992) Fuzzy classification of soil profiles and horizons from the Lockyer Valley, Queensland, Australia. Geoderma 52:173–197. doi:10.1016/0016-7061(92)90082-I

Rao CR (1948) The utilization of multiple measurements in problems of biological classification. J R Stat Soc Ser B (Methodol) 10:159–193

Rayner JH (1966) Classification of soils by numerical methods. J Soil Sci 17:79–92. doi:10.1111/j.1365-2389.1966.tb01454.x

Ribeiro MV, Cunha LMS, Camargo HA, Rodrigues LHA (2014) Applying a fuzzy decision tree approach to soil classification. In: Laurent A, Strauss O, Bouchon-Meunier B, Yager RR (eds) Information processing and management of uncertainty in knowledge-based systems, communications in computer and information science. Springer, Cham, pp 87–96

Robinove, C.J., 1979. Integrated terrain mapping with digital landsat images in Queensland, Australia (Geological Survey Professional Paper No. 1102). United States Geological Survey, Washington, D.C.

Roubens M (1982) Fuzzy clustering algorithms and their cluster validity. Eur J Oper Res 10:294–301. doi:10.1016/0377-2217(82)90228-4

Roudier, P., Manderson, A., Hedley, C., 2016. Advances towards quantitative assessments of soil profile properties, In: Hartemink, A.E., Minasny, B. (Eds.), Digital soil Morphometrics, progress in soil science. Springer International, pp. 113–132. doi:10.1007/978-3-319-28295-4_8

Rousseeuw PJ, Kaufman L, Trauwaert E (1996) Fuzzy clustering using scatter matrices. Comput Stat Data Anal 23:135–151. doi:10.1016/S0167-9473(96)00026-6

Saboya Jr F, da Glória Alves M, Dias Pinto W (2006) Assessment of failure susceptibility of soil slopes using fuzzy logic. Eng Geol 86:211–224. doi:10.1016/j.enggeo.2006.05.001

Sarkar PK, Bidwell OW, Marcus LF (1966) Selection of characteristics for numerical classification of soils. Soil Sci Soc Am Proc 30:269–272. doi:10.2136/sssaj1966.302269x

Schulz K, Huwe B (1999) Uncertainty and sensitivity analysis of water transport modelling in a layered soil profile using fuzzy set theory. J Hydroinf 1:127–138

Scull P, Franklin J, Chadwick OA, McArthur D (2003) Predictive soil mapping: a review. Prog Phys Geogr 27:171–197. doi:10.1191/0309133303pp366ra

Simonson, R.W., Gardiner, D.R., 1960. Concept and functions of the pedon. Presented at the 7th International Congress of Soil Science, pp. 127–131

Sneath PHA, Sokal RR (1962) Numerical taxonomy. Nature 193:855–860. doi:10.1038/193855a0
Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification, A series of books in biology. W. H. Freeman and Company, San Francisco
Soil Survey Staff (1993) Soil Survey Manual, U. S. Department of Agriculture Handbook 18. United States Department of Agriculture Soil Conservation Service
Soil Survey Staff (1999) Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys, 2nd ed. United States Department of Agriculture Natural Resources Conservation Service
Soil Survey Staff (2016) Soil survey geographic (SSURGO) database. United States Department of Agriculture Natural Resources Conservation Service. https://sdmdataaccess.sc.egov.usda.gov. Accessed 9 Dec 2016
Sokal RR (1961) Distance as a measure of taxonomic similarity. Syst Zool 10:70–79
Sokal RR (1963) The principles and practice of numerical taxonomy. Taxon 12:190–199
Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438
Sokal RR, Sneath PHA (1963) Principles of numerical taxonomy, A series of books in biology. W. H. Freeman and Company, San Francisco
Stace HCT, Hubble GD, Brewer R, Northcote KH, Sleeman JR, Mulcahy MJ, Hallsworth EG (1968) A handbook of Australian soils. Rellim Technical Publications, Glenside
Steinley D (2006) K-means clustering: a half-century synthesis. Br J Math Stat Psychol 59:1–34. doi:10.1348/000711005X48266
Sun X-L, Zhao Y-G, Wang H-L, Yang L, Qin C-Z, Zhu A-X, Zhang G-L, Pei T, Li B-L (2012) Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters. Geoderma 171–172:24–34. doi:10.1016/j.geoderma.2011.03.016
Triantafilis J, Ward WT, Odeh IOA, McBratney AB (2001) Creation and interpolation of continuous soil layer classes in the lower Namoi valley. Soil Sci Soc Am J 65:403–413
Triantafilis J, Odeh IOA, Minasny B, McBratney AB (2003) Elucidation of physiographic and hydrogeological features of the lower Namoi valley using fuzzy k-means classification of EM34 data. Environ Model Softw 18:667–680. doi:10.1016/S1364-8152(03)00053-7
Valeriano MM, Epiphanio JCN, Formaggio AR, Oliveira JB (1995) Bi-directional reflectance factor of 14 soil classes from Brazil. Int J Remote Sens 16:113–128. doi:10.1080/01431169508954375
van Alphen BJ, Stoorvogel JJ (2000) A functional approach to soil characterization in support of precision agriculture. Soil Sci Soc Am J 64:1706–1713. doi:10.2136/sssaj2000.6451706x
Verheyen K, Adriaens D, Hermy M, Deckers S (2001) High-resolution continuous soil classification using morphological soil profile descriptions. Geoderma 101:31–48. doi:10.1016/S0016-7061(00)00088-4
Verma P, Singh P, George KV, Singh HV, Devotta S, Singh RN (2009) Uncertainty analysis of transport of water and pesticide in an unsaturated layered soil profile using fuzzy set theory. Appl Math Model 33:770–782. doi:10.1016/j.apm.2007.12.004
Vrindts E, Mouazen AM, Reyniers M, Maertens K, Maleki MR, Ramon H, de Baerdemaeker J (2005) Management zones based on correlation between soil compaction, yield and crop data. Biosyst Eng 92:419–428. doi:10.1016/j.biosystemseng.2005.08.010
Webster R (1968) Fundamental objections to the 7th approximation. J Soil Sci 19:354–366. doi:10.1111/j.1365-2389.1968.tb01546.x
Webster R (1977) Quantitative and numerical methods in soil classification and survey. Oxford University Press, Oxford
Webster R, Burrough PA (1974) Multiple discriminant analysis in soil survey. J Soil Sci 25:120–134. doi:10.1111/j.1365-2389.1974.tb01109.x
Webster R, Oliver MA (1990) Statistical methods in soil and land resource survey, spatial information systems. Oxford University Press, Oxford
Williams WT (1976a) Attributes. In: Williams WT (ed) Pattern analysis in agricultural science. CSIRO, Melbourne, pp 31–36

Williams WT (1976b) Types of classification. In: Williams WT (ed) Pattern analysis in agricultural science. CSIRO, East Melbourne, pp 76–83

Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13:841–847. doi:10.1109/34.85677

Yang M-S, Ko C-H (1996) On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets Syst 84:49–60. doi:10.1016/0165-0114(95)00308-8

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353. doi:10.1016/S0019-9958(65)90241-X

Zhang S, Zhu A-X, Liu W, Liu J, Yang L (2013) Mapping detailed soil property using small scale soil type maps and sparse typical samples. Chin Geogr Sci 23:680–691. doi:10.1007/s11769-013-0632-7

Chapter 9
Soil Profile Classes

Nathan P. Odgers, Alex. B. McBratney, and Florence Carré

> *". . . and He shall separate them one from another, as a shepherd divideth his sheep from the goats: and He shall set the sheep on His right hand, but the goats on the left".*
>
> Matthew 25:32–33 (KJV)

The previous chapter discussed the possibility of using pedometric techniques to make numerical classifications of soil material and soil layers. Of course it is not a step too far to use pedometric techniques to make classifications of entire profiles also. That is the subject of this chapter.

One of the main differences between classification of soil material and classification of soil profiles lies in how we abstract and encapsulate soil profiles as structured data. The simplest view treats whole profiles as individuals complete in themselves, but more sophisticated approaches consider profiles as composite objects made up of sequences of soil layer individuals. We will see that the choice of approach determines how the distances between soil profiles are computed.

Once distances have been computed, there are several ways to identify classes of soil profiles. Methods of classification may be viewed as either hierarchical or non-hierarchical.

N.P. Odgers (✉)
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia

Soils and Landscapes Team, Manaaki Whenua – Landcare Research, PO Box 69040,
Lincoln, 7640, New Zealand
e-mail: OdgersN@landcareresearch.co.nz

A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: alex.mcbratney@sydney.edu.au

F. Carré
INERIS, Parc Technologique Alata, BP2, 60550 Verneuil en Halatte, France
e-mail: Florence.CARRE@ineris.fr

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_9

9.1 Profiles as a Sequence of Horizons

Like the soil layer, another basic soil entity that is recognised universally is the soil profile (Simonson 1968). Soil profiles can be described in a number of ways, for example, as a sequence of horizon classes or as a sequence of depth functions for various soil properties. One of the problems of describing profiles as sequences of horizons is that the sequences themselves are quite short, because profiles have at least one to seldom more than seven horizons.

9.1.1 Homology

When one wants to compare one profile with another one, it is worth considering which of the horizons are homologous—in other words, which horizons perform the same function. This is really quite a difficult problem, and no quantitative solution has yet been found. Homology could be established based on the horizons which are most similar to each other in different profiles (e.g. Rayner 1966), on the basis of a common sequence of horizons (e.g. Dale et al. 1970; Little and Ross 1985) or with respect to the presence or absence of diagnostic horizon classes. To gain a better understanding of the possible importance of homology, consider a podzol with the usual horizon sequence developed somewhere in Scotland which is 50 cm thick and a giant podzol developed in the sand dunes of eastern Australia which is 20 m thick. Both have the same sequence of horizons, so on this basis, these profiles are completely homologous with each other. On the other hand, if we compare the soil profiles on a depth-by-depth basis, they would be considered completely different from each other. Homology on the basis of diagnostic horizons presents its own problems, discussed briefly in the previous chapter and at length by FitzPatrick (1976). So for quantitative understanding of soil profiles and their relation with each other, a quantification of homology is required.

9.1.2 Transition Matrices

One way of describing a soil profile as a sequence of horizons is to construct for each profile a transition matrix which describes the number of transitions from any particular horizon class to any other horizon class (e.g. Dale et al. 1970; Moore et al. 1972; Norris and Dale 1971). By convention, the class that defines the column succeeds that which defines the row in the sequence. Any two profiles that have the same sequence would have the same transition matrix. The problem with this approach is that the matrices are rather sparse—in other words they contain a lot of zero entries. This means that the information content of the matrices will be low. By way of illustration, a Red Dermosol (Isbell 1996) under viticulture in the lower

Table 9.1 Transition matrix for a Red Dermosol under viticulture at Pokolbin in the lower Hunter Valley in New South Wales, Australia

	A	B21	B22	B3	C
A	0	1	0	0	0
B21	0	0	1	0	0
B22	0	0	0	1	0
B3	0	0	0	0	1
C	0	0	0	0	0

Hunter Valley in New South Wales, Australia, might have a horizon sequence of A B21 B22 B3 C. Its transition matrix is depicted in Table 9.1.

The information content of the matrix, I, is computed as (Dale et al. 1970):

$$I = \mathrm{G} \ln \mathrm{G} - \sum_{i=1}^{n} \sum_{j=1}^{m} f_{\mathrm{ij}} \ln f_{\mathrm{ij}} \tag{9.1}$$

where f_{ij} is the entry in row i and column j of the transition matrix and G is the sum of all f_{ij}.

The information content of the transition matrix in Table 9.1 is 5.55.

9.2 Profiles as Depth Functions

From a quantitative point of view, it is easier and seems more natural to describe soil profiles by the depth functions of the individual soil properties although this approach must exclude some pedological insight that horizons carry. Depth functions may be used for several purposes. For example, they enable the estimation of the value of a soil attribute at any depth in the soil profile and even the total amount of an attribute, such as carbon, over a whole region of a profile (e.g. Arrouays and Pelissier 1994; Mestdagh et al. 2004).

Depthwise distributions of soil attributes are sometimes represented as step functions when attribute measurements or predictions are made for each soil layer (e.g. Jobbagy and Jackson 2000). In practice the measurement made for a layer is often made on soil material derived from several locations in the layer and thus represents its average condition. As a result, the variation in the soil attribute that the step function describes has been dampened by virtue of the sampling and measurement process (Ponce-Hernandez et al. 1986). In other words the step function likely does not reflect the true extremes of the soil attribute distribution within the profile.

In reality, the depthwise distributions of soil attributes are usually continuous (Colwell 1970). These distributions have been modelled by a wide range of mathematical functions; some are examined in the remainder of this section.

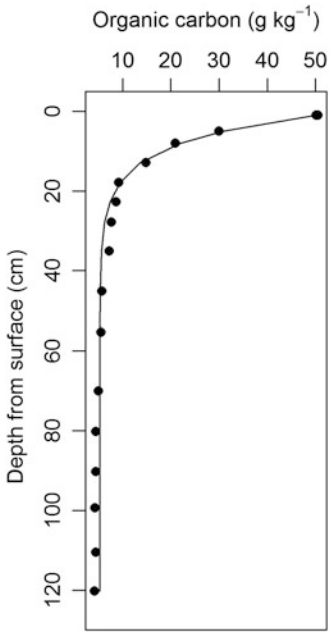


Fig. 9.1 Exponential depth function of organic carbon for an alluvial soil in New South Wales, Australia, after Minasny et al. (2016) (Data from Walker and Green 1976)

9.2.1 Exponential Functions

The most popular choice of depth function has been the exponential function (Minasny et al. 2016). They are most useful for modelling the trend in properties that decay from the surface (Fig. 9.1). Researchers have fit exponential functions to a range of attributes including clay content (Brewer 1968), root density (Dwyer et al. 1996; Kalisz et al. 1987), radionuclide distribution (He and Walling 1997; Koarashi et al. 2012) and, notably, organic carbon (Arrouays and Pelissier 1994; Bernoux et al. 1998; Malone et al. 2009; Mestdagh et al. 2004; Mikhailova and Post 2006; Mishra et al. 2009; Russell and Moore 1968; Taylor and Minasny 2006).

The general form of the exponential function is:

$$y = a\mathrm{e}^{bd} \tag{9.2}$$

where y is the value of the soil property, d is the depth below the soil surface, a is a constant relating to the initial value at the surface of the profile and b controls the rate of decay.

9.2.2 Polynomials

Polynomial functions model the trend with depth of attributes that have a maximum somewhere other than at the surface, such as the clay content of a soil with an argillic B horizon. Quadratic polynomials require relatively few data to fit but are not the best solution where a greater degree of flexibility is required. Cubic or higher-order polynomials are more flexible in shape but require more data to fit adequately.

Although their application has not been universally successful (Campbell et al. 1970), Colwell (1970) found that polynomial functions adequately characterised the depthwise trend in attributes relating to soil fertility across four great soil groups in New South Wales, Australia. Other researchers have used them to model the depthwise trend in attributes such as soil particle size fractions (Wickramagamage 1986), penetration strength (Veronesi et al. 2012), soil moisture (Tabatabaeenejad et al. 2015) and radionuclide activity (Alewell et al. 2014; Ramzaev et al. 2007).

9.2.3 Wetting Fronts

During an infiltration event, a plot of soil water content against depth in the profile displays a characteristic shape due to the wetting front. The wetting front is the zone of interface between the unsaturated soil below, with low soil water content, and the zone of transmission, above, with high soil water content. Researchers have tried to model these numerically (Haverkamp et al. 1977; Hills et al. 1989). Functions of other soil phenomena display a similar shape (Fig. 9.2). For example, weathering processes may be examined by analyses of solute concentration with depth, whose functions are the so-called weathering fronts or reaction fronts (Brantley et al. 2008; Kirkby 1985).

These fronts can in general be modelled using a sigmoid function (Minasny et al. 2016):

$$C(d) = \frac{C_{\max}}{1 + \frac{C_{\max} - C_{\min}}{C_{\min}} \exp(\alpha d)} \tag{9.3}$$

where $C(d)$ is the concentration at depth d and α is an empirical parameter. $C_{\max}$ is the maximum concentration, which is at the top of the profile for a depletion trend and at the bottom of the profile for an addition trend. Likewise $C_{\min}$ is the minimum concentration, which is at the bottom of the profile for a depletion trend and at the top of the profile for an addition trend.

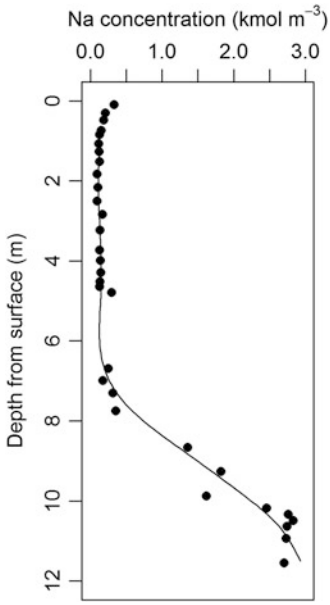


Fig. 9.2 Concentration of Na in a granite soil in Georgia in the United States displays a wetting front-type depth function (After Minasny et al. 2016 Data from Brantley et al. 2008)

9.2.4 Peak Functions

The depthwise distributions of certain soil attributes, such as clay content, often display a peak at some point in the soil profile. Oftentimes these distributions display a normal (Wetselaar 1962) or lognormal (Jury 1982) shape, but in other situations, the distributions are not symmetrical throughout their depth or may contain more complex features (Fig. 9.3). In other circumstances, such as when surface mixing and translocation processes are operating, or when there is excessive bioturbation, the depth function of a soil attribute may display both a minimum and a maximum in the profile (Minasny et al. 2016). In these situations typical polynomial parametric functions may be unable to adequately model these shapes because attribute measurements down the profile are usually sparse and few in number, among other reasons. Peak functions, being non-linear parametric functions that are also non-linear in their parameters, are also susceptible to these faults, which have led some researchers to recommend that attributes be measured at high resolution using proximal sensors (Myers et al. 2011). Non-parametric methods such as splines are better able to model complex shapes but are unable to quantitatively describe features present in them (Myers et al. 2011).

Some peak functions are asymmetric, meaning that the shape of the curve on either side of the peak can vary. Myers et al. (2011) demonstrated the use of the Pearson IV and logistic power peak functions to model depth functions of soil attributes. The Pearson IV function has six parameters, three of which control peak shape, whereas the logistic power peak function has five parameters, two of

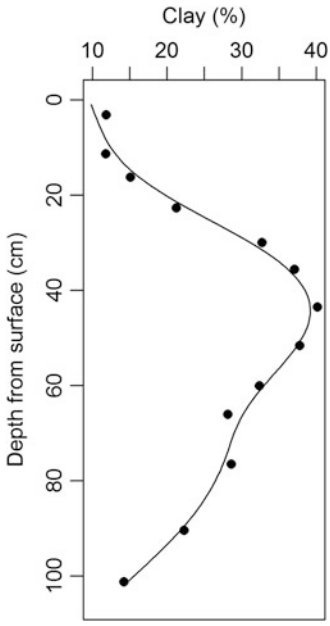


Fig. 9.3 Clay content often displays a peak-type depth function (After Minasny et al. 2016)

which control peak shape. Peak functions are usually fit using non-linear regression techniques that attempt to optimise these parameters.

Myers et al. (2011) demonstrated that both peak functions fit very well to clay content, silt content and pH profiles although they preferred the logistic power peak function because of its parsimony with respect to the fewer parameters. The logistic power peak function is expressed as follows (Myers et al. 2011):

$$\mathbf{y} = \alpha + \frac{\beta}{\mu}\left(1+\exp\left(\frac{\mathbf{d}+\delta\ln(\varepsilon)-\mu}{\delta}\right)\right)^{\frac{-\varepsilon-1}{\varepsilon}} \exp\left(\frac{\mathbf{d}+\delta\ln(\varepsilon)-\mu}{\delta}\right)(\varepsilon+1)^{\frac{\varepsilon+1}{\varepsilon}} \tag{9.4}$$

where $\mathbf{d}$ is a vector of depths, $\mathbf{y}$ is a vector of soil attribute values, α is the intercept, β is the amplitude, $\alpha+\beta$ is the magnitude of the peak, μ is the peak centre and δ controls the width of the peak and interacts with ε to control the asymmetry of the peak.

Due to the interactions between parameters, different sets of parameters can produce curves that are similar in shape (equifinality). Despite this, Myers et al. (2011) were able to link parameters with soil-landscape processes and parameters. For example, they were able to demonstrate a relationship between the parameter in the logistic power peak function that controlled peak width and position along a hillslope.

9.2.5 *Splines*

Some (Webster 1978) have pointed out several difficulties associated with the fitting of polynomial or exponential functions to soil attribute data. Objections include the lack of theory from which to determine a suitable model—although it is noted that some properties, such as organic carbon, often follow an exponentially decreasing trend through the profile (Minasny et al. 2006; Russell and Moore 1968)—the difficulty in fitting a polynomial function to more than a few data points and the fact that because the goodness of fit is sensitive to individual measurements, local variation or error in a single measurement can affect the fit of the function to the whole profile.

Splines were introduced as a way of mitigating some of these deficiencies. Instead of fitting a single function to a whole profile, splines typically fit a series of polynomial functions piecewise through measurements made down the profile. Locations where consecutive polynomials join are called *knots*, and constraints on curvature are usually applied at the knots to ensure the curve remains smooth through the transition. Splines may be *exact fitting* or *smoothing*. In exact-fitting splines, the curve passes through the data points exactly, and the knots are located at the data points. Erh (1972) fits exact-fitting cubic splines to the water retention curve. Others (e.g. Kastanek and Nielsen 2001; Wesseling et al. 2008) performed similar analyses, but in general the exact-fitting splines have not been widely applied in soil science.

We noted at the start of this section that soil attributes are often measured as average values within soil horizons. For the purpose of fitting the depth functions described previously in this section, it is normally assumed that these averages correspond to the depth at the centres of the horizons. Depth functions fit to horizon averages in this way will almost certainly display less variance in the soil attribute than the true depth function if it could be known accurately, even if the splines are exact fitting.

To mitigate this effect, Ponce-Hernandez et al. (1986) proposed an equal-area smoothing spline method. In smoothing splines, unlike exact-fitting splines, the curve does not pass exactly through the data points, and the knots are located at the depth boundaries of the calibration data supports. Equal-area smoothing splines are still fit through horizon averages, but a constraint is added so that the area to the left of the spline (the X_i in Fig. 9.4) equals the area to the right of the spline (the Y_i in Fig. 9.4) through each horizon. In this way the average value of the property through the horizon is preserved, but the equal-area spline should reflect the actual variation in the soil property more accurately.

Bishop et al. (1999) demonstrated that equal-area quadratic smoothing splines were better able to predict the true depth functions of a range of soil properties including pH, clay content and organic carbon content than quadratic polynomials, exponential functions and linear regressions and simple horizon averages. Their model required that the depthwise support of the calibration data is contiguous with depth. This requirement is satisfied when the calibration data are a set of horizon averages; however, samples are often taken at depth intervals that may

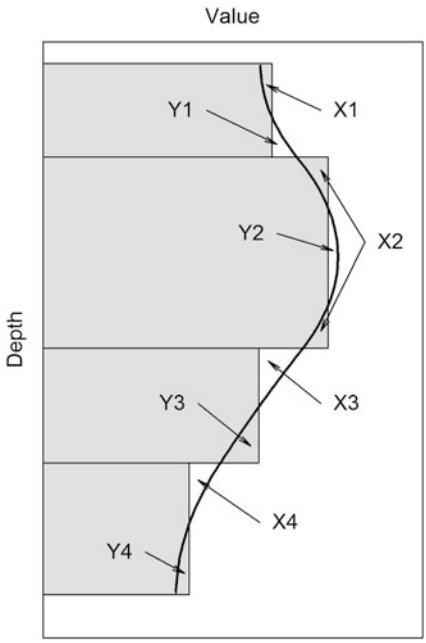


Fig. 9.4 An equal-area quadratic spline, after Ponce-Hernandez et al. (1986) and Bishop et al. (1999)

not be contiguous with depth. For example, Crépin and Johnson (1993) recommended several methods of sampling discontiguous depth intervals in mechanically disturbed profiles where horizonation may not be evident. Therefore Malone et al. (2009) extended the work of Bishop et al. (1999) in order to relax the contiguity requirement. The updated spline model remained quadratic across calibration supports but became linear between them. We refer readers to the relevant publications for complete mathematical implementations of these techniques.

9.3 Classification of Soil Profiles

If we have a soil profile described by series of properties at fixed depth or as a collection of depth functions or as a sequence of horizons, one of the fundamental problems in soil science is to create a classification of such profiles. The possible procedures are described in Fig. 9.5, and all require the computation of an *association matrix* that describes the degree of association between all pairs of soil profiles being compared. This matrix is usually square and symmetric, and each element describes the degree of association between a particular pair of profiles. We saw in Chap. 8 that the most commonly used inter-profile association metrics are those that are distances in a soil attribute space and strictly compute the degree of *dissimilarity* between two individuals because their values increase with decreasing similarity. In such cases the association matrix may be called a *between-profile* or *inter-profile distance matrix*.

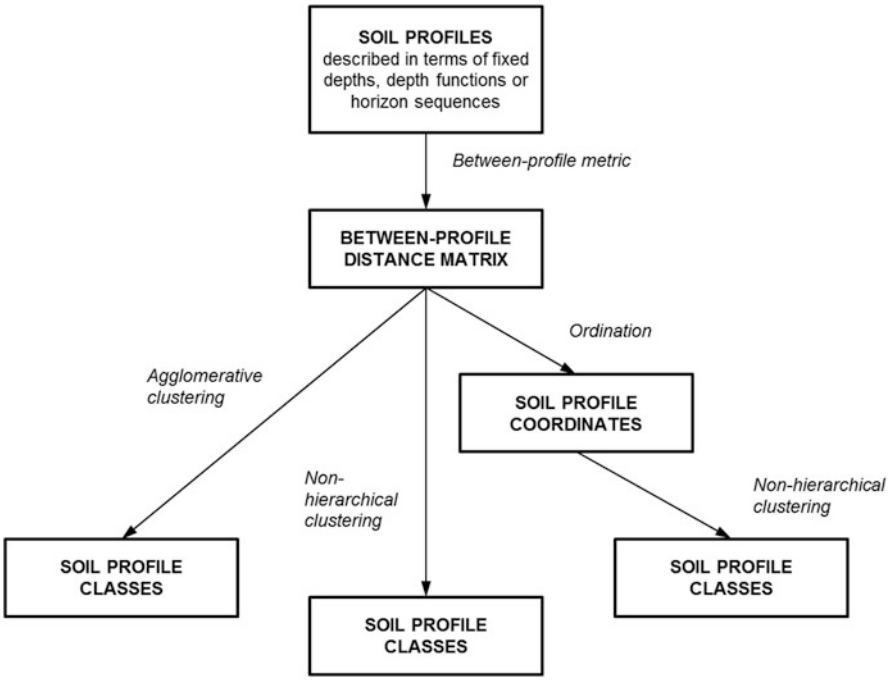


Fig. 9.5 Soil profiles may be classified in several ways

9.3.1 Between-Profile Metrics

One of the chief problems in profile classification is calculation of the inter-profile distance between two soil profiles. It is relatively easy to compute distances between a pair of individual objects such as soil horizons, but a means by which these metrics can be used to calculate the distance between two profiles is not immediately obvious. This is because soil profiles can be viewed not as individual objects but as *sequences of horizon objects* where each horizon has its own soil attribute information. Researchers have devised several methods of computing the inter-profile distance. We briefly discuss them in this section.

9.3.1.1 Reference Horizons

We may choose to recognise the existence of reference horizons or layer classes yet for practical reasons still treat a profile as a single object for the purpose of computing the inter-profile distance. Thus a profile may be represented as a vector of attributes drawn from various horizons throughout the profile. Calculation of an inter-profile distance by regular Euclidean, Mahalanobis or other means is therefore simple.

Due to its relative simplicity, it is perhaps not surprising that this approach was the first used in the numerical classification of soil profiles. For example, Hole and Hironaka (1960) computed the inter-profile similarity on data on 25 attributes drawn from throughout the profile, including thickness of the A2 horizon, ratio of the maximum B horizon clay content to minimum A horizon clay content, average 0–36-in. pH and SiO_2/R_2O_3 ratio of the C or D horizon. Bidwell and Hole (1964) performed similar computations on 30 attributes including A horizon pH, B horizon hue, C horizon clay content and the proportion of the solum exhibiting colour mottles. Young and Hammer (2000) used a range of horizon and profile attributes.

Bottner et al. (1975) performed a more complex analysis than the previous researchers. Like Hole and Hironaka (1960) and Bidwell and Hole (1964), they assembled data on a range of attributes drawn from various reference horizons although about half were attributes of the B horizon. In addition they tried to take account of the presence or absence of relevant horizons in the soil profile by creating rules to penalise the overall inter-profile distance; one or other of a pair of profiles being compared did not possess a specific reference profile. This ensured that profiles with different sequences of horizons had a larger inter-profile distance.

9.3.1.2 Fixed Depths

Notwithstanding the pedological information they contain, we may choose not to use pedological horizons as the basis for computation of the inter-profile distance. We may be concerned, for example, about the subjectivity in horizon designation, the lack of comparative horizons in certain soils or the fact that horizons of like class have different depth boundaries (Russell and Moore 1968). An alternative approach is to compare pairs of layers of fixed or varying thickness down the profile instead.

For instance, Russell and Moore (1968) treated soil profiles as sequences of 1-in.-thick layers whose attributes were derived from those of the soil horizons they intersected. They calculated the inter-profile distance between a pair of profiles by computing the average of the Euclidean distances D_E between layer pairs. Consider two profiles X and Y with $i = 1,2,\ldots,m$ layers. It is assumed that both profiles have the same number of layers, m, and that the layers have equal thickness. The lower depths of their layers x_i and y_i are denoted d_{xi} and d_{yi}, respectively, with the surface at d_{x0} and d_{y0}. Mathematically the inter-profile distance D(X,Y) is calculated as:

$$D(X,Y) = \frac{1}{m}\sum_{i=1}^{m} W_i D_E(x_i, y_i) \tag{9.5}$$

where W_i is an optional weighting factor. Russell and Moore (1968) experimented with assigning differential weights to the layers using a negative exponential function on the basis that the upper layers are more important than those lower in the profile for some purposes. Using the negative exponential function, layers higher in the profile would receive greater weight in the inter-profile distance calculation than those lower in the profile.

Other methods search for the layers in Y that are most similar to those in X and vice versa. There are several variants of this approach. Assume now that the number of layers in X can be different from the number of layers in Y, so that now their layers are denoted x_i and y_j, respectively, where $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. The approach that Grigal and Arneman (1969) devised consists of three steps:

1. Compare each layer x_i in X with three layers in Y: (i) the layer at a comparable depth, y_j, the layer immediately above this layer, y_{j-1}, and (iii) the layer immediately below this layer, y_{j+1}. The layer in Y that is most similar to the one in X in the Euclidean sense is considered to be its analogue.
2. Repeat (1) with respect to all layers in Y.
3. Compute the inter-profile distance as the arithmetic mean of the X-to-Y and Y-to-X analogous layer distances:

$$D(X,Y) = \frac{1}{m+n}\sum_{i=1}^{m}\min_{j\in j-1<j<j+1} D_E\left(x_i, y_j\right) + \frac{1}{m+n}\sum_{j=1}^{n}\min_{i\in i-1<i<i+1} D_E\left(y_j, x_i\right) \tag{9.6}$$

A rule must be devised to govern the mapping of the x_i to the corresponding y_j and vice versa.

Rayner (1966) devised a similar method except the comparison in step (1) that involved all the layers in Y at and below the corresponding layer in X.

Yet other methods contain elements of all the above approaches. Carré and Girard (2002) devised a distance similar to the *utilitarian distance* described later by Carré and Jacobson (2009) that uses the distance between pairs of adjacent layers weighted by the thickness of overlap (Fig. 9.6). Now the depth of the deepest profile is $d_{\max}$, and $D_{\max}$ is the largest interlayer distance. With respect to X, overlap between layers in X and Y occurs when any of the following conditions are satisfied:

$$d_{\mathrm{yj}-1} \le d_{\mathrm{xi}-1} < d_{\mathrm{yj}} \& d_{\mathrm{yj}-1} < d_{\mathrm{xi}} \le d_{\mathrm{yj}} \tag{9.7}$$

$$d_{\mathrm{yj}-1} \le d_{\mathrm{xi}-1} < d_{\mathrm{yj}} \& d_{\mathrm{yj}-1} < d_{\mathrm{xi}} \le d_{\mathrm{yj}} \tag{9.8}$$

$$d_{\mathrm{xi}-1} \le d_{\mathrm{yj}-1} < d_{\mathrm{yj}} \& d_{\mathrm{yj}-1} < d_{\mathrm{xi}} \le d_{\mathrm{yj}} \tag{9.9}$$

The distance $D(X, Y)$ between the profiles in Fig. 9.6 is thus:

$$D(X,Y) = \{D(x_1,y_1)\times(d_{x1}-d_{x0}) + D(x_2,y_1) \times \left(d_{y1}-d_{x1}\right) + D(x_2,y_2)\times\left(d_{x2}-d_{y1}\right) + D_{\max}\times\left(d_{y2}-d_{x2}\right)\}/d_{\max} \tag{9.10}$$

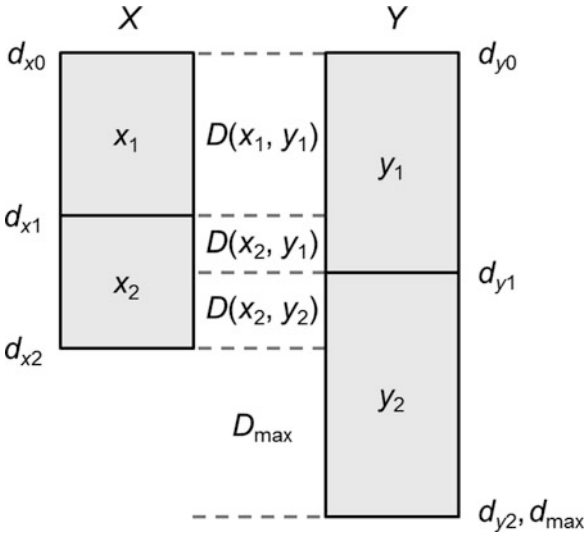


Fig. 9.6 Elements of the inter-profile distance calculation of Carré and Girard (2002)

More generally we can think of this inter-profile distance calculation in terms of the number of overlapping layer segments, $g = 1,2,\ldots,l$. Thus if the upper and lower depths of a zone of overlap are p and q, respectively, the inter-profile distance is:

$$D(X,Y) = \frac{1}{d_{\max}} \sum_{g=1}^{l} \begin{cases} D(x_g, y_g) \times (q_g - p_g), & g < l \\ D_{\max} \times (q_g - p_g), & g = l \end{cases} \tag{9.11}$$

where $D(x_g, y_g)$, the interlayer distance between x_g and y_g, is the Euclidean distance divided by the number of layer attributes (Carré and Girard 2002). Later, Carré and Jacobson (2009) further developed this work to implement two additional inter-profile distances they called *pedological* and *joint*. The pedological distance considers the interlayer distances (Euclidean or Manhattan) of pairs of layers depthwise, in sequence, though layer thickness is not taken into account. The joint distance is computed in the same manner as the utilitarian distance, but the layer boundary depths of both profiles are first scaled so that each profile has a depth of 1.0.

9.3.1.3 Collection of Depth Functions

Another approach is to calculate the inter-profile distance using the parameters of the depth functions as attributes. Since the depth functions already account for continuous variation within and between soil layers, conventional metrics like the Euclidean or Mahalanobis may be used to quantify the inter-profile distance. Campbell et al. (1970) did this for coastal soils in Western Australia and Moore

et al. (1972) for soils from Queensland, Australia. Other examples appear to be scarce. The success of a classification made using this method appears to be closely tied to how well the depth functions fit the observed data. For example, Campbell et al. (1970) found that the classification they made by clustering depth function parameters was not like those they made by clustering the original profile data, which they attributed to ill-fitting depth functions. Perhaps a better classification could have been made if a more appropriate depth function model was used.

9.3.1.4 Sequence of Horizons

Transition Matrices

We saw in Sect. 9.1.2 how the sequence of layer classes in a soil profile can be represented as a transition matrix. The inter-profile distance between a pair of profiles can be computed as the element-wise sum of the Euclidean distances between their transition matrices. Dale et al. (1970) and Norris and Dale (1971) demonstrated how the similarity between two profiles could also be computed in terms of the information content of their transition matrices.

The transition matrix approach recognises that individual soil layers are not independent of each other; furthermore it explicitly takes into account the sequence of layers down a profile. This is in contrast to most other inter-profile distance metric schemes. On the other hand, we must consider a couple of factors before applying this approach. First, since soil attribute data are not formally utilised in the inter-profile similarity calculations, relationships between attributes are implicit in the layer class labels used to encode the matrices. The choice of layer classification system therefore plays an important role because the same layer class label can have different connotations in different kinds of soil. For example, the pedogenic B horizon implies a different set of processes in vertisols than it does in spodosols. It is therefore likely to be better to use a soil layer classification in which the class labels have a consistent meaning across all kinds of soil. A numerical soil layer classification system (e.g. Odgers et al. 2011b; Triantafilis et al. 2001) may satisfy this requirement if the universe of soils it encompasses is large enough. Norris and Dale (1971) first classified soil layers into eight numerical classes prior to creating transition matrices for their soil profiles.

The second factor that should be considered since the transition matrices of soil profiles are often sparse is whether the transition matrices contain enough information to separate soil profiles adequately. Two characteristics of soil profiles that work to confound the information content of their transition matrices are (i) the fact that they often possess only short sequences, since the number of horizons is typically less than 7, and (ii) the fact that transitions between layer classes (e.g. A→B, A1→A2) tend to occur only once in a given profile. This leads to the dominance of 1s and 0s in soil profile transition matrices. Therefore since $1 \times \ln 1 = 0$ and $0 \times \ln 0$ is undefined, the information content according to Dale et al.

Table 9.2 Horizon sequence of a brown Dermosol from the lower Hunter Valley, New South Wales, Australia

Horizon	Depth range (cm)
An	0–8
B21	8–40
B22	40–84
B3	84–103
BC	103–112

Table 9.3 Depth interval sequence of a brown Dermosol from the lower Hunter Valley, New South Wales, Australia

Horizon	Depth interval (cm)
An	0–10
B21	10–20
B21	20–30
B21	30–40
B22	40–50
B22	50–60
B22	60–70
B22	70–80
B3	80–90
B3	90–100
BC	100–110
BC	110–120

(1970) is largely driven by G (Eq. 9.1), which will be relatively small since the number of transitions between layers is typically relatively small.

We may increase the information content of a transition matrix by increasing the resolution at which the soil profile is segmented—in other words, by using, say, 10-cm-thick depth intervals rather than pedogenic horizons—as this will increase the chance of larger transition frequencies if, for example, a thick B horizon is segmented into three or four depth intervals. This has the additional benefit of implying some information about the relative thickness of each layer. For example, consider a Brown Dermosol from the lower Hunter Valley in New South Wales, Australia, with a pedogenic horizon sequence of An→B21→B22→B3→BC. Its horizon boundaries are given in Table 9.2 and the horizon designations of its 10-cm-depth intervals in Table 9.3. Using Eq. 9.1, the information content of the transition matrix based on Table 9.2 is 5.55 and that based on Table 9.3 is 21.69.

Levenshtein Metric

Profiles can be represented as sequences of horizon or layer designations. Given a set of layer classes $c = 1,2,\ldots,k$, a string representing the sequence of layers according to their layer classes could be the following:

22244335557

When we remove repeated layer codes, the compressed form (Little and Ross 1985) of the sequence is then:

24357

If all profiles can be represented in this way, we may ponder how the inter-profile distance can be computed on such a representation. One representation of this distance is the number of additions, deletions or substitutions of layer classes it would take to transform the compressed layer class sequence of *X* into that of *Y*. This is the so-called Levenshtein metric, devised by Levenshtein (1966) to describe the distance between two syntactic patterns and extended by Lu and Fu (1978), among others. Little and Ross (1985) applied the Levenshtein metric to soil profile classification.

9.3.2 Agglomerative Clustering of Distance Matrix

Once the inter-profile distance matrix has been computed, it may be subjected to an agglomerative cluster analysis using the techniques described in Chap. 8. Agglomerative techniques were initially used to examine numerical relationships between profiles allocated to the classes of established classification systems (e.g. Anderson 1971; Rayner 1966). The "ramifying linkage method" was probably the first agglomerative technique employed in soil science (Bidwell and Hole 1964). It is equivalent to the average linkage method (Arkley 1976; Webster 1977) and has its foundations in psychological research (Cattell 1944; Horn 1944) but came to the attention of soil researchers through its application in the biological domain (Sokal and Michener 1958).

There is no reason why the profiles being classified must be labelled a priori since construction of the taxonomic hierarchy involves the computation of the association between individuals in some form regardless of what class they belong to. Thus agglomerative clustering could be applied to an unlabelled collection of profiles to determine the relationships between them, but an appropriate *working level* (Wickramagamage and Fisher 1988) of the resulting hierarchy must be determined at which classes can be identified (Fig. 9.7). This is usually not the lowest level of the hierarchy; otherwise all profiles would be allocated to their own class.

9.3.3 Non-hierarchical Clustering

9.3.3.1 Preceded by Ordination of the Distance Matrix

Ordination is used to examine the structure of multivariate populations. In general terms ordination techniques can be thought of as ordering a population of individuals in attribute space so that similar individuals are closely spaced and vice

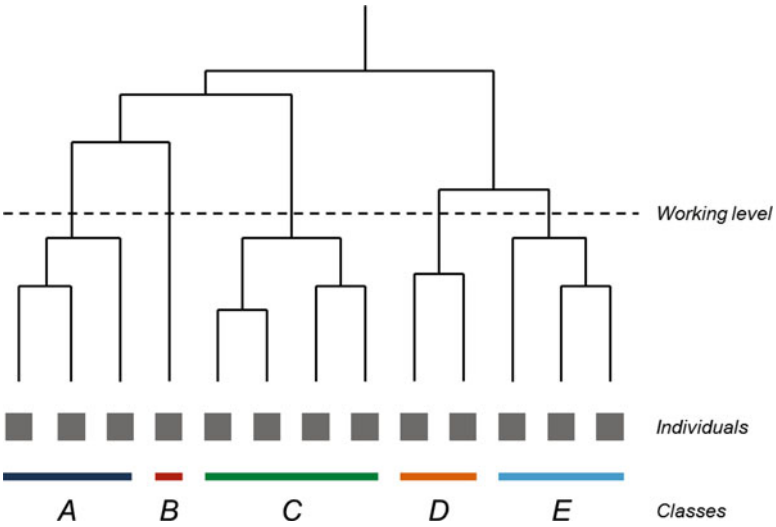


Fig. 9.7 A taxonomic hierarchy must be cut at a working level before classes based on it can be formed and centroids computed

versa. Two of the more well-known techniques are principal components analysis (PCA; Chap. 8) and principal coordinates analysis (PCoA) (Gower 1966). PCA is an R-type multivariate technique since it examines correlations between variables (i.e. attributes) across all individuals, whereas PCoA can be considered its Q-type complement because it examines correlations between individuals across all variables (Blackith and Reyment 1971).

In PCoA the starting point is the so-called N × N association matrix, **A**, whose elements a_{ij} are some estimates of resemblance between pairs of N individuals. The attribute data on which **A** is calculated are not required. Gower (1966) preferred to populate **A** with coefficients of association such as the Jaccard coefficient (Jaccard 1908) or simple matching coefficient, but any kind of similarity measure can be used as long as the matrix **B** (see below) has no large negative eigenvalues (Sneath and Sokal 1973). In PCoA the objective is to find the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_N$, and eigenvectors, $\mathbf{b}_1$, $\mathbf{b}_{2\ldots}\mathbf{b}_N$, of the matrix **A**. The eigenvectors form the matrix **B** and are normalised so that the sums of squares of their elements are equal to the corresponding latent roots:

$$\sum_{r=1}^{N} b_{\mathrm{ir}}^2 = \lambda_r \tag{9.12}$$

The rows of the normalised matrix give the coordinates of the individuals on their principal axes in N-dimensional Euclidean space. In fact it is possible to reduce the dimensionality of this space by examination of the eigenvalues of **B**. If a λ_r is small, then the b_{ir} are relatively small, and the contribution of $(b_{\mathrm{ir}} - b_{\mathrm{jr}})^2$to the distance between individuals i and j will also be small. Likewise if a λ_r is large but there is

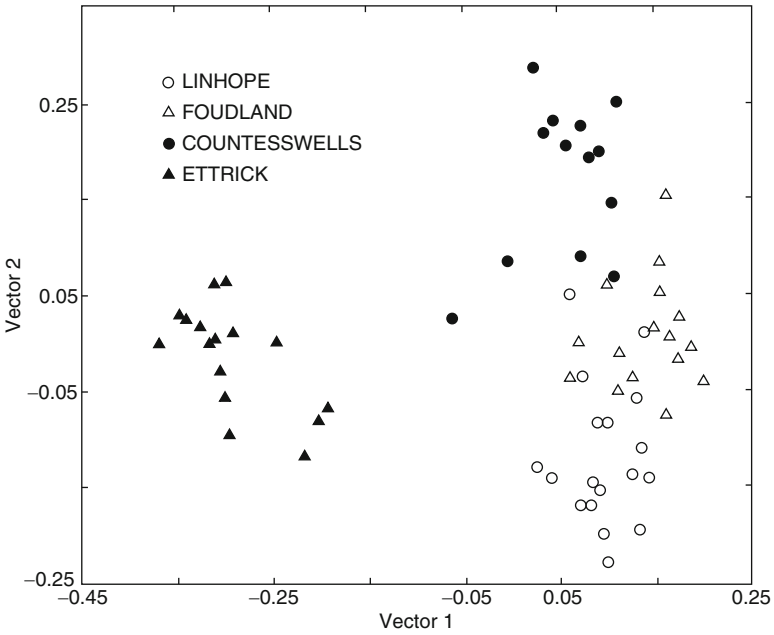


Fig. 9.8 Principal coordinates biplot of several soil series from Scotland (Muir et al. 1970, reproduced with permission)

not much variation in the size of the corresponding b_{ir}, then $(b_{ir} - b_{jr})$ will also be small. On the other hand, if λ_r is large and the corresponding b_{ir} vary in size, then $(b_{ir} - b_{jr})^2$ will contribute more to the distance between individuals i and j. Thus it is possible that the distances between individuals may be adequately expressed by only two or three principal coordinates (Blackith and Reyment 1971; Gower 1966).

PCoA was initially employed in the numerical classification of soil profiles in order to visualise the relationships between soils via biplots of the first two principal coordinates, as Fig. 9.8 demonstrates (e.g. Banfield and Bascomb 1976; Campbell et al. 1970; Rayner 1966; Webster and Butler 1976). It was seen as a companion to agglomerative clustering which performed the actual numerical classification.

Once non-hierarchical clustering became available, researchers realised that the principal coordinates themselves can be used as attributes in a non-hierarchical classification procedure such as k-means or fuzzy k-means. We will attempt to show that this is serendipitous for soil taxonomists. Recall that the k-means algorithms recompute centroids dynamically at the end of every iteration (Chap. 8). It is easy to perform this calculation if the data structure describing a soil profile is consistent across all soil profiles being classified. This is the case if soil profiles are represented as a set of parameters of a depth function or if profiles are represented as a set of depth slices of consistent thickness (Beaudette et al. 2013) because the centroids are merely the average of each of the parameters or each of the depth slices of the profiles within each cluster.

On the other hand, the method of computing a centroid is not obvious if the data structure is inconsistent across all profiles being classified, such as when profiles are represented as a sequence of horizons, since profiles vary in the number and thickness of horizons. Should the centroid have the average number of horizons of all individuals in the class? How should those horizons be attributed? Since relevant inter-profile distances are available (Carré and Jacobson 2009), it is easier to compute the principal coordinates of these profiles and use these data in a non-hierarchical classification procedure.

There are few examples of principal coordinates clustering in the pedometric literature. Examples are more frequent in ecology (e.g. Anderson and Clements 2000; Lefkovitch 1976; Mora and Iverson 2002; Schmidtlein et al. 2010; Tichý et al. 2011). In soil science, Oliver and Webster (1989) used the *k*-means algorithm to cluster the principal coordinates of a distance matrix that had been modified to weight the distances according to separation distance in order to reduce spatial fragmentation of the resulting classes. Frogbrook and Oliver (2007) performed a similar analysis to identify management zones within agricultural fields. Finally, Leblanc et al. (2016) clustered the principal coordinates of soil series where potato fertiliser trials had been conducted in Québec.

9.3.3.2 Without Prior Ordination of the Distance Matrix

A third class of soil profile classification techniques performs non-hierarchical classification without prior conversion of a distance matrix to a principal coordinate matrix. The OSACA algorithm (Carré and Girard 2002; Carré and Jacobson 2009) exemplifies this class of technique. OSACA classifies soil profiles using Diday's (1971) implementation of the hard *k*-means algorithm. The user can choose from one of three inter-profile distance metrics, described previously in Sect. 9.3.1.2. The choice of distance metric is largely driven by the intended use of the resulting classification. The algorithm converges when the allocation does not change between iterations or when the ratio of the mean distance between individuals and their respective centroids to the mean inter-centroid distance falls below a threshold.

The method of computing cluster centres varies depending on the type of inter-profile distance. When the pedological distance is used, the attributes of the cluster centres are the arithmetic mean of the attributes of the corresponding horizons of the individuals in the cluster. When the utilitarian or joint distances are used, it is not possible to average the horizon attributes. Instead, the cluster centres become the individuals that have the minimum total distance to the other individuals in the cluster. Since in these cases the cluster centres are real profiles, they are really *exemplars* or *medoids* rather than centroids in the traditional sense (Kaufman and Rousseeuw 1987).

The earlier version of OSACA (Carré and Girard 2002) incorporated a non-hierarchical cluster analysis of soil horizons in the algorithm. It was implemented prior to the classification of the soil profiles so that classes of modal horizons were found first followed by classes of modal profiles. The aim was to identify modal

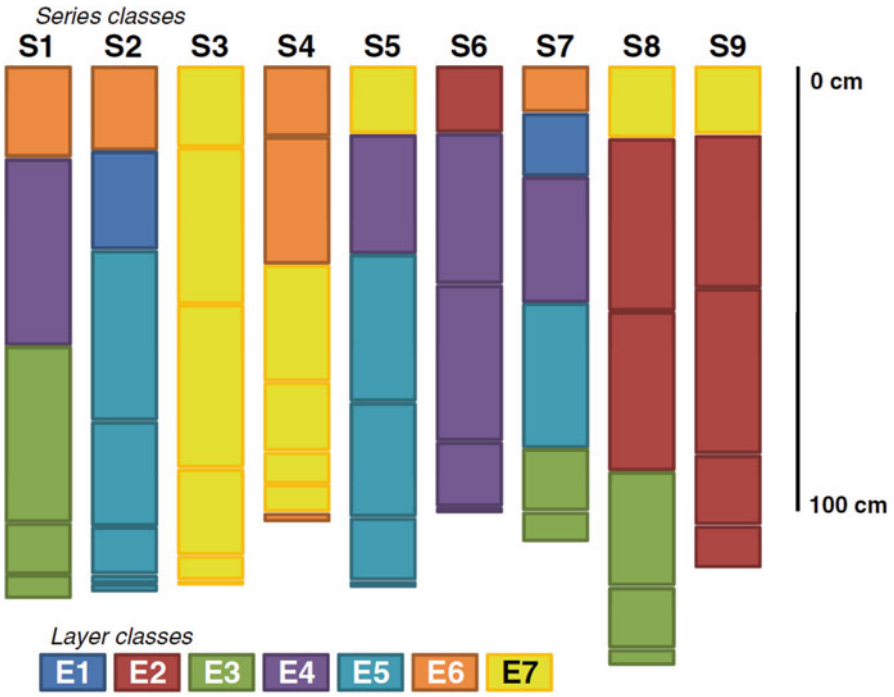


Fig. 9.9 Modal soil profiles of a set of numerical soil classes from the lower Hunter Valley in New South Wales, Australia, created using the OSACA algorithm (Reproduced with permission from Odgers et al. 2011a)

profiles that consisted of common sequences of the layer classes, but later work simplified the algorithm by removing this functionality (Carré and Jacobson 2009). Odgers et al. (2011a) demonstrated that some of this functionality could be restored by attributing the horizons of the profiles to be classified with fuzzy memberships to a set of layer classes created in a separate analysis (Fig. 9.9; Odgers et al. 2011b).

Massawe et al. (2016a, b) used machine learning techniques to spatially predict soil profile classes generated by OSACA across a valley in Tanzania. Rizzo et al. (2014) used OSACA to classify some Brazilian soil profiles by using the first four principal components of visible near-infrared spectra (350–2,500 nm wavelength at 1 nm resolution) as the attributes of each pedogenetic horizon. They compared the classes that OSACA generated with classes at several levels of the Brazilian soil classification and found that the OSACA classes were most similar to classes at the suborder level.

References

Alewell C, Meusburger K, Juretzko G, Mabit L, Ketterer ME (2014) Suitability of 239+240Pu and 137Cs as tracers for soil erosion assessment in mountain grasslands. Chemosphere 103: 274–280. doi:10.1016/j.chemosphere.2013.12.016

Anderson AJB (1971) Numeric examination of multivariate soil samples. Math Geol 3:1–14

Anderson MJ, Clements A (2000) Resolving environmental disputes: a statistical method for choosing among competing cluster models. Ecol Appl 10:1341–1355. doi:10.1890/1051-0761(2000)010[1341:REDASM]2.0.CO;2

Arkley RJ (1976) Statistical methods in soil classification research. Adv Agron 28:37–70. doi:10.1016/S0065-2113(08)60552-0

Arrouays D, Pelissier P (1994) Modeling carbon storage profiles in temperate forest humic loamy soils of France. Soil Sci 157:185–192

Banfield CF, Bascomb CL (1976) Variability in three areas of the Denchworth soil map unit. II. Relationships between soil properties and similarities between profiles using laboratory measurements and field observations. J Soil Sci 27:438–450. doi:10.1111/j.1365-2389.1976.tb02013.x

Beaudette DE, Roudier P, O'Geen AT (2013) Algorithms for quantitative pedology: a toolkit for soil scientists. Comput Geosci 52:258–268. doi:10.1016/j.cageo.2012.10.020

Bernoux M, Arrouays D, Cerri CC, Bourennane H (1998) Modeling vertical distribution of carbon in Oxisols of the western Brazilian Amazon (Rondonia). Soil Sci 163:941–951

Bidwell OW, Hole FD (1964) An experiment in the numerical classification of some Kansas soils. Soil Sci Soc Am Proc 28:263–268

Bishop TFA, McBratney AB, Laslett GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45. doi:10.1016/S0016-7061(99)00003-8

Blackith RE, Reyment RA (1971) Multivariate morphometrics. Academic, London

Bottner P, Grandjouan G, Nedelka E (1975) Classification des sols par une méthode multivariable. Application à une séquence bioclimatique méditerranéo-alpine sur roches-mères calcaires. Geoderma 14:15–46. doi:10.1016/0016-7061(75)90011-7

Brantley SL, Bandstra J, Moore J, White AF (2008) Modelling chemical depletion profiles in regolith. Geoderma 145:494–504. doi:10.1016/j.geoderma.2008.02.010

Brewer R (1968) Clay illuviation as a factor in particle-size differentiation in soil profiles. In: 9th international congress of soil science. Adelaide, South Australia, p 489–499

Campbell NA, Mulcahy MJ, McArthur WM (1970) Numerical classification of soil profiles on the basis of field morphological properties. Aust J Soil Res 8:43–58. doi:10.1071/SR9700043

Carré F, Girard MC (2002) Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110:241–263. doi:10.1016/S0016-7061(02)00233-1

Carré F, Jacobson M (2009) Numerical classification of soil profile data using distance metrics. Geoderma 148:336–345. doi:10.1016/j.geoderma.2008.11.008

Cattell R (1944) A note on correlation clusters and cluster search methods. Psychometrika 9: 169–184. doi:10.1007/BF02288721

Colwell JD (1970) A statistical-chemical characterization of four great soil groups in southern New South Wales based on orthogonal polynomials. Aust J Soil Res 20:221–238

Crépin J, Johnson RL (1993) In: Carter MR (ed) Soil sampling for environmental assessment. CRC Press, Boca Raton, pp 5–18

Dale MB, MacNaughton-Smith P, Williams WT, Lance GN (1970) Numerical classification of sequences. Aust Comput J 2:9–13

Diday E (1971) Une nouvelle méthode en classification automatique et reconnaissance des formes la méthode des nuées dynamiques. Rev Stat Appl 19:19–33

Dwyer LM, Ma BL, Stewart DW, Hayhoe HN, Balchin D, Culley JLB, McGovern M (1996) Root mass distribution under conventional and conservation tillage. Can J Soil Sci 76:23–28

Erh KT (1972) Application of the spline function to soil science. Soil Sci 114:333–338
FitzPatrick EA (1976) Soil horizons and homology. Classif Soc Bull 3:68–89
Frogbrook ZL, Oliver MA (2007) Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. Soil Use Manag 23:40–51. doi:10.1111/j.1475-2743.2006.00065.x
Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338. doi:10.1093/biomet/53.3-4.325
Grigal DF, Arneman HF (1969) Numerical classification of some forested Minnesota soils. Soil Sci Soc Am Proc 33:433–438
Haverkamp R, Vauclin M, Touma J, Wierenga PJ, Vachaud G (1977) A comparison of numerical simulation models for one-dimensional infiltration. Soil Sci Soc Am J 41:285–294. doi:10.2136/sssaj1977.03615995004100020024x
He Q, Walling DE (1997) The distribution of fallout 137Cs and 210Pb in undisturbed and cultivated soils. Appl Radiat Isot 48:677–690. doi:10.1016/S0969-8043(96)00302-8
Hills RG, Porro I, Hudson DB, Wierenga PJ (1989) Modeling one-dimensional infiltration into very dry soils: 1. Model development and evaluation. Water Resour Res 25:1259–1269. doi:10.1029/WR025i006p01259
Hole FD, Hironaka M (1960) An experiment in ordination of some soil profiles. Soil Sci Soc Am Proc 24:309–312
Horn D (1944) A study of personality syndromes. Character Pers 12:257–274. doi:10.1111/j.1467-6494.1944.tb01962.x
Isbell RF (1996) The Australian soil classification. CSIRO Publishing, Australia, p 152
Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaudoise Sci Nat 44:223–270. doi:10.5169/seals-268384
Jobbagy EG, Jackson RB (2000) The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol Appl 10:423–436. doi:10.2307/2641104
Jury WA (1982) Simulation of solute transport using a transfer function model. Water Resour Res 18:363–368. doi:10.1029/WR018i002p00363
Kalisz PJ, Zimmerman RW, Muller RN (1987) Root density, abundance, and distribution in the mixed mesophytic forest of eastern Kentucky. Soil Sci Soc Am J 51:220–225. doi:10.2136/sssaj1987.03615995005100010045x
Kastanek FJ, Nielsen DR (2001) Description of soil water characteristics using cubic spline interpolation. Soil Sci Soc Am J 65:279–283. doi:10.2136/sssaj2001.652279x
Kaufman L, Rousseeuw PJ (1987) Clustering by means of medoids. North Holland/Elsevier, Amsterdam, pp 405–416
Kirkby MJ (1985) A basis for soil profile modelling in a geomorphic context. J Soil Sci 36:97–121. doi:10.1111/j.1365-2389.1985.tb00316.x
Koarashi J, Atarashi-Andoh M, Matsunaga T, Sato T, Nagao S, Nagai H (2012) Factors affecting vertical distribution of Fukushima accident-derived radiocesium in soil under different land-use conditions. Sci Total Environ 431:392–401. doi:10.1016/j.scitotenv.2012.05.041
Leblanc MA, Gagné G, Parent LE, Hartemink AE, McBratney AB (2016) Numerical clustering of soil series using profile morphological attributes for potato. In: Hartemink AE, Minasny B (eds) Progress in soil science. Springer International, p 253–266
Lefkovitch LP (1976) Hierarchical clustering from principal coordinates: an efficient method for small to very large numbers of objects. Math Biosci 31:157–174. doi:10.1016/0025-5564(76)90047-X
Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals. Sov Phys—Dokl 10:707–710
Little IP, Ross DR (1985) The Levenshtein metric, a new means for soil classification tested by data from a sand-podzol chronosequence and evaluated by discriminant function analysis. Aust J Soil Res 23:115–130. doi:10.1071/SR9850115
Lu S-Y, Fu KS (1978) A sentence-to-sentence clustering procedure for pattern analysis. IEEE Trans Syst Man Cybern 8:381–389. doi:10.1109/TSMC.1978.4309979

Malone BP, McBratney AB, Minasny B, Laslett GM (2009) Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154:138–152. doi:10.1016/j.geoderma.2009.10.007

Massawe BHJ, Slater BK, Subburayalu SK, Kaaya AK, Winowiecki L (2016a) In: Lal R, Kraybill D, Hansen DO, Singh BR, Mosogoya T, Eik LO (eds) Updating legacy soil maps for climate resilient agriculture: a case study of Kilombero Valley, Tanzania. Springer International, Cham, pp 345–364

Massawe BHJ, Subburayalu SK, Kaaya AK, Winowiecki L, Slater BK (2016b) Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. Geoderma. in press. doi:10.1016/j.geoderma.2016.11.020

Mestdagh I, Lootens P, van Cleemput O, Carlier L (2004) Soil organic carbon stocks in Flemish grasslands: how accurate are they? Grassl Forage Sci 59:310–317. doi:10.1111/j.1365-2494.2004.00432.x

Mikhailova EA, Post CJ (2006) Organic carbon stocks in the Russian Chernozem. Eur J Soil Sci 57:330–336. doi:10.1111/j.1365-2389.2005.00741.x

Minasny B, McBratney AB, Mendonça Santos M d L, Odeh IOA, Guyon B (2006) Prediction and digital mapping of soil carbon storage in the lower Namoi Valley. Aust J Soil Res 44:233–244. doi:10.1071/SR05136

Minasny B, Stockmann U, Hartemink AE, McBratney AB, Hartemink AE, McBratney AB (2016) Measuring and modelling soil depth functions. In: Hartemink AE, Minasny B (eds) Progress in soil science. Springer International, p 225–240

Mishra U, Lal R, Slater BK, Calhoun F, Liu D, Van Meirvenne M (2009) Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. Soil Sci Soc Am J 73:614–621. doi:10.2136/sssaj2007.0410

Moore AW, Russell JS, Ward WT (1972) Numerical analysis of soils: a comparison of three soil profile models with field classification. J Soil Sci 23:193–209. doi:10.1111/j.1365-2389.1972.tb01653.x

Mora F, Iverson L (2002) A spatially constrained ecological classification: rationale, methodology and implementation. Plant Ecol 158:153–169. doi:10.1023/A:1015534615415

Muir JW, Hardie HGM, Inkson RHE, Anderson AJB (1970) The classification of soil profiles by traditional and numerical methods. Geoderma 4:81–90. doi:10.1016/0016-7061(70)90035-2

Myers DB, Kitchen NR, Sudduth KA, Miles RJ, Sadler EJ, Grunwald S (2011) Peak functions for modeling high resolution soil profile data. Geoderma 166:74–83. doi:10.1016/j.geoderma.2011.07.014

Norris JM, Dale MB (1971) Transition matrix approach to numerical classification of soil profiles. Soil Sci Soc Am Proc 35:487–491. doi:10.2136/sssaj1971.353487x

Odgers NP, McBratney AB, Minasny B (2011b) Bottom-up digital soil mapping. II. Soil series classes. Geoderma 163. doi:10.1016/j.geoderma.2011.03.013

Odgers NP, McBratney AB, Minasny B, (2011a) Bottom-up digital soil mapping. I. Soil layer classes. Geoderma 163. doi:10.1016/j.geoderma.2011.03.014

Oliver MA, Webster R (1989) A geostatistical basis for spatial weighting in multivariate classification. Math Geol 21:15–35. doi:10.1007/BF00897238

Ponce-Hernandez R, Marriott FHC, Beckett PHT (1986) An improved method for reconstructing a soil profile from analyses of a small number of samples. J Soil Sci 37:455–467. doi:10.1111/j.1365-2389.1986.tb00377.x

Ramzaev V, Mishine A, Golikov V, Brown JE, Strand P (2007) Surface ground contamination and soil vertical distribution of 137Cs around two underground nuclear explosion sites in the Asian Arctic, Russia. J Environ Radioact 92:123–143. doi:10.1016/j.jenvrad.2006.10.001

Rayner JH (1966) Classification of soils by numerical methods. J Soil Sci 17:79–92. doi:10.1111/j.1365-2389.1966.tb01454.x

Rizzo R, Demattê JA, da Silva Terra F (2014) Using numerical classification of profiles based on Vis-NIR spectra to distinguish soils from the Piracicaba Region, Brazil. Rev Bras Ciência Solo 38:372–385. doi:10.1590/S0100-06832014000200002

Russell JS, Moore AW (1968) Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data. In: 9th International congress of soil science. Adelaide, South Australia, p 205–213

Schmidtlein S, Tichý L, Feilhauer H, Faude U (2010) A brute-force approach to vegetation classification. J Veg Sci 21:1162–1171. doi:10.1111/j.1654-1103.2010.01221.x

Simonson RW (1968) Concept of soil. Adv Agron 20:1–47

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification, A series of books in biology. W. H. Freeman and Company, San Francisco

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409–1438

Tabatabaeenejad A, Burgin M, Duan X, Moghaddam M (2015) P-band radar retrieval of subsurface soil moisture profile as a second-order polynomial: first AirMOSS results. IEEE Trans Geosci Remote Sens 53:645–658. doi:10.1109/TGRS.2014.2326839

Taylor JA, Minasny B (2006) A protocol for converting qualitative point soil pit survey data into continuous soil property maps. Aust J Soil Res 44:543–550. doi:10.1071/SR06060

Tichý L, Chytrý M, Smarda P (2011) Evaluating the stability of the classification of community data. Ecography (Cop) 34:807–813. doi:10.1111/j.1600-0587.2010.06599.x

Triantafilis J, Huckel AI, Odeh IOA (2001) Comparison of statistical prediction methods for estimating field-scale clay content using different combinations of ancillary variables. Soil Sci 166:415–427

Veronesi F, Corstanje R, Mayr T (2012) Mapping soil compaction in 3D with depth functions. Soil Tillage Res 124:111–118. doi:10.1016/j.still.2012.05.009

Walker PH, Green P (1976) Soil trends in two valley sequences. Aust J Soil Res 14:291–303. doi:10.1071/SR9760291

Webster R (1977) Quantitative and numerical methods in soil classification and survey. Oxford University Press, Oxford

Webster R (1978) Mathematical treatment of soil information. 11th Int. Congr. Soil Sci

Webster R, Butler BE (1976) Soil classification and survey studies at Ginninderra. Aust J Soil Res 14:1–24. doi:10.1071/SR9760001

Wesseling JG, Ritsema CJ, Stolte J, Oostindie K, Dekker LW (2008) Describing the soil physical characteristics of soil samples with cubical splines. Transp Porous Media 71:289–309. doi:10.1007/s11242-007-9126-3

Wetselaar R (1962) Nitrate distribution in tropical soils: III. Downward movement and accumulation of nitrate in the subsoil. Plant Soil 16:19–31. doi:10.1007/BF01378155

Wickramagamage P (1986) Characterization of soil profiles for numerical classification. J Natl Sci Counc Sri Lanka 14:233–250

Wickramagamage P, Fisher GC (1988) The numerical classification of soils: a case study using data from West Sussex, England. J Soil Sci 39:139–153. doi:10.1111/j.1365-2389.1988.tb01201.x

Young FJ, Hammer RD (2000) Defining geographic soil bodies by landscape position, soil taxonomy, and cluster analysis. Soil Sci Soc Am J 64:989–998

Part V
Soil Variation in Space and Time

"Sir Walter Blunt, new-lighted from his horse,
Stain 'd with the variation of each soil
Betwixt that Holmedon and this seat of ours."

Shakespeare,
King Henry IV (Part I)

A good introductory explanation of geostatistics is given by Webster and Oliver (2007) and also by Isaaks and Srivastava (1989). A much more advanced treatment illustrated with soil chemical data is contained in Goovaerts (1997).

One of the most important aspects of pedometrics is the quantitative description modelling and prediction of the spatial and temporal variation of soil properties.

As was discussed in Chap. 1, soil variation has been known since recorded history. With an increasing demand for both better economic management and environmental management of the soil resource, the need for quantitative information about soil within agricultural fields within catchments across regions and continents has grown apace. The approaches we describe in Chaps. 9, 10, 12 and 13 are relevant to these problems and scales. When we move to finer scales, more geometric approaches required as described in Chap. 14.

We also recognise that soil variation, which many see as problematic, can be regarded positively as pedodiversity – in the same way as we think of biodiversity. Chapter 15 deals with the measurement and representation of pedodiversity.

References

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Isaaks E, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York

Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester

Chapter 10
Classical Soil Geostatistics

R. Murray Lark and Budiman Minasny

"All the business of war, and indeed all the business of life, is to endeavour to find out what you don't know by what you do".

Arthur Wellesley
Duke of Wellington

10.1 Introduction

In 1971 Beckett and Webster reviewed the information on soil variability available to them. Their interest was, to use their term, in *lateral* variability of the soil. That is to say in the variation of the soil from place to place across the landscape (Beckett and Webster 1971). One way to capture this notion is by considering the variability of the soil as measured by sample variance, standard deviation or coefficient of variation (CV) within regions of different sizes. If the spatial location of soil observations is immaterial to their variation, then the variance and other quantities will be the same within regions of any size. Beckett and Webster (1971) found that, typically, half the variation of a soil property within an agricultural field, measured from conventional core samples, may be found within an area of 1 m^2. This shows us that the variability of the soil is spatially very intricate but also that a significant amount of variation is spatially structured over the intermediate scales. The fact that CV depends on area tells us something about the variability of soil. However, in the absence of intensive sensor measurements, the CV/area relationship is a cumbersome tool for describing spatial variability, requiring, in principle, sets of random samples from

R.M. Lark (✉)
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
e-mail: mlark@bgs.ac.uk

B. Minasny
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: budiman.minasny@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_10

within randomly selected areas of different size. In a pioneering study Beckett and Bie (1976) sampled the soil on regular transects. They then grouped the observations into successive pairs, groups of three samples, groups of four samples, groups of n samples, etc. and for each $n = 2, 3, \ldots$ calculated the mean variances within the groups of n observations. Variance was then plotted against the length of the aggregated groups (log scale). It was shown that the form of the graphs related soil variability at different scales to vegetation and geomorphology in a predictable way.

This approach to analysis of spatial variability was innovative, developed from statistical literature available at the time (e.g. Yates 1948). However, other older statistical work had shown how the formalisms of analysis of variance could be used to describe spatial variation. This is the work of Youden and Mehlich (1937) who devised a sampling scheme to allow a partition of variance of a soil property between contrasts over different spatial intervals. This was achieved by a nested sampling design permitting a nested analysis in which, at each stage, the variance is partitioned into that observable between and within sampling stations separated by a particular interval. Webster and Butler (1976) revived the methodology and applied it to the analysis of soil variation in Australian Capital Territory, in a landscape where soil surveyors had struggled to represent the variation of soil by conventional soil maps. Webster and Lark (2013) discuss this sampling scheme and analysis in more detail.

While soil scientists were developing these innovations, statisticians in the mining industry, geostatisticians, were developing a formalism for the description of spatial variation which embraced and unified both the nested sampling of Youden and Mehlich and the transect analysis of Beckett and Bie. Their objective was to predict ore grades locally from a limited number of boreholes to support mine planning. The methods that they developed constitute what we might call classical geostatistics or mining geostatistics. The textbook of Journel and Huijbregts (1978) sets out the stall of these methods. Around the time that it was published, soil scientists became aware of geostatistics and the potential of geostatistical methods to solve the problems of soil survey: spatial prediction of soil conditions from limited sets of observations. Webster (2015) tells the story of how the connection between the armoury of mining geostatistical methods and the requirements of soil survey were first made.

This chapter discusses the classical methods of geostatistics as applied to soil science. It provides background to some of the more modern methods described in Chap. 11, but much of this standard methodology remains serviceable for the problem of spatial prediction. We first describe the random model of soil variation that underlies geostatistics and then show how this can be used for spatial prediction and inference.

10.2 Random Models of Soil Variation

This section presents a summary of the principles of geostatistical method, as set out by standard texts such as Journel and Huijbregts (1978), Isaaks and Srivastava (1989), Goovaerts (1997), Chiles and Delfiner (1999) and Webster and Oliver

(2007). Classical references on this methodology in the soil science literature are the papers of Burgess and Webster (1980a, b), McBratney and Webster (1983) and Webster and Burgess (1980).

10.2.1 Stationary Random Functions

A powerful approach to problems of inference is to treat data as if they are a realisation of a set of random variables. A simple example of a random variable is the number given by the throw of an unbiased die which is a random number, Y, which takes some value from the set $\{1,2,\ldots 6\}$. In the case of soil, we assume that a soil property s at location $\mathbf{x}$ is a realisation of a random variable. However, we can only ever observe one realisation, $s(\mathbf{x})$, the actual value of s at $\mathbf{x}$, which is not sufficient information to characterise a random variable. We can make progress if we assume there is a realisation not of a single variable but rather of a random variable $S(\mathbf{x})$ when the argument of the random function is location in space. If we can assume that certain properties of this function are constant for all $\mathbf{x}$ of interest, then a set of observations at different locations contains some replicated information from which we may infer properties of the random function.

10.2.1.1 Stationarity

The simplest, and strongest, assumption of stationarity is that the joint distribution of the random function over a set of locations, $\{S(\mathbf{x}_1),S(\mathbf{x}_2),\ldots S(\mathbf{x}_n)\}$, is identical to that for a set $\{S(\mathbf{x}_1+\mathbf{h}),S(\mathbf{x}_2+\mathbf{h}),\ldots S(\mathbf{x}_n+\mathbf{h})\}$ where $\mathbf{h}$ is any displacement or lag vector. This means that all moments of the distribution are constant so the mean $\mu = \mathrm{E}[S(\mathbf{x})]$, for all $\mathbf{x}$ and the covariance $\mathrm{E}[\{S(\mathbf{x})-\mu\}\{S(\mathbf{x}+\mathbf{h})-\mu\}]$, is constant for all $\mathbf{x}$ and $\mathbf{h}$, as are all higher moments. Note that E[.] denotes the statistical expectation of a random quantity in the square brackets. If we restrict these assumptions to the first two moments of the joint distribution (mean and covariances), then we have a weaker stationarity assumption – second-order or weak stationarity. Under the assumption of stationarity, it is therefore possible to define a covariance function

$$C(\mathbf{h}) = E\left[\{S(\mathbf{x})-\mu\}\{S(\mathbf{x}+\mathbf{h})-\mu\}\right]. \tag{10.1}$$

At lag zero the covariance function is equal to the variance of the random function. If the random variable shows no spatial dependence (it is a 'white noise' process in the terms of signal analysis or 'pure nugget' in geostatistical terms), then it is zero for all lag vectors which are non-zero. Spatial dependence is shown when the covariance declines with increasing lag separation reaching value zero for lags at which values of the random function are independent. Note that in some circumstances the covariance may not be strictly decreasing with lag distance, there may be a periodic component to the variation or quasiperiodic 'hole effect'.

It is possible to make a stationarity assumption weaker still (and so a plausible assumption about a broader class of soil variables). Intrinsic stationarity is, in effect, a second-order stationarity assumption about the increments $S(\mathbf{x}_i) - S(\mathbf{x}_i + \mathbf{h})$ $S(\boldsymbol{x}_i) - S(\boldsymbol{x}_i + \boldsymbol{h})$. So we assume that the mean increment is zero everywhere and that the variance of the increment is constant everywhere. Hence

$$E\left[\{S(\mathbf{x}) - S(\mathbf{x} + \mathbf{h})\}\right] = 0$$

and

$$E\left[\{S(\mathbf{x}) - S(\mathbf{x} + \mathbf{h})\}^2\right] = 2\gamma(\mathbf{h}). \tag{10.2}$$

The function $\gamma(\mathbf{h})$ given above is the semi-variogram. It is related to the covariance function in the second-order stationary case by

$$C(\mathbf{h}) = C(0) - \gamma(\mathbf{h}). \tag{10.3}$$

The covariance at lag 0, $C(0)$, is the variance of independently drawn values of S(**x**), also known as the a priori variance. We can use the variogram in a wider case of processes than the second- or higher-order stationary ones that can be described by the autocorrelation or covariance functions.

10.2.1.2 Variogram

Under the intrinsic hypothesis of stationarity, $\frac{1}{2}\{s(\mathbf{x}) - s(\mathbf{x} + \mathbf{h})\}^2$ and $\frac{1}{2}\{s(\mathbf{x}') - s(\mathbf{x}' + \mathbf{h})\}^2$ are both estimates of $\gamma(\mathbf{h})$. We may, therefore, combine all observations over lag **h** into an estimate of $\gamma(\mathbf{h})$. This is illustrated for a transect in Fig. 10.1. In the top row, the pairs of observations are combined to estimate the variogram for the lag equal to the basic sample interval. In the second row, the pairs

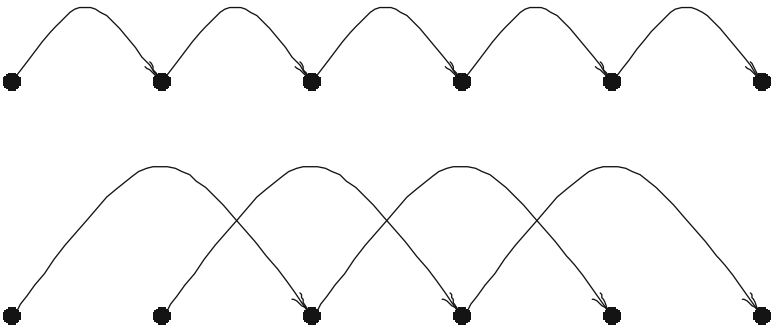


Fig. 10.1 Pair comparisons between sample points on a linear transect at intervals of (*top*) one times the basic spacing and (*bottom*) two times the basic spacing

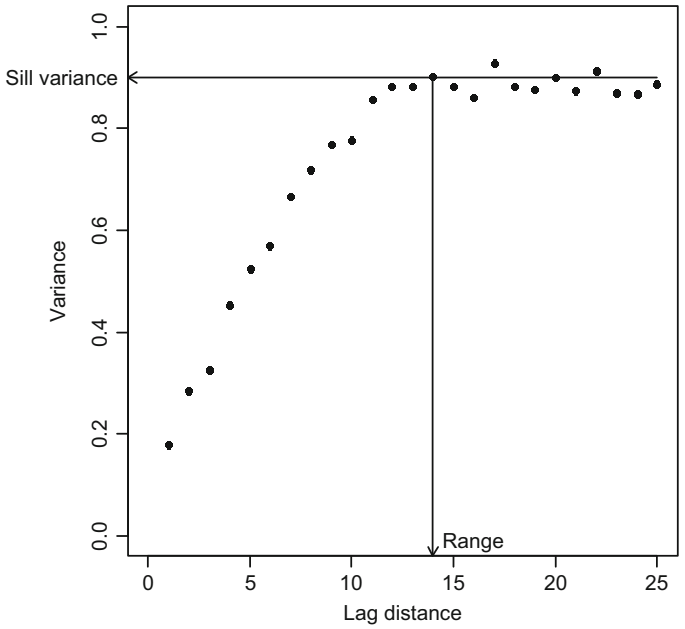


Fig. 10.2 An empirical variogram plot illustrating the concepts of sill variance and range

of observations are separated by twice the sample interval. On a transect the lag is a scalar and is written h. If h is some integer multiple of the basic sampling interval on the transect, then $\gamma(h)$ is estimated by taking all the N_h pairs separated by h and calculating from them half the mean-squared paired difference:

$$\widehat{\gamma}(h) = \frac{1}{2}\sum_{i=1}^{N_h}\{s(x_i) - s(x_i + h)\}^2. \tag{10.4}$$

We call this the empirical, or experimental variogram. Note that we can use the scalar lag in the analysis of data in two or more dimensions if we assume that the spatial autocorrelation of our variable depends on distance only and not direction. This is the assumption of isotropy. We consider isotropic variograms in most of the following discussion and address how to model directional dependence later in the chapter.

Consider Fig. 10.2. This empirical variogram increases with lag distance to a maximum value – the *sill* variance or *a priori* variance of the random function. In this particular figure, this happens at a particular distance – the *range*. If we consider Eq. 10.3, we see that at lags larger than the range, the autocovariance is zero. That is to say two points in space separated by a distance larger than the range are uncorrelated. At shorter distances, they will tend to be correlated. In this way the variogram describes the structure of spatial variability.

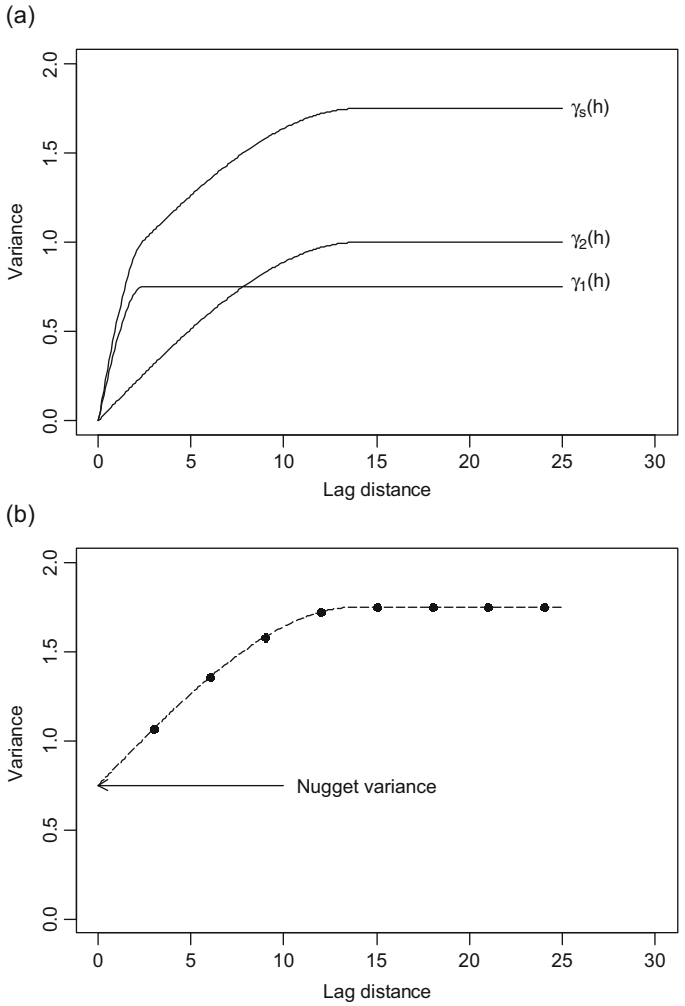


Fig. 10.3 (**a**) Two variograms, $\gamma_1(h)$ and $\gamma_2(h)$ for two mutually independent random variables with contrasting ranges of spatial dependence and the variogram of the sum of these two variables, $\gamma_s(h)$. (**b**) Points from the expected empirical variogram of the sum of the two variables, when sampled at basic interval 3 units. Note that the spatial structure of $\gamma_1(h)$ cannot be resolved, and its variance contributes an apparent intercept to the empirical variogram, the nugget variance

Imagine two isotropic random fields, statistically independent of each other, but with variograms $\gamma_1(h)$ and $\gamma_2(h)$. If we form a new random function by adding these two, then its variogram, $\gamma_s(h)$, is simply the sum, $\gamma_1(h) + \gamma_2(h)$. We describe the random function as a nested random function with a nested variogram. See Fig. 10.3. Note that, while nested variogram functions are commonly used in classical geostatistics and have an intuitive appeal representing combined effects of random processes at contrasting scale, they have been criticised in the setting of model-based geostatistics, a treatment of which is given in Chap. 11. See also Stein (1999).

Consider a soil process which can be regarded as the sum of two independent random functions with different ranges and with variograms $\gamma_1(h)$ and $\gamma_2(h)$. Its variogram, $\gamma_s(h)$, as seen above and as illustrated in Fig. 10.3a is the sum of the two variograms. If we sampled at a basic interval between the ranges of the two variograms, then we may obtain point estimates of the variogram as shown in Fig. 10.3b. The empirical variogram obtained under this sampling scheme resolves the structure of component 2 but not 1. The effect of random function 1 is that the experimental variogram appears to have a non-zero intercept, equal to the sill. This is called the *nugget* variance. Nugget variance will include elements due to measurement error, but as illustrated above, it also includes components of spatial variation which are spatially dependent at scales too fine to be resolved by our sampling scheme. It is possible to partition the nugget between these sources only if we have some independent estimate of the measurement error.

10.2.1.3 Variogram Models

Equation 10.4 above generates point estimates of the variogram for particular lags. In practice we need values of the variogram for any lag. This requires that we can express the variogram as a function of lag. At first glance this might seem like a simple problem in curve fitting. In reality there is a complication. The covariance function allows us to compute the variance of any linear combination of values of a random variable. Now the variance of a linear combination of values of a random function, subject to random variation, must have a variance which is positive and non-zero. We could write down a function of h that does not guarantee this (and so cannot actually be the covariance function of any actual random function). Such a covariance function is said to be non-positive definite. We avoid this by using variograms that are negative semi-definite (negative because the variogram is equal to the a priori variance minus the covariance function and semi-definite because the covariance function may not be zero at all lags). Variogram functions which meet this criterion are said to be authorised. We now describe some commonly used models.

10.2.1.4 Nugget

Imagine a situation where all the spatial dependence for random function is at scales finer than the basic sampling interval. In signal analysis such variation is called 'white noise'. The variogram will be flat, pure nugget. In practice such variograms are rare, but nugget models are almost always included as an additive term in a nested model to describe the variation which has not been resolved by sampling. The nugget variogram model may be written as

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \mathbf{h} = 0 \\ c_0 & \text{otherwise} \end{cases}. \tag{10.5}$$

Note that in practice, the nugget term appears as an intercept (i.e. a value at lag zero), but all variogram models must be zero at lag zero by definition. This is a practical importance since the variogram at lag zero appears in the kriging equation.

The nugget model is a model of discontinuance variation since the variogram jumps from zero to a finite value for any finite lag. This discontinuity can affect predictions of soil properties using a variogram with a significant nugget. When we invoke a nugget model, however, we are not necessarily claiming that the variation of a property is discontinuous at the limit, although it might be. The term nugget effect comes from the case of a nugget of pure gold embedded in rock with a discontinuous step in grade at its edge. All we can say in practice is that variation appears discontinuous at the scale of resolution of our data.

10.2.1.5 The Exponential Variogram

Imagine a linear transect on the ground intercepted at random by boundaries between regions within which the value of the soil property is a uniform value drawn from a random process of variance c_1. If the probability that two points on the transect separated by lag distance h lie either side of at least one such boundary is $p(h)$, then it is clear that the variogram of the process will be

$$\gamma(h) = c_1 p(h). \tag{10.6}$$

Let us assume that the boundaries occur at random as a Poisson process and that the mean interval between two boundaries is a. The mean number of boundaries falling on an interval of the transect of length h is therefore h/a. Under the Poisson distribution, the probability that *no* boundary falls on such an interval is

$$\left\{\left(\frac{h}{a}\right)^0 e^{-h/a}\right\} \Big/ 0! = e^{-h/a}. \tag{10.7}$$

The probability at one or more boundaries falls as such an interval is therefore $1 - e^{-h/a}$, so our variogram function is

$$\gamma(h) = c_1 \left\{1 - e^{-h/a}\right\}. \tag{10.8}$$

This is the exponential variogram for a process of variance c_1 and with distance parameter a. As the derivation shows, it describes the most basic concept of spatial randomness. The exponential function is a bounded one, with an upper bound at c_1, but it approaches this asymptotically so we cannot define a finite range at which $\gamma(h) = c_1$. Since $\gamma(h) \approx 0.95 c_1$ where $h = 3a$, it is common to define $3a$ as the *effective* range. Note that while we may argue from the Poisson boundary process to the exponential variogram, it is hazardous to reverse the argument when such a

variogram is found. Other spatial random processes will give rise to an exponential variogram – for example, a first-order auto-regressive process:

$$s(x_i) = \alpha + \beta\, s(x_{i-1}) + \varepsilon_i \tag{10.9}$$

where x_{i-1}, x_i are successive, equally spaced locations on a transect defines, at the limit as the interval between successive values approaches zero, a random process with an exponential variogram. The terms α and β are coefficients and ε_1 is an independent random variable of mean zero.

10.2.1.6 Bounded Linear Model

We might divide a transect into intervals of equal length, a, and then allocate to all sites within any segment a uniform value drawn from a random process with variance c_1. From Eq. 10.2 we may write the variogram of this process as

$$\gamma(h) = c_1 \left\{ \begin{array}{ll} h/a & \text{if } h \leq a \\ 1 & \text{otherwise} \end{array} \right\}. \tag{10.10}$$

This is a bounded linear variogram of range a. Burrough (1983) discusses this model and fitted it to data where a more or less regular pattern of geological boundaries was a dominant source of soil variation. As with the exponential model, however, this variogram can arise from a contrasting kind of spatial process. Imagine that we generated a random function on a transect with independent random values. This would have a nugget variogram. Let us now smooth this process by replacing each value by the sample average of all the values at locations within $\pm a/2$. Webster and Oliver (2007) demonstrate that such a process will have a bounded linear variogram.

Note that the model can only describe spatial variation in one dimension. It is not negative semi-definite in two or more dimensions.

10.2.1.7 Circular and Spherical Models

We can extend the bounded linear model of spatial variability to two dimensions by imagining a field of independent random variables filtered by replacing each value with the simple average of all values within a distance a. This new field has a circular variogram function:

$$\gamma(h) = c_1 \left\{ \begin{array}{ll} 1 - \frac{2}{\pi}\cos^{-1}\{h/a\} + \frac{2h}{\pi\, a}\sqrt{1 - h^2/a^2} & \text{if } h \leq a \\ 1 & \text{otherwise} \end{array} \right\}. \tag{10.11}$$

This is negative semi-definite in one or two dimensions but not in three dimensions. Like the bounded linear model, it reaches a distinct range at distance a beyond which it is flat, although its slope decreases as it approaches the range. The three-dimensional equivalent, of course, is a spherical model:

$$\gamma(h) = c_1 \left\{ \begin{array}{cc} \frac{3h}{2a} - \frac{1}{2}\{h/a\}^3 & \text{if } h \leq a \\ 1 & \text{otherwise} \end{array} \right\}. \tag{10.12}$$

This is negative semi-definite in one, two or three dimensions. Although it is fundamentally a three-dimensional random process, it is commonly used to describe variograms of two-dimensional data when these have a distinct range. By extending into five dimensions, the pentaspherical model is defined.

So far the variogram models described are fairly straightforward in their behaviour, but they will not always fit observed variograms comfortably. At this juncture we must consider some niceties of the behaviour of variograms of random processes. A more detailed account is given by Webster and Oliver (2007).

Consider a random function $S(x)$. Because this function is random, we cannot compute a derivative $\frac{\mathrm{d}}{\mathrm{d}x}S(x)$, it is not differentiable. For this reason the variogram of a function is not differentiable at the origin, it has an approximate linear approach to the origin. On the other hand two derivatives of a parabolic function can be obtained at the origin. A parabolic variogram would describe smooth variation, i.e. variation that is entirely predictable so not in any sense random. A variogram which approaches this behaviour may arise because of some short-range deterministic variation (*drift*) or perhaps as an artefact arising from a measurement process which has a strong smoothing effect over short distances (yield monitor data, remote sensors and electromagnetic inductance measurements of the soil's electrical conductivity might have such an effect). Some variogram functions may, therefore, appear concave upwards near the origin, suggesting possible local drift. As seen above, parabolic behaviour of the origin is not consistent with random variation. We may define a power function variogram,

$$\gamma(h) = c_1 h^{\alpha}. \tag{10.13}$$

This is subject to the constraint that $0 < \alpha < 2$. If $\alpha = 1$ then the variogram is an unbounded linear function (which unlike the bounded linear function is negative semi-definite in two dimensions). For any authorised power function, the variability appears to increase without bound so the function cannot describe a weakly stationary process, only an intrinsically stationary one.

Some variograms are concave upwards near the origin but are bounded. The Stable model is often used to describe such behaviour:

$$\gamma(h) = c_1 \left\{1 - e^{-\frac{h^{\kappa}}{a^{\kappa}}}\right\}, \tag{10.14}$$

where $\kappa < 2$ is a constant. The Gaussian model is the Stable model with $\kappa = 2$. Although widely used, the behaviour of the Gaussian model at the origin is inconsistent with random variation; it should not be used for the geostatistical methods of prediction which are described below because of this behaviour which can lead to artefacts. A more satisfactory alternative for spatial variation which shows a certain smoothness is the Matérn function (Matérn 1986). The Matérn variogram model is

$$\gamma(h) = c_1 \left\{ \frac{1}{2^{\nu}\Gamma(\nu)} (\varphi\, h)^{\nu} K_{\nu}(\varphi\, h) \right\} \tag{10.15}$$

where $\Gamma(\cdot)$ and $K_{\nu}(\cdot)$ are, respectively, the gamma function and a modified Bessel function (second kind) of order ν. Smoothness of the random process is controlled by parameter ν. If $\nu \geq 0$, then the process is continuous. If $\nu = 0.5$, the Matérn variogram is the exponential. With larger ν it is smoother than the exponential, and as $\nu \to \infty$, the variogram function approaches the Gaussian. The Matérn function has been used increasingly in soil science (Minasny and McBratney 2005) but largely in the setting of model-based geostatistics. It is important that the model be used only when there are adequate observations at short lag distances.

The experimental variogram may appear to vary in a periodic way when the soil variation is controlled in part by some regular pattern such as ridge-and-furrow variation. A sine function is a semi-negative definite model in one dimension. It has gradient zero at lag zero which is unacceptable, but a sine function in combination with another semi-negative definite variogram will constitute an acceptable model. In two dimensions or more, the sine function is not negative semi-definite, but a damped function in which the fluctuations diminish the distance is negative semi-definite. A damped sinusoidal variogram is said to show a 'hole affect'. However, apparent fluctuations may be artefacts. Webster and Oliver (2007) recommend that periodic models are not used unless there is strong evidence for periodic behaviour, perhaps because of our a priori knowledge of the process. Some phenomena in soils can give rise to periodic fluctuations in the variogram, in particular "patterned ground" phenomena as formed in the Gilgai landscapes of Australia or in soils affected at some stage of their development by periglacial conditions.

It follows from our previous discussion of nested processes – additive combination of independent random functions – that two or more of the models discussed here may be combined to describe an experimental variogram. In fact most model variograms are a combination of a nugget variogram and a spatially structured one. But since any combination of negative semi-definite functions is itself negative semi-definite, other combinations are possible. Two models with different distance parameters may be combined to describe a variable with variability with different scales caused by different processes. A double spherical model is a common example of the nested structure with two distinct ranges.

10.2.1.8 Anisotropy

In our examples so far, the lag is a distance in space, i.e. a scalar, and we explained that a random function for which the variogram depends only on the scalar lag distance is said to be isotropic. In practice the direction of a comparison may affect the variance of the difference between two points, so our lags are vectors with both a distance and direction. This is called anisotropy. There are two kinds of anisotropy. In *geometric* or *affine* anisotropy, the sill variances are independent of direction, but the variogram does depend on direction at short distances. It therefore approaches the sill with different slopes at different directions. An affine transformation of the coordinate system will transform the variogram to isotropy. Under the model a locus that is the set of all locations $\mathbf{x}_i \in \mathbf{X}$ about a fixed point, $\mathbf{x}_0$, such that $\gamma(\mathbf{x}_0 - \mathbf{x}_i), \mathbf{x}_i \in \mathbf{X}$ is constant (and less than the sill variance) describes an ellipse, which can be transformed to a circle by the affine transformation of coordinates.

The other kind of anisotropy is *zonal* anisotropy. In zonal anisotropy the sill depends on direction. Such variation is not difficult to imagine. The variation and landscape scale where contrasting rocks outcrop with parallel strike will be largest perpendicular to strike and smallest parallel to the strike, but anisotropy will diminish as the lag tends to zero.

10.2.2 Estimating and Modelling the Variogram

10.2.2.1 Variogram Cloud

The variogram cloud is a plot of the individual values, $\frac{1}{2}\{s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})\}^2$, against the scalar lag, $h = |\mathbf{h}|$. Plotting and examining the variogram cloud can be useful in exploratory spatial analysis of the data. In particular, we may use it to examine evidence for anisotropy or to identify effects of a few outlying observations (Ploner 1999).

10.2.2.2 Lag Classes

In the example above, we illustrated the application of Matheron's (1962) estimator of the variogram with an idealised example of regularly sampled data in one dimension. In practice the problem is more complex, particularly in two dimensions or more when data are irregularly sampled and/or we wish to account for anisotropy. Matheron's estimator is now applied to pair comparisons $\{s(\mathbf{x}) - s(\mathbf{x} + \mathbf{h})\}$ where the lag vector is the mean or central vector of a lag class. Ignoring anisotropy the lag class may be defined by a range of lag distances or lag bin, a central distance plus a tolerance of $\pm w/2$ where w is the width of the bin. We may then take the mean lag distance within each class as the representative lag interval. Defining lag classes requires care and an element of trial and error. If the classes are defined

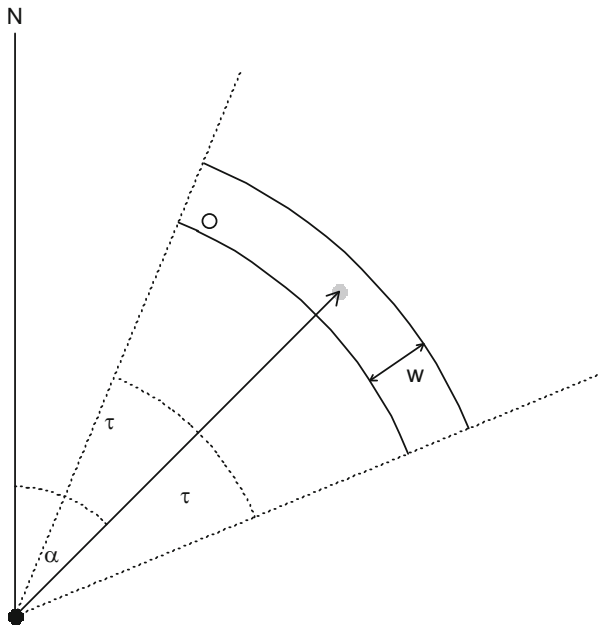


Fig. 10.4 Definition of a lag bin in two dimensions

too narrowly, then each will contain too few pair comparisons, and the resulting variogram estimates will be too noisy. If the classes are too wide, however, then they will smooth the spatial structure which is seen at lags within each interval and may obscure the underlying spatial dependence.

When anisotropy is a concern, lag classes must be defined relative to a compass bearing as well as to a distance. This is illustrated in Fig. 10.4. Consider the two points in space shown by a black and a grey disc. These are separated by a lag vector of scalar value (distance) h along a bearing of α from due north. With irregular sampling it is necessary to define lag bins with some tolerance into which to combine pair comparisons to form point estimates of the variogram. The tolerance is defined by a width of the lag distance bin ($\pm w/2$ in the Figure) and an angular tolerance of $\pm\tau$. With the tolerances specified, the lag between the black and the grey disc is the central lag for a lag bin such that any point within the region between the two solid arcs and the dotted lines indicating the angular tolerance would be separated from the black disc by a lag which belongs in the bin. The open circle illustrates such a point.

10.2.2.3 Estimating Variogram Parameters

The simplest variogram model to fit is an unbounded linear model, which may be fitted with an ordinary least squares criterion. However, most variogram models have non-linear parameters which must be estimated by more complex methods.

Nonetheless, the least squares criterion can be applied; that is to say we find a vector of variogram parameters, $\boldsymbol{\theta}$, such that the mean-squared difference between our point estimates of the variogram, $\widehat{\gamma}(h)$, and the corresponding model values, $\gamma(h|\boldsymbol{\theta})$, is minimised:

$$\text{MSE} = \frac{1}{N_l}\sum_{i=1}^{N_l}\{\widehat{\gamma}(h_i) - \gamma(h_i|\boldsymbol{\theta})\}^2, \tag{10.16}$$

where h_i is the representative lag for the ith lag class of which there are N_l, $\widehat{\gamma}(h)$ is the estimated variogram for this lag class and $\gamma(h|\boldsymbol{\theta})$ is the fitted model. We need an iterative algorithm to minimise this criterion for models with non-linear parameters.

Models can be fitted by eye and there is software available to help this process. However, fitting by eye is generally best avoided particularly where the variogram estimates are very variable and where the number of pair comparisons in each lag class varies so that different lag classes have different uncertainty. Visual inspection can be useful to guide the choice of a class of model and also to check that a fitted model seems reasonable.

Most variogram models are fitted by a weighted least squares method using a standard non-linear fitting algorithm to minimise a generalisation of *MSE*. If the experimental variogram for lag class i is supported by n_i pair comparisons, then one weighted *MSE* is

$$\text{MSE}_1 = \frac{1}{N_l}\sum_{i=1}^{N_l} n_i\{\widehat{\gamma}(h_i) - \gamma(h_i|\boldsymbol{\theta})\}^2, \tag{10.17}$$

so that we give more weight to a point estimate supported by many pair comparisons. Cressie (1985) proposed a development of this:

$$\text{MSE}_2 = \frac{1}{N_l}\sum_{i=1}^{N_l} \frac{n_i}{\gamma(h_i|\boldsymbol{\theta})^2}\{\widehat{\gamma}(h_i) - \gamma(h_i|\boldsymbol{\theta})\}^2. \tag{10.18}$$

This gives greater weight to the lags with small semi-variance, which is reasonable since the reliability of an estimate of variance is inversely related to its size, and the variogram at short lags is generally more influential when used in particular applications. Note that, since the model variogram appears in the weight, and it also depends on the parameters that we are estimating, the process must be done iteratively. The iterative procedure means that after one run of the algorithm, we use the fitted model to specify the modelled values. Generally only one iteration is required before the solution converges to a single set of model parameters.

As we have seen, there are many variogram models which can be fitted to data. Visual inspection of a set of point estimates and consideration of the factors causing the variation will generally enable us to narrow the field to a few plausible models, but we still have to choose between them. One basis for choice is to fit

all possible models and select that one for which the *MSE* is smallest. This is a rational procedure when all models have the same number of parameters (e.g. if we are comparing an exponential with a spherical model). However, we may have a choice between a double spherical model (two independent spherical processes) and an exponential. If we exclude the nugget (common to both), then the double spherical has four parameters to the exponential's two. These additional parameters give the model more flexibility and mean that the *MSE* for the double spherical will almost inevitably be smaller. This does not mean that the model is necessarily better, however. The question is whether the inclusion of more parameters is justified by the improvement of the fit which they allow. The Akaike information criterion (AIC) is a basis for making this judgement (Akaike 1973; McBratney and Webster 1986). For a given set of point estimates for the variogram, AIC will be least for the model for which

$$\widehat{A} = N_l \ln(MSE) + 2P \tag{10.19}$$

is smallest where P is the number of parameters. That model for which $\widehat{A}$ is smallest is judged most parsimonious. This means that an exponential model will be favoured over a double spherical unless the *MSE* for the double spherical, MSE_{DSP}, is so much smaller than that for the exponential, MSE_{Exp}, that

$$\ln \frac{MSE_{\mathrm{DSP}}}{MSE_{\mathrm{Exp}}} < -4N_l. \tag{10.20}$$

10.2.3 Departures from the Standard Model

10.2.3.1 Trends

It is worth recapitulating from Chap. 1 the basic model of soil variation with which we are working. In most general terms $S(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x})$ where $f(\mathbf{x})$ is some deterministic function of space and $\varepsilon(\mathbf{x})$ is a random variable, so $S(\mathbf{x})$ too has random properties. In our discussion so far, we assume that $f(\mathbf{x})$ is a constant, μ, and that $\varepsilon(\mathbf{x})$ is a spatially dependent random variable that can be regarded as a realisation of a stationary (strongly, weakly or intrinsically) random variable or as the sum of two or more independent such variables. The soil variation which we observed in reality is deterministic, the result of many processes, and may be poorly described by such a model. First μ may have to be replaced by a deterministic function $f(\mathbf{x})$. This could represent a pronounced trend, for example, a trend in the particle size distribution along a catena or in water content down a slope. Strong trends cannot be ignored. If $\gamma(\mathbf{h})$ is the variogram of $\varepsilon(\mathbf{x})$, then it is clearly not equal to $\frac{1}{2}E\left[\{(f(\mathbf{x}) + \varepsilon(\mathbf{x})) - (f(\mathbf{x}+\mathbf{h}) + \varepsilon(\mathbf{x}+\mathbf{h}))\}^2\right]$ when $f(\mathbf{x})$ is not a constant.

The experimental variogram will be affected by deterministic variation and becomes concave upwards. We may also find short-range drift or local trend which disappears at longer ranges. In either case we have to disentangle the variation into components which may be treated as deterministic and those which are treated as random. There is no unique way of doing this in any case; it is a matter of finding a particular model which is suitable given the variation and the uses to which we want to put it. We discuss how this might be done later, since it forms an important part of the process of spatial prediction.

10.2.3.2 Non-stationarity in the Random Variation

The term $\varepsilon(\mathbf{x})$ has constant mean by definition. However, in practice it may not be plausible to assume (as in the intrinsic hypothesis or any stronger stationarity assumption) that the variance is uniform. This may originate in many ways. Consider a landscape in which, as we move upslope from the river channel, the parent material changes from sorted alluvium to terrace gravels to clay to sandstone with wet flush zones and then chalk with pockets of overlying clay. It is clear that the variability of a soil property such as the saturated hydraulic conductivity or the water potential will change over the sequence. The development of sensors that generate large data sets, for example, on soil electrical conductivity, will allow us to model such complexity more effectively.

One approach is to compute local variograms within a moving window (Haas 1990; Corstanje et al. 2008). This allows us to account both for changes in the magnitude and spatial scale of variation. As the window moves across the landscape in one part of the landscape, the sill of the variogram might be relatively low (the magnitude of the variation is small); in another part of the landscape, there may be a longer range to the variogram than we see on average (the spatial scale is different). This approach has been incorporated into the Vesper software package discussed in more detail below (Minasny et al. 2006). However, there is greater scope for managing non-stationarity of the variance within the model-based geostatistical framework discussed in Chap. 11.

10.2.3.3 Contaminated Fields

Consider the spatial variation of the concentration of some metal in the soil of a region. This variation may be complex. It will have its origins in the variable composition of the parent material from which the soil is derived: solid rock, drift, loess, gravels and alluvium of varying mineralogical composition. There may also be different sources of pollution through atmospheric deposition or deposition by flood water. These factors considered so far are likely to give rise to a more or less continuous variation of the concentration of the metal in the soil which we might reasonably regard as a realisation of a random function and describe by a variogram. However, other sources of variation are possible. In particular we consider quasi-

point processes, effects of very intensive contamination which have a very localised effect. We say a quasi-point process because the actual area of soil affected by such an event will be of finite dimensions, but at a practical sampling intensity, such a patch will be represented by a maximum of one sample point. A patch, therefore, appears in our sample as a random event of probability η.

Let us assume for the moment, although it is not necessary, that the continuous background process is a normal random variable S, with mean μ_b and standard deviation σ_b. We may also assume that the contaminant process is normal but typically of larger mean μ_b and with standard deviation σ_c. The overall distribution function is, therefore, $(1-\eta)N(\mu_b, \sigma_b) + \eta N(\mu_c, \sigma_c)$. Matheron's (1962) estimator of the variogram is unduly influenced by extreme values since it is based on squared differences. Also, one contaminated value may appear in several comparisons in each of several lag classes. The contaminant process will therefore have a large effect on the estimate of the variogram in the case of such data. Ideally we should like to decompose our data into background and contaminant components. One way of doing this in spatial analysis is to use robust estimators of the variogram which estimate the variation of the background process.

Estimation of the variogram assumes that our $s(\mathbf{x})$ are realisations of an intrinsically stationary random function $S(\mathbf{x})$. Since $E[S(\mathbf{x}_i) - S(\mathbf{x}_i + \mathbf{h})] = 0$, we can estimate $\gamma(\mathbf{h})$ by $\frac{1}{2}\text{Var}[s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})]$ where Var[] denotes the variances of the term in the brackets. Matheron (1962) uses the standard method of moments estimator, the mean square difference. We may estimate the variogram robustly by using an alternative variance estimator, of which several are available. Lark (2000) reviews some of the principal ones, but we consider one in detail for illustration. The mean-squared error is non-robust because it is an arithmetic average. If just one number in a set of data becomes very large, so does the average. Medians, however, are robust. The median value of a set of data is that value such that 50% of the date are smaller and 50% are larger. If there is an even number of data, $2n$, then the median is the average of the nth and the $n+1$th value. The example in the box below shows how substituting one datum with a very large value can dramatically affect the mean, while the median is only slightly affected or not affected at all. For this reason the median absolute deviation has been proposed as a robust measure of variability as an alternative to the mean square error. The median absolute deviation (MAD) is a robust estimator of the variance based on the median absolute difference between each data value considered in turn and the median of the whole data set. For a set of n values of a variable, $s_i, i = 1, 2, \ldots, n$,

$$\text{MAD} = 2.198 \ \text{median}_{i=1}^{n}\left\{\left|s_i - \text{median}_{i=1}^{n}\{s_i\}\right|\right\}^2 \tag{10.21}$$

where $\text{median}_{i=1}^{n}\{\}$ denotes the median value of the n terms in the brackets. The constant 2.198 requires explanation. It is a consistency correction which, on the assumption that s is a normal variable of variance σ^2, ensures that $E[\text{MAD}] = \sigma^2$. Dowd (1984) proposed using median absolute deviation to estimate variograms. His estimate may be written

$$2\widehat{\gamma}_D(\mathbf{h}) = 2.198\{\text{median}(|y_i(\mathbf{h})|)\}^2, \tag{10.22}$$

where $y_i(\mathbf{h}) = s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h}), i = 1, 2, \ldots, N(\mathbf{h})$. Note that, since we are assuming an intrinsically stationary process which is bivariate normal so that $s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})$ $N\{0, 2\gamma(h)\}$, we have the information that the median value $\text{median}_{i=1}{}^n \{s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})\}$ is zero, which is implicit in this estimator.

Using this robust estimator entails stronger assumptions than does Matheron's estimator, specifically the bivariate normality of $[s(\mathbf{x}_i), s(\mathbf{x}_i + \mathbf{h})]$. The median absolute deviation is also a less efficient estimator than the mean square error in the statistical sense. All the data apart from the median value itself (or the nth and $n + 1$th in the even case) only influence the median absolute deviation by their relative values and so their order. The information contained in the actual values, which the mean square error uses, is not used. This is a penalty of robustness and is one reason why robust estimators should not be used without specific reason. A useful tool to examine variogram models is cross-validation. In cross-validation we use the ordinary kriging method, described below, to compute a prediction of each observation from all the others. Ordinary kriging returns, along with the prediction $\widehat{S}(\mathbf{x}_i)$ at location $\mathbf{x}_i$, the expected squared error of the kriging prediction, $\sigma^2_{\text{OK}}(\mathbf{x}_i)$, which depends on the variogram. Since we know the observed value at each location, we can compute the standardised square prediction error:

$$\text{SSPE}(x_i) = \frac{\left\{s(x_i) - \widehat{S}(x_i)\right\}}{\sigma^2_{\text{OK}}(x_i)}. \tag{10.23}$$

The expected value of this statistic over a set of data is one, and, assuming normal prediction errors, the expected median is 0.455. Lark (2000) suggested that the median standardised squared prediction error from cross-validation is used to select between variogram models fitted to the standard method of moment estimates and alternatives from robust estimators.

10.3 Geostatistical Spatial Prediction

We may sample the soil at discrete locations and analyse the collected material. In practice many problems that the pedometrician has to tackle can be expressed as how to use this information on a very small volume of soil from a region to make predictions about locations or subregions which have not been sampled. In general this problem emerges in two forms. First we may require an estimated value for the soil variable at a site or over a block, the latter being any region, regular or irregular in two or more dimensions such as a field, or a square panel corresponding to the pixel of a remote sensor image. This predicted value may then be used to make some management decision, for example, how much fertiliser to add. Second, we may be more interested in whether the true value of the point or over a block

exceeds some threshold value. This could be a regulatory threshold, for example. In such a case simple comparison of the predicted value with a threshold value is not entirely adequate since the prediction has attendant uncertainty. We need to have some idea of the risk that the true value exceeds or is less than the threshold so that we know how strong is the evidence that the soil at a location should be treated as contaminated (von Steiger et al. 1996) or salt affected (Wood et al. 1990) or deficient in a nutrient (Lark and Ferguson 2004). In mathematical terms we need to know the probability the true value exceeds the threshold value conditional on our observations. Geostatistics has solutions to these problems, and they have been used by soil scientists since the early 1980s. During this time the variants on the basic kriging equations have diversified. We outline here the key methods which soil scientists have found useful.

10.3.1 Kriging Predictions and their Uncertainty

We wish to obtain an estimate of the value of the soil property at an unsampled location $\mathbf{x}_0$. We call the estimate $\widehat{S}(\mathbf{x}_0)$. Later we extend the problem to the estimation of the mean value of S over a block. We have observations $s(\mathbf{x}_i), i = 1, 2, \ldots n$ from which our estimate is to be obtained. We assume that our $s(\mathbf{x}_i)$ are drawn from a realisation of a random function $S(\mathbf{x})$. At any location this has mean $\mu(\mathbf{x}_i)$.

The kriging estimate is based on familiar regression model. We may write

$$\widehat{S}(\mathbf{x}_0) - \mu(\mathbf{x}_0) = \sum_{i=1}^{n_0} \lambda_i \{S(\mathbf{x}_i) - \mu(\mathbf{x}_i)\}, \tag{10.24}$$

where λ_i is a coefficient or kriging weight. This equation can be rearranged giving

$$\widehat{S}(\mathbf{x}_0) = \sum_{i=1}^{n_0} \lambda_i S(\mathbf{x}_i) + \mu(\mathbf{x}_0) - \sum_{i=1}^{n_0} \lambda_i \mu(\mathbf{x}_i). \tag{10.25}$$

How we proceed from here depends on assumptions which we choose to make about the behaviour of $\mu(\mathbf{x}_i)$.

10.3.2 Ordinary Kriging (OK)

Ordinary kriging is the most widely used kriging method. It proceeds on the assumption that $\mu(\mathbf{x}_i)$ is constant at least for all n_0 sample sites at $\mathbf{x}_i$ within the vicinity of $\mathbf{x}_0$ which we use to estimate $s(\mathbf{x}_0)$. It is also assumed that this mean is unknown to us.

On the basis of these assumptions, we may reduce Eq. 10.25 to

$$\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0) = \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} S(\mathbf{x}_i) + \mu(\boldsymbol{x}_0)\left(1 - \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}}\right). \tag{10.26}$$

Now if we specify that $\sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} = 1$, then the unknown mean is filtered from the ordinary kriging estimator and $\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0) = \sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} S(\mathbf{x}_i)$. This means that $\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0)$ is an unbiased estimate of $\mu(\mathbf{x}_0)$, i.e. on average $\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0)$ is equal to the local mean, under the assumption of a stationary mean within the neighbourhood. We may also define a quantity $\sigma^2\mathrm{OK}(\mathbf{x}_0) = E\left[\left\{\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0) - s(\mathbf{x}_0)\right\}^2\right]$. This is the ordinary kriging variance, the mean-squared error of the kriging estimate. We find the weights λ_i by solving a set of linear equations which minimise this value. The OK estimate is an optimal estimate in the least square sense; that is to say it is a best linear unbiased estimator, given the assumption of a locally constant (but unknown) mean.

The kriging equations are essentially obtained by writing expression for the kriging variance in terms of the covariances of $S(\mathbf{x}_0)$ and all the $S(\mathbf{x}_i)$ and the covariance among all the $S(\mathbf{x}_i)$. These depend on the covariance function $C(\mathbf{h})$ defined in Eq. 10.1. We then write partial derivatives of the kriging variance with respect to the weights λ_i, which are zero where the kriging variance is minimised. See Isaaks and Srivastava (1989) for further details. At first glance the problem may appear simple. We have obtained n_0 unknowns, the weights, λ_i and obtained n_0 equations from setting the n_0 partial derivatives to zero However, this ignores an additional equation, namely, the constraint used to filter out the unknown mean which requires that all the weights sum to one. This $n_0 + 1$th equation requires that we add an additional unknown the Lagrange parameter, $\psi(\mathbf{x}_0)$, which means that the OK equations actually minimise an auxiliary function ${\sigma^2}_{\mathrm{OK}}(\mathbf{x}_0) - 2\psi(\mathbf{x}_0)\left\{\sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} - 1\right\}$. This generates the ordinary kriging equations:

$$\begin{array}{l} \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} C\left(\mathbf{x}_i - \mathbf{x}_j\right) + \psi(\mathbf{x}_0) = C\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} = 1. \end{array} \tag{10.27}$$

At this point the reader might reasonably wonder why we have discussed ordinary kriging in terms of covariance functions. Why the previous focus on the variogram? But remember that $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$. This allows us to rewrite the OK equations as

$$\begin{array}{l} \sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} \left\{C(0) - \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right)\right\} + \psi(\mathbf{x}_0) = C(0) - \gamma\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} = 1. \end{array} \tag{10.28}$$

The unbiasedness constraint causes the term $C(0)$ to drop out, so

$$\begin{array}{l} \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right) + \psi\left(\mathbf{x}_0\right) = \gamma\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} = 1. \end{array} . \tag{10.29}$$

We still seem to depend on the assumption of weak stationarity since it is assumed that the covariance function can be defined. In fact OK can be conducted for intrinsically stationary random variables by substituting for $C(0)$ some arbitrary large value C_A which is filtered out. The expression C_A -$C(\mathbf{h})$ is called the pseudo-covariogram. Solving Eq. 10.29 above for the weights allows us to generate the kriging estimator with kriging variance

$$\sigma^2{}_{\mathrm{OK}}\left(\mathbf{x}_0\right) = \sum_{i=1}^{n_0} \lambda_i{}^{\mathrm{OK}} \gamma\left(\mathbf{x}_i - \mathbf{x}_0\right) + \psi\left(\mathbf{x}_0\right). \tag{10.30}$$

OK can be extended to the estimation of a regional mean, i.e. the average value of s over some block R. This can be conveniently expressed as an integral

$$s(R) = \frac{1}{R} \int_{x \in R} s(x) \mathrm{d}x.$$

The interpretation of $s(R)$ may require some care. It is the population mean that we would estimate by design-based sampling of R, measuring s at randomly selected locations in R. For some properties we can think of $s(R)$ as the value we would obtain if all the soil in R were taken and homogenised and then analysed for properties in which the arithmetic average values of a set of discrete samples is equivalent to the aggregate property of the sample. This includes compositional properties such as the clay content or volumetric water content, but not scale-dependent physical properties such as the saturated hydraulic conductivity or properties such as the soil solution concentration of an element which may depend on variable and non-linear exchange processes.

Our block estimate $\widehat{S}(R)$ is found as a linear combination of the data $s(\mathbf{x}_i)$. The weights are found by solving a similar set of equations to those for point estimates.

$$\begin{array}{l} \sum_{i=1}^{n_0} \lambda_{R,i}^{\mathrm{OK}} \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right) - \psi(R) = \overline{\gamma}\left(\mathbf{x}_j - R\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} \lambda_{R,i}^{\mathrm{OK}} = 1. \end{array} . \tag{10.31}$$

These equations are solved for the weights. The block kriging variance is

$$\sigma^2{}_{\mathrm{OK}}(R) = \sum_{i=1}^{n_R} \lambda_{R,i}^{\mathrm{OK}} \overline{\gamma}\left(\mathbf{x}_i, R\right) + \psi\left(R_0\right) - \overline{\gamma}\left(R, R\right). \tag{10.32}$$

The two components $\overline{\gamma}(\mathbf{x}_i, R)$ and $\overline{\gamma}(R, R)$ require explanation. The former represents the mean value of the variogram between $\mathbf{x}_i$ and some point in R. The latter is the mean value of the variogram between points in R. This latter quantity is called the dispersion variance or within-block variance. It is equal to the mean-squared difference form the mean of the values of s drawn at random from within R. These two values can be calculated for regular blocks using specific functions, the auxiliary functions (see Journel and Huijbregts 1978). However, with modern computers they are usually obtained by numerical integration which can also be done efficiently for regions of any shape or size.

10.3.3 Simple Kriging (SK)

We assume that the mean is everywhere constant and known. The kriging weights are calculated to minimise the prediction variance subject to an unbiasedness constraint that $E\left[\widehat{S}(\mathbf{x}_0) - s(\mathbf{x}_0)\right] = 0$. From this we obtain n_0 equations

$$\sum_{i=1}^{n_0} \lambda_i^{\mathrm{SK}} C\left(\mathbf{x}_i - \mathbf{x}_j\right) = C\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j. \tag{10.33}$$

In SK the value of $C(0)$ is not filtered out in the kriging equation. This means that the variogram function must be bounded, i.e. the random function is assumed to be second-order stationary. This constraint, and the condition that the mean is known, makes SK generally less attractive than OK, and it is much less widely used in soil science. SK is most commonly used in special circumstances where the mean is predetermined. We shall touch on some of these shortly.

10.3.4 Non-linear Kriging

Ordinary kriging and simple kriging, as discussed above, are linear methods which return an estimate of a soil variable at an unsampled site or over a block, the estimate being a linear combination of the data. Information about the soil for a point or block is often needed to make decisions about its management. This management may involve an intervention, for example, the addition of a fertiliser or other amendments to the soil or the removal or remediation of contaminated soil. This intervention entails a cost, but if it is not undertaken where it is actually needed, then a further cost may be incurred through loss of yield or environmental damage which may result in a fine. Regulations or rules are commonly stated in terms of threshold values s_t. These lead to rules:

If $s(R) > s_t$, then action (A), else ($s(R) \leq s_t$) action (B).

If we follow the rule, substituting our estimate $\widehat{S}(R)$ for $s(R)$ above, then errors in the estimate will inevitably result in some incorrect decisions. Should the uncertainty in $\widehat{S}(R)$ affect the decision-making rule above? This depends on the distribution of the estimation errors and the shape of the loss function. By the latter we mean the additional cost entailed as a result of an error (positive or negative) in $\widehat{S}(R)$ when the rule above is applied, the loss being expressed as a function of the error. If the loss function is symmetrical (the cost of an overestimate of s by x units is equal to the cost of an underestimate by the same amount) and the distribution of the errors is symmetrical (e.g. normal), then the best procedure is to follow the rule with the estimate $\widehat{S}(R)$. The only way to reduce the costs incurred as a result of error is to reduce the error variance of $\widehat{S}(R)$ (e.g. by sampling more intensively).

In practice, however, loss functions are usually asymmetric. The fine for leaving a region of land unremediated where remediation was, in fact, the correct decision may exceed the cost of remediation. The yield loss on under-fertilising a region of a field would generally exceed the cost of overfertilising. In these conditions the correct decision in the presence of uncertainty about the true value of $s(R)$ requires that we can quantify the uncertainty, conditional on the observations of the variable that we do have.

There are two general groups of kriging techniques that have come to prominence for tackling this problem, both are non-linear – a conventional linear kriging estimator is applied to the data after these have been transformed non-linearly. The first group are indicator methods (Journel 1983) of which the basic tool is ordinary indicator kriging (IK). Here the non-linear transform is to a binary indicator variable. These techniques have been widely applied by soil scientists (e.g. Meirvenne and Goovaerts 2001; Halvorson et al. 1995). The second group of techniques involve a non-linear transformation of the data to a continuous variable, usually a normally distributed variable. This method is exemplified by disjunctive kriging (Matheron 1976), but multi-Gaussian kriging is a similar method. These techniques have also found widespread use in soil science (e.g. Wood et al. 1990; von Steiger et al. 1996).

Indicator kriging (IK) As previously soil property s at location $\mathbf{x}$ takes value $s(\mathbf{x})$. An indicator transform of $s(\mathbf{x})$ can be defined by

$$\omega_t(\mathbf{x}) = \begin{array}{l} 1 \text{ if } s(\mathbf{x}) \leq s_t \\ 0 \quad \text{otherwise,} \end{array} \tag{10.34}$$

where s_t is a threshold value of the property such as one of the management thresholds mentioned above. In indicator geostatistics, $\omega_t(\mathbf{x})$ is regarded as realisation of the random function $\Omega_t(\mathbf{x})$. It can be seen that

$$\text{Prob}\left[S(\mathbf{x}) \leq s_t\right] = E\left[\Omega\left\{S(\mathbf{x}), s_t\right\}\right] \tag{10.35}$$

where Prob[] denotes, respectively, the probability of the event within the square brackets and $G\{S(\mathbf{x}), s_t\}$ is the cumulative distribution function of $S(\mathbf{x})$ at the threshold value s_t. In indicator kriging we estimate the conditional probability that

$s(\mathbf{x})$ is smaller than or equal to the threshold value s_t, conditional on a set of observations of s at neighbouring sites, by kriging $\omega_t(\mathbf{x})$ from a set of indicator-transformed data.

A set of data on s is transformed to the indicator variable $\omega_t(\mathbf{x})$ using Eq. 10.34 above. The variogram of the underlying random function $\Omega_t(\mathbf{x})$ is then estimated and modelled in the usual way. An estimate of the indicator random function may then be obtained for a location $\mathbf{x}_0$ by kriging from the neighbouring indicator transform data. Ordinary indicator kriging is equivalent to simple kriging of the indicator variable $\Omega_t(\mathbf{x})$ using the mean within the kriging neighbourhood as the expectation. Goovaerts (1997) and Webster and Oliver (2007) give more details. The IK estimate $\Omega_t(\mathbf{x})$ is an estimate of the conditional probability that $s(\mathbf{x}) \leq s_t$ and of the conditional cumulative distribution function (ccdf) $G(S(\mathbf{x}), s_t)$. This direct estimate $\widehat{\Omega}_t(\mathbf{x})$ is not generally used as the conditional probability. Instead it is recommended to obtain estimates $\widehat{\Omega}_t(\mathbf{x})$for several s_t that include the threshold of practical importance. There is no guarantee that the estimates $\widehat{\Omega}_t(\mathbf{x})$ will meet the order relation constraint for a cumulative distribution function, that is, $\widehat{\Omega}_{t1}(\mathbf{x}) \leq \widehat{\Omega}_{t2}(\mathbf{x}) \leq \widehat{\Omega}_{t2}(\mathbf{x})$ for any $s_{t1} < s_{t2} < s_{t3}$. Because of this the original set of estimates $\widehat{\Omega}_{t1}(\mathbf{x}), \widehat{\Omega}_{t2}(\mathbf{x}) \widehat{\Omega}_{t2}(\mathbf{x})$ must be smoothed to give a set of revised estimates $\tilde{\Omega}_{t1}(\mathbf{x}), \tilde{\Omega}_{t2}(\mathbf{x}) \tilde{\Omega}_{t2}(\mathbf{x})$ such that the order relations hold. Deutsch and Journel (1998) describe methods for doing this. The estimate of the conditional probability that $s(\mathbf{x}) \leq s_t$ is now given by $\tilde{\Omega}_t(\mathbf{x})$. It is likely that $\tilde{\Omega}_t(\mathbf{x})$ is a better estimate of the conditional probability for a cut-off than is $\widehat{\Omega}_t(\mathbf{x})$ since it incorporates information from adjacent parts of the ccdf. Variants of ordinary indicator kriging are available. In particular we can cokrige the indicator variables for the different thresholds (see the later discussion of cokriging) although Goovaerts (1994) found no substantial benefit from doing so.

Disjunctive kriging (DK) Disjunctive kriging is an alternative to indicator kriging. It is based on the assumption that our data are a realisation of a process with a second-order stationary distribution. Further we assume that the underlying process is a diffusion process (Rivoirard 1994; Webster and Oliver 2007). That is, it varies continuously so that if the variable takes values s_1 and s_2 at locations $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, then all intervening values between s_1 and s_2 must occur at locations on a straight line between $\mathbf{x}_1$ and $\mathbf{x}_2$. The commonest model which we use for DK is the normal diffusion process. Webster and Oliver (2007) explain how the plausibility of this assumption is tested. Since data may often not resemble a normal random variable, the first step in DK is to apply a non-linear transform to the data to achieve normality. This is done using hermite polynomials, a procedure described in more detail by Rivoirard (1994). A variogram is estimated for the hermite transform data in the usual way and modelled. The hermite polynomials are then kriged to target locations of interest. From these may then be estimated the original soil variable $s(\mathbf{x})$ and the conditional probability for specified thresholds, $\widehat{\Omega}_t^{\mathrm{DK}}(\mathbf{x})$.

The relative merits of IK and DK may be summarised as follows. IK may be implemented using widely available kriging software. Since indicators may be estimated by simple kriging or ordinary kriging, we need not assume that the mean of the indicator variable is known. The DK on the other hand is restricted to simple

kriging. The indicator transform of the continuous soil variable discards a good deal of information, and it is tedious to estimate model indicator variograms for a large number of threshold values. The fact that IK estimates of the ccdf do not necessarily conform to the order relations and generally require an arbitrary correction to meet this constraint is a further disadvantage.

DK requires only one variogram model for the hermite transform data. Indicator variograms and cross-variograms must also be estimated to test the assumption of a diffusion process, but these need not be modelled. The hermite transformation transforms a continuous variable monotonically to another continuous variable, so, unlike the indicator transform, there is no loss of information. Further DK returns an estimate of $s(\mathbf{x})$ along with the conditional probabilities. An estimate of $s(\mathbf{x})$ is not generated by IK without additional effort. On the other hand, DK requires stricter assumptions of stationarity than does IK. It can only be implemented if the hermite transform data have a bounded variogram. DK is a complex procedure, and friendly software is not widely available, although the related procedure of multi-normal kriging is available in the GSLIB package (Deutsch and Journel 1998).

In practice we make the choice between IK and DK on practical considerations such as some of those above. In general we might expect that DK or other non-linear kriging techniques that use a continuous transform of the original data will perform better than IK since the indicator transform will inevitably lose some information, while a continuous transform retains all the information in the original data. There have been few studies to compare IK and DK on real data. Those of which we are aware (Papritz and Dubois 1999; Lark and Ferguson 2004) suggest that there may be very little difference between the results obtained with the two methods.

10.3.5 Kriging with a Nonstationary Mean

Again we return to our basic model of soil variations $s(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x})$. So far we have considered $f(\mathbf{x})$ to be a constant value, the mean. In practice it may be necessary for $f(\mathbf{x})$ to express sources of variation in the soil which are not constant and cannot be regarded as any kind of random variable. It is not difficult to think of examples where the variation of a soil property is systematically linked to location and space. Consider the familiar figures in pedological textbooks of the depth of weathered material and how it changes from a crest down a slope to the foot slope and toe slope. Consider again the systematic variation of soil texture, mineralogy, redox potential and organic carbon content associated with the familiar catenas of Central Africa. In all these cases, $f(\mathbf{x})$ must express some deterministic relationship between location in space and a soil variable.

Generally we recognise a distinction between two kinds of systematic variation although it is more a distinction of degree and spatial scale than a fundamental difference. The first kind of systematic variation includes long-range trends where the component represented by $f(\mathbf{x})$ is a broad variation from large to small values of a variable between different parts of a landscape. The second kind of systematic

variation is where short-range differences are dominated by what appear to be deterministic processes. This results in the parabolic form of the experimental variogram that we discussed above. It should be noted that variation that we treat as a trend within field scale might be reasonably attributed to random variation if we are sampling a catchment. This underlines the fact that our model of variation is in part a consequence of the scale and intensity of our sampling.

Various approaches have been taken to the problem of kriging in the presence of a nonstationary mean; these include universal kriging, regression kriging and the method of intrinsic random functions of order k. In universal kriging (UK) (e.g. Webster and Burgess 1980), we express $f(\mathbf{x})$ as some function linear in polynomials of the elements of $\mathbf{x}$ (i.e. trend surfaces), and the kriging estimator implicitly estimates the coefficients of these terms along with a minimum variance estimate of the random term $\varepsilon(\mathbf{x})$. This requires that we have a variogram of $\varepsilon(\mathbf{x})$. Regression kriging (RK) is formally equivalent to UK, but we start by finding an ordinary least squares estimate of the coefficients of the trend function and then estimate the variogram from the residuals (e.g. Odeh et al. 1994). We may then use the variogram to compute a weighted least squares estimate of the coefficients and iterate this process to convergence. The RK prediction of $s(\mathbf{x}_0)$ is then computed by first calculating the value of the trend at $\mathbf{x}_0$ and then estimating the random component by simple kriging (with a known mean of zero) from the residuals from the trend surface at the observation sites. RK and UK are formally equivalent, given the variogram of $\varepsilon(\mathbf{x})$, but it is this variogram that is the problem for both methods. We saw above that the experimental variogram of a variable with a trend will be affected by this trend, and what we need is a variogram of the random component only. This might be obtained for UK by only estimating the variogram from pair comparisons that are not strongly affected by the trend (e.g. over lags perpendicular to the direction of the trend), but this is not always possible, and neither is it efficient. The variogram obtained from residuals from a trend surface is biased (Cressie 1993), so the RK solution is not satisfactory either, although it is reasonably robust when applied to large data sets.

Intrinsic random functions of order k (IRFk) are generalised increments of our data which filter out the trend. The concept is easily grasped if we consider a linear trend in one dimension. It is clear that the differences between adjacent and regularly sampled observations of a variable with a linear trend and an intrinsically stationary random function will be intrinsically stationary since the trend component is replaced by its first derivative. If the trend is of higher order, then further differencing will remove it (and the lower-order components become zero). The method of IRFk works in this way, and a key stage is the determination of the order of increments that can be regarded as intrinsically stationary. This is done by examining generalised covariance functions, in effect variograms of differenced data. Chiles and Delfiner (1999) provide more detail, and Buttafuoco and Castrignano (2005) give an example of the application of this approach in soil science.

In RK and UK we use an unbiased linear model of the trend for prediction in combination with an unbiased minimum variance estimate of the random component of variation. UK does this simultaneously and RK separates the two stages. Given a variogram for the random component, the predictors are the best linear unbiased

predictor (BLUP) or empirical BLUP (E-BLUP), empirical in that it is conditional on a variance model derived from data. We noted above that the key problem is obtaining this variance model, a variogram of $\varepsilon(\mathbf{x})$. The state-of-the-art approach to this problem is to use residual maximum likelihood (REML), and it is the REML–E-BLUP that we advocate for spatial prediction in the presence of a trend. This, however, is out with the scope of the current chapter and is discussed in Chap. 11.

10.3.6 Sampling for Estimation by Kriging

McBratney et al. (1981) and Burgess et al. (1981) showed how a geostatistical survey can be designed to ensure that the variable may be estimated with a prespecified precision. They considered regular grids and assumed that the maximum kriging variance occurred at the centre of each regular grid cell. Thus, they ignored the increased kriging variances close to the boundaries of the region. They sought the maximum interval between observations on the regular grid such that the kriging variance was less than the prespecified tolerance, σ^2_{T}. Thus, for variogram parameter vector $\boldsymbol{\theta}$, they calculated the optimal interval, $I(\boldsymbol{\theta})$, such that

$$I(\boldsymbol{\theta}) = \max\left[i, \text{such that } \max\left\{\sigma_K^2\left[\boldsymbol{\theta}; \mathbf{x}_g(i)\right]\right\} < \sigma_T^2\right], \tag{10.36}$$

where ${\sigma^2}_K|\boldsymbol{\theta}, \mathbf{x}_g(i)$ denotes the kriging variance at the centre of a cell within a square grid of interval i. The same approach can be followed when using cokriging (McBratney and Webster 1983) and when planning composite sampling (Webster and Burgess 1984).

It is possible (Marchant and Lark 2007) to write an expression for the OK prediction error variance that accounts both for the distance to neighbouring observations and the error in the estimated variogram parameters. It is possible to minimise the average value of this statistic across a proposed sample region by modifying the location of sample points. Brus and Heuvelink (2007) show how the approach can be extended to universal kriging.

The problem here is that in most cases, the variogram is unknown when sampling is designed. We may use approximate variograms, as discussed in Chap. 10. An alternative is the adaptive approach proposed by Marchant and Lark (2006). This is appropriate for circumstances where the sampling can be done in phases (ideally when the data can be collected in real time with a sensor). For example, in their Bayesian adaptive approach, spatial simulated annealing is used to minimise an objective function which is the mean square error of the kriging variance at the centre of a notional grid cell. This objective function is obtained as a Bayesian integration over the space of possible variogram parameters with a probability density function for these parameters. The sampling proceeds in phases, and the pdf of the variogram parameters is updated as new data are collected. At the end of each sampling phase, we consider the cumulative distribution function of grid spacings which are sufficient to ensure some maximum kriging variance. When the uncertainty on this interval is sufficiently small, the sampling is completed

by finding those locations needed to complete a sample grid of adequate intensity across the study region.

10.3.7 Kriging for Large Data Sets

Sensors such as gamma-ray spectrometer and electromagnetic induction instruments attached to a vehicle with a GPS have been used to collect intensive soil data for high-resolution mapping. These on-the-go proximal soil sensors can collect a large volume of data over an area (approximately 5,000–100,000 data points per km^2). Data collected from these sensors need to be interpolated to a regular grid.

Kriging of such data poses several shortcomings: the time taken to calculate an empirical variogram can be excessive; in addition, solving the kriging equations for a data set of size n involves inversion of an $(n \times n)$ covariance matrix, which requires $O(n^3)$ operation. A way to circumvent this is to use the spherical model (a finite range variogram), where a sparse matrix can be approximated for the variance–covariance matrix (Barry and Kelley Pace 1997). Others called for covariance matrix tapering, where covariances at large distances are set to zero (Kaufman et al. 2008). Kriging using a single variogram model for the whole area usually resulted in a smooth map, where local variation captured by data can be lost.

Another solution is kriging that takes into account the local spatial structure (Haas 1990). This is implemented as kriging with local variograms, also known as kriging and automated variogram modelling within a moving window. It involves searching for the closest neighbourhood for each prediction site, estimating the empirical variogram from the neighbourhood, fitting a variogram model to the data automatically by a non-linear least squares approach, kriging with local neighbourhood and variogram parameters and calculating the uncertainty of kriging prediction. All these steps need to be done automatically, and thus the program adapts itself spatially in the presence of distinct differences in local structure over the whole field. Local variogram estimation and kriging can preserve the local spatial variation in the predictions. In most cases, local variograms could circumvent the problems of anisotropy and the need for trend analysis.

Minasny et al. (2006) developed a program called Vesper (Variogram Estimation and Spatial Prediction plus Error), a PC-Windows software program that can calculate and model global local variograms and do global and local kriging in either punctual or block form. Sun et al. (2012) extended this approach to local regression kriging to take into account both the local relationship between the covariates and soil observations and the spatial variance of the residuals.

Cressie and Kang (2010) regarded this local kriging method as an ad hoc solution and suggested the fixed ranked kriging (FRK) approach. FRK uses covariance functions that are flexible through a set of r basis functions, where the $(n \times n)$ variance–covariance matrix can be approximated by $(r \times r)$ positive-definite matrices (Cressie and Johannesson 2008). Cressie and Kang (2010) demonstrated the application of FRK on proximally-sensed gamma counts ($n = 34{,}266$) in a field of 2.66 km^2. They used a flexible, nonstationary spatial covariances represented as

77 basis functions, for which exact kriging can be carried out. Nevertheless, the FRK approach still needs tuning with respect to the type of basis functions and estimation of the parameters of those functions.

10.4 A Case Study

We now illustrate some of the key concepts introduced in previous sections with a case study. This uses data on the concentration of copper in the topsoil of part of the east of Scotland. The data were first described by McBratney et al. (1982). Figure 10.5 shows histograms of the raw data and of the data transformed to natural logarithms, and Table 10.1 presents summary statistics of these two variables.

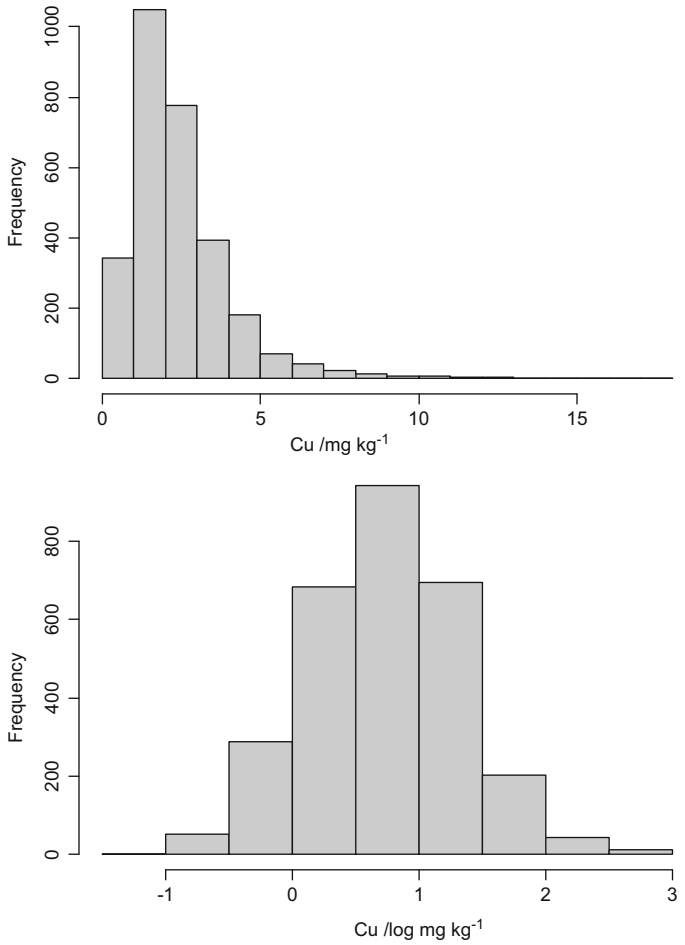


Fig. 10.5 Histograms of topsoil copper content in soils from the east of Scotland on (*top*) original and (*bottom*) natural log scales

Table 10.1 Summary statistics of data on topsoil copper content from the east of Scotland

Variable	Mean	Median	Standard deviation	Minimum	Maximum	Skewness
Copper/mg kg^{-1}	2.48	2.10	1.68	0.3	18	2.66
Copper/$\log_e$ mg kg^{-1}	0.73	0.74	0.59	−1.2	2.89	0.09

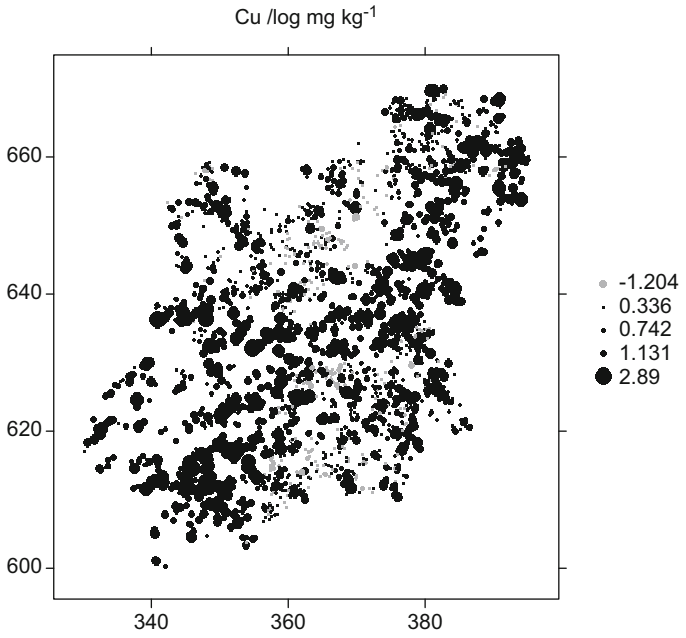


Fig. 10.6 Quintiles of soil copper content (legend is on log scale) shown on a post-plot of the copper data. Coordinates are kilometres relative to the origin of the British National Grid

The histograms and the summary statistics show that the data appear symmetrically distributed on a log scale, so this is used for further analysis. Figure 10.6 shows the distribution of the observations.

Figure 10.7 shows the empirical variogram of the log-transformed data, estimated for four different directions. There is no evidence of systematic anisotropy, so isotropic variograms were estimated and modelled. Matheron's estimator (Eq. 10.4) and the robust estimator of Dowd, described in Eq. 10.22, were both used, and the resulting empirical variograms were fitted with double spherical variogram models, which were preferred to simpler ones on the grounds of Akaike information criterion. Both models were cross-validated, and the standardised squared prediction error, Eq. 10.23, was computed. The estimates, models and values of *SSPE* (see Eq. 10.23) are shown in Fig. 10.8. On the basis of the cross-validation, the model fitted to estimates from Dowd's estimator was chosen. Figure 10.9 shows the ordinary kriging estimates of soil copper content (log scale) across the study area and the kriging variance. As expected the variance is largest where the data are most sparse.

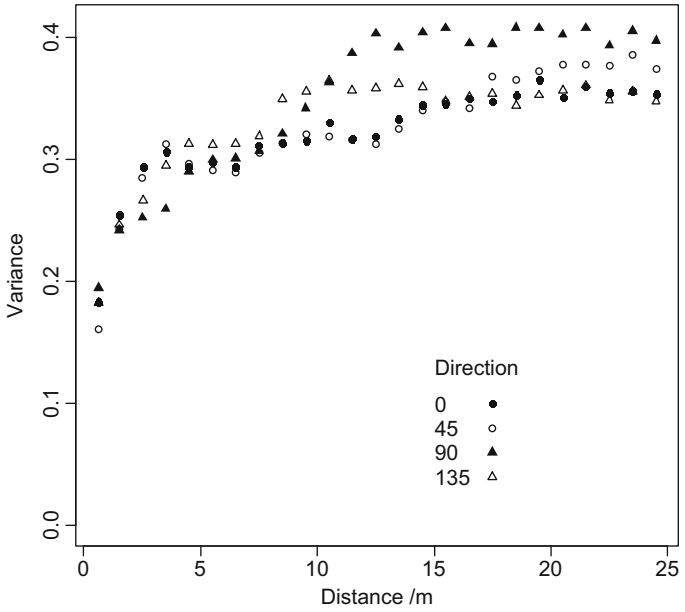


Fig. 10.7 Directional empirical variograms of log copper content

Figure 10.10 shows the ordinary kriging variance for (log) soil copper content, computed at the centre of a cell of a notional square sampling grid with spacing ranging from 500 m to 5 km. This graph, following McBratney et al. (1981), allows one to identify the grid spacing which would be required in a survey in comparable conditions if one wanted to achieve a kriging variance no larger than 0.2. The graph shows that a grid spacing of no more than 1.5 km is required to achieve this target.

10.5 Spatial Covariation and Coprediction

The methods we have discussed so far are all univariate. They consider just one variable and its variation in space. The geostatistical model of spatial variation can be readily extended to two or more variables. Why should we wish to do this? First, because there are cases where we are interested both in estimates of variables and in estimates of some linear function of the variables. An example from soil science quoted by Webster and Oliver (2007) is where one variable, $s_1(\mathbf{x})$, is the depth to the top of a particular soil horizon and the second, $s_2(\mathbf{x})$, is the depth to the bottom of the horizon. A linear combination of these variables is their difference, the thickness of the horizon. If $\widehat{S}_1(\mathbf{x})$ and $\widehat{S}_2(\mathbf{x})$ are estimates of the two variables at $\mathbf{x}$, then one estimate of the horizon thickness at $\mathbf{x}$ is $\widehat{S}_1(\mathbf{x}) - \widehat{S}_2(\mathbf{x})$. However, we could also krige the difference variables $\{s_1(\mathbf{x}) - s_2(\mathbf{x})\}$ directly from the differences at

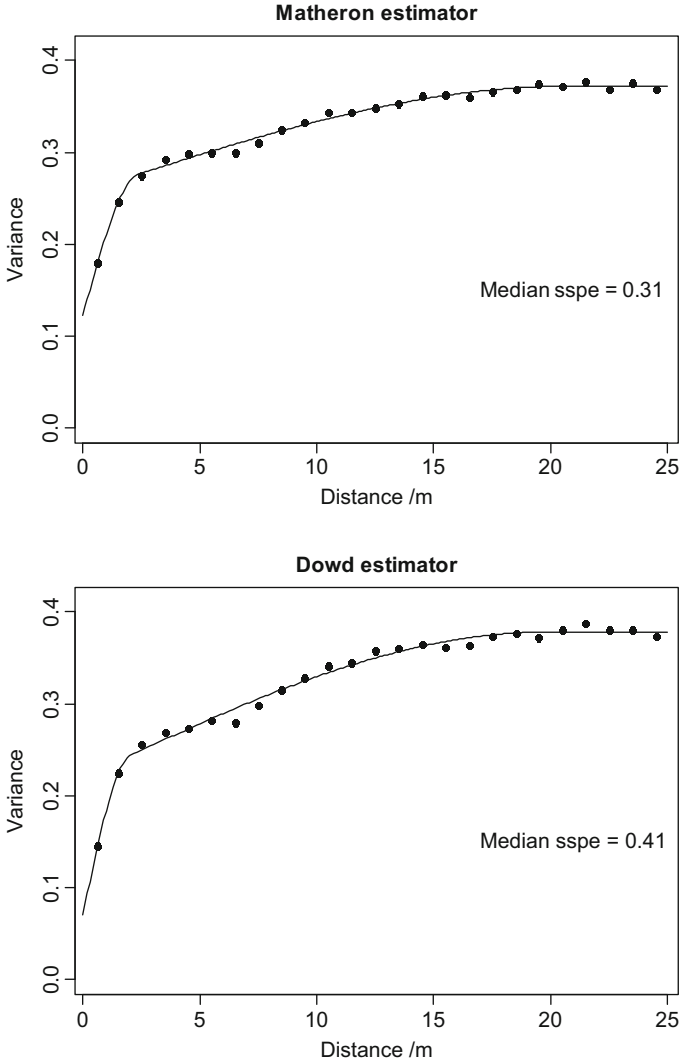


Fig. 10.8 Estimated isotropic variograms of soil copper content with fitted double spherical models and median standardised squared prediction error using Matheron's or Dowd's variogram estimators

our observation sites. There is no guarantee that the two approaches will yield the same estimate if we use the univariate kriging equations described above. That is to say, the kriging estimate of $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$ is not guaranteed to be equivalent to the kriging estimate of any linear combination of $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$. The kriging estimates are not coherent. For some applications this may be important.

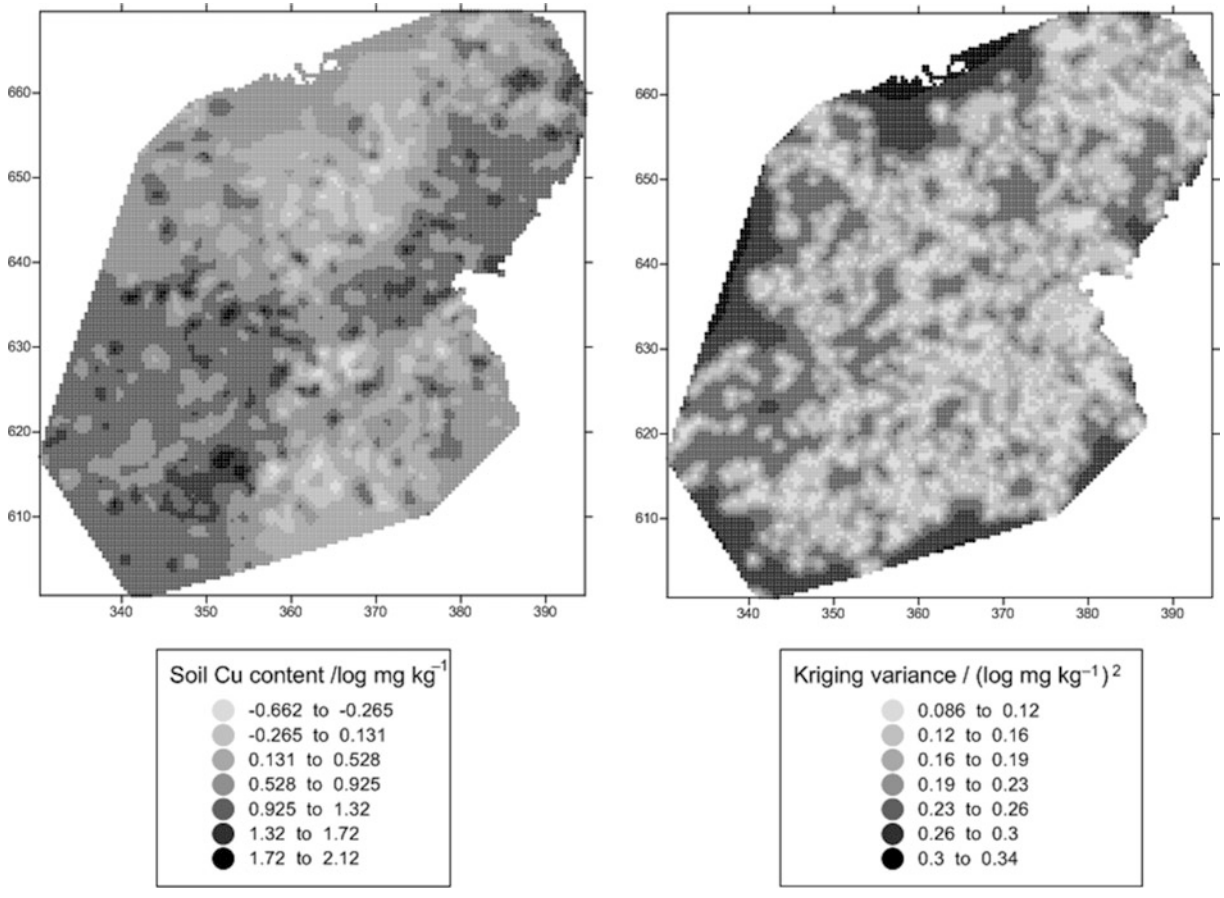


Fig. 10.9 Ordinary kriging prediction of topsoil copper content (log scale) and ordinary kriging variance across the east of Scotland

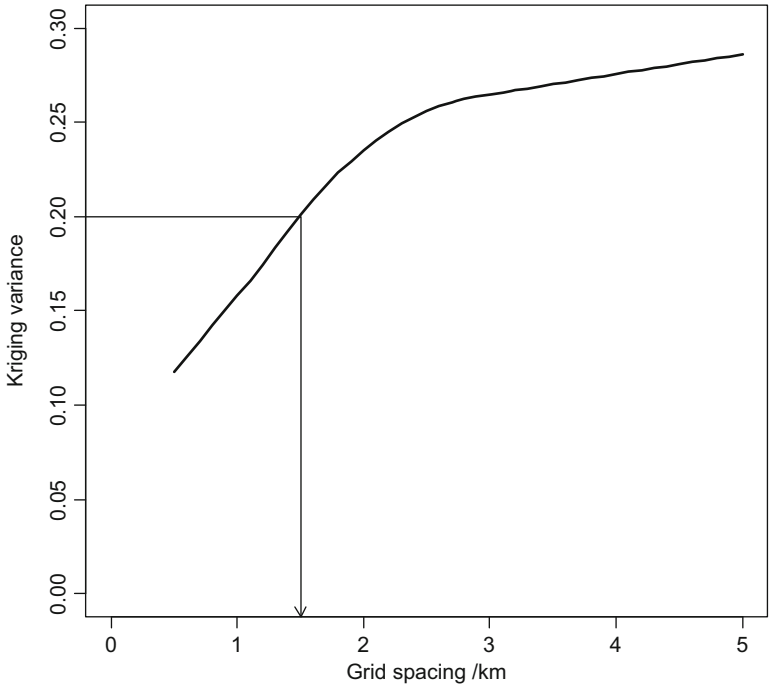


Fig. 10.10 Ordinary kriging variance at the centre of a cell of square grids of different spacing, based on the variogram for topsoil copper content in the east of Scotland

The second reason for considering multivariate statistical methods is probably the commonest. In many circumstances it may be possible to supplement a set of relatively costly direct measurements of a soil variable with a denser but cheaper set of measurements of a second correlated variable. For example, remote sensor measurements of the earth's surface may be correlated with a soil variable of interest. If this can be exploited through a geostatistical method, then we may be able to obtain better predictions of the soil variable by incorporating the cheapest second variable, without added costly measures.

10.5.1 Spatial Co-regionalisation

Let $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$ denote measurements of two soil variables at location $\mathbf{x}$. We assume that these are realisations of, respectively, random functions $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$. It is assumed that these are both intrinsically stationary, and there exists a cross-variogram $\gamma_{2,1}(\mathbf{h})$ dependent only on $\mathbf{h}$ where

$$\gamma_{2,1}(\mathbf{h}) = E\left[\{S_1(\mathbf{x}_i) - S_1(\mathbf{x}_i + \mathbf{h})\}\{S_2(\mathbf{x}_i) - S_2(\mathbf{x}_i + \mathbf{h})\}\right]. \tag{10.37}$$

The cross-variogram can take negative values unlike the auto-variograms $\gamma_{1,1}(\mathbf{h})$ and $\gamma_{2,2}(\mathbf{h})$ which are positive by definition. In fact the ratio

$$\frac{\gamma_{2,1}(\mathbf{h})}{\sqrt{\gamma_{1,1}(\mathbf{h})\,\gamma_{2,2}(\mathbf{h})}} \tag{10.38}$$

is known as the co-dispersion coefficient and measures the correlation of variables $S_1(\mathbf{x})$ and $S_2(\mathbf{x})$ at lag $\mathbf{h}$ which may be positive (large values of $S_1(\mathbf{x})$ are associated with large values of $S_2(\mathbf{x})$) or negative (small values of $S_1(\mathbf{x})$ are associated with large values of $S_2(\mathbf{x})$).

The cross-variogram may be estimated in much the same way as the ordinary variogram for a single variogram, sometimes called the auto-variogram. If we define a lag class centred at scalar lag h with N_h pair differences, then an estimate of the cross-variogram for the lag class is

$$\widehat{\gamma}_{2,1}(h) = \frac{1}{2}\sum_{i=1}^{N_h}\{s_2(\mathbf{x}_i) - s_2(\boldsymbol{x}_i + h)\}\{s_1(\mathbf{x}_i) - s_1(\mathbf{x}_i + h)\}. \tag{10.39}$$

This estimator is non-robust, like the comparable auto-variogram estimator we have already discussed (of which it is a generalisation). Lark (2003) has proposed and demonstrated robust estimators of the cross-variogram which should be used with either s_1 or s_2 or both s_1 and s_2 containing outlying values.

It is notable that the cross-variogram estimator requires that we have paired comparisons of both variables made at the same location. There are circumstances in which few or none of our measurements on different variables are made at the same location. For example, Papritz and Webster (1995) point out that when monitoring change of soil properties over time (with s_1 equal to the variable at time 1 and s_2 the variable at time 2), there can be such problems since at the limit, we cannot measure the same soil twice by destructive sampling and analysis. In these circumstances we may estimate the pseudo-cross-variogram introduced by Clark et al. (1989) defined by Myers (1991) as

$$\widehat{\gamma}^{P}_{2,1}(h) = \mathrm{Var}\left[s_2(\mathbf{x}_i) - s_1(\mathbf{x}_i + h)\right], \tag{10.40}$$

where Var[] denotes the variance of the term in brackets. While the pseudo-cross-variogram allows us to extract information on spatial covariation where variables are not measured at the same locations, it does require more restrictive assumptions, and in practice these amount to the assumption of weak rather than only intrinsic stationarity.

In order to use the cross-variogram, we must model it along with the corresponding auto-variograms. In the multivariate case, it is necessary but not sufficient to use authorised variogram functions; since a model co-regionalisation comprising authorised functions fitted separately to the cross, the auto-variograms is not necessarily positive definite. The most widely recommended strategy to ensure a positive-definite model of co-regionalisation is to use a linear model of co-

regionalisation (LMCR). In the LMCR it is assumed that all the constituent random functions are linear combinations of a common set of independent random functions of mean zero and unit variance, $Y_k^l(\mathbf{x})$ (l is an index not a power). Thus

$$S_u(\mathbf{x}) = \mu_u + \sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{u,k}^l Y_k^l(\mathbf{x}), \tag{10.41}$$

where μ_u is the mean of the random function. The coefficients $a_{u,k}^l$ are specific to the random functions $S_u(\mathbf{x})$ and $Y_k^l(\mathbf{x})$. As stated above the random functions $Y_k^l(\mathbf{x})$ are mutually independent, but if two such functions have a common index l, then they have the same spatial correlation structure. There are n_l such structures with variogram functions $g_l(\mathbf{h})$. It can be shown that the cross-variograms of any two of the correlated random variables S_u and S_v can be expressed as a linear combination of the $L+1$ basic variogram functions:

$$\gamma_{\mathrm{uv}}(\mathbf{h}) = \sum_{l=0}^{L} b_{u,v}^l g_l(\mathbf{h}), \tag{10.42}$$

where

$$b_{u,v}^l = \sum_{k=1}^{n_l} a_{u,k}^l a_{v,k}^l.$$

Thus the cross-variogram matrix for N_w variables,

$$\mathbf{G}(\mathbf{h}) \equiv \begin{bmatrix} \gamma_{1,1}(\mathbf{h}) & \gamma_{1,2}(\mathbf{h}) & \ldots & \gamma_{1,N_w}(\mathbf{h}) \\ \gamma_{2,1}(\mathbf{h}) & . & \ldots & \gamma_{2,N_w}(\mathbf{h}) \\ . & . & \ldots & . \\ & & & . \\ \gamma_{N_w,1}(\mathbf{h}) & \gamma_{N_w,2}(\mathbf{h}) & \ldots & \gamma_{N_w,N_w}(\mathbf{h}) \end{bmatrix},$$

can be written as

$$\mathbf{G}(\mathbf{h}) = \sum_{l=0}^{L} g_l(\mathbf{h}) \mathbf{S}_l, \tag{10.43}$$

where

$$S_l \equiv \begin{bmatrix} b^l{}_{1,1} & b^l{}_{1,2} & \ldots & b^l{}_{1,N_w} \\ b^l{}_{2,1} & . & \ldots & b^l{}_{2,N_w} \\ . & . & \ldots & . \\ & & & . \\ b^l{}_{N_w,1} & b^l{}_{N_w,2} & \ldots & b^l{}_{N_w,N_w} \end{bmatrix}.$$

If the $L + 1$ matrices $\mathbf{S}_l$, co-regionalisation matrices, are all positive semi-definite and the $L + 1$ basic variogram functions are all authorised, then the LMCR has a positive-definite covariance structure.

To fit an LMCR, we must find a set of variogram functions in the corresponding co-regionalisation matrices that optimise a measure of the fit of $\mathbf{G}(\mathbf{h})$ to the point estimates of the variograms, subject to the constraints on the co-regionalisation matrices which we describe above. This is not trivial, particularly if the number of variables is increased. Goovaerts (1997) describes how an LMCR may be fitted by visual assessment and trial and error. A semiautomated procedure was devised by Goulard and Voltz (1992). If the distance parameters of the basic variogram functions are given, then this algorithm will find estimates of the corresponding co-regionalisation matrices optimising the measure of the goodness of fit and meeting the constraints on the co-regionalisation matrices. Kunsch et al.(1997) describe a way of fitting the LMCR using a non-linear regression method to fit the distance parameter and Goulard and Voltz's (1992) algorithm to fit the co-regionalisation matrices. Lark and Papritz (2003) show how the model can be fitted subject to these constraints using simulated annealing. Fitting the LMCR with estimates of the pseudo-cross-variogram is more complex but can be done under certain assumptions discussed by Papritz et al. (1993) and by Lark (2002).

10.5.2 Cokriging

When we have an LMCR for two variables, then we can proceed to spatial prediction by cokriging. The cokriging estimator is a linear predictor, like the kriging estimator we have already discussed. If soil variable s_1 is to be estimated at location $\mathbf{x}_0$ from observations of $s_1, s_2, \ldots, s_m$ at neighbouring locations, then the general linear predictor is

$$\widehat{s}_1(\mathbf{x}_0) - \mu_1(\mathbf{x}_0) = \sum_{i=1}^{n_{\mathbf{x}_0,1}} \lambda_{i,1} \{s_1(\mathbf{x}_i) - \mu_1(\mathbf{x}_i)\} + \sum_{j=2}^{m} \sum_{i=1}^{n_{\mathbf{x}_0,j}} \lambda_{i,j} \left\{s_j(\mathbf{x}_i) - \mu_j(\mathbf{x}_i)\right\}, \tag{10.44}$$

where $\mu_j(\mathbf{x}_i)$ is the expected value of s_j at $\mathbf{x}_i$. The cokriging estimator is the best linear unbiased estimator best in the sense that the cokriging variance

$$\sigma_C^2(\mathbf{x}_0) = E\left[\{\widehat{s}_1(\mathbf{x}_0) - S_1(\mathbf{x}_0)\}^2\right] \tag{10.45}$$

is minimised and unbiased in that $E[\widehat{s}_1(\mathbf{x}_0) - S_1(\mathbf{x}_0)] = 0$. As with univariate kriging, there are two alternative treatments of the local mean. The first is to regard it either as a known constant value for all locations (simple cokriging) or as an unknown value fixed within the neighbourhood about $\mathbf{x}_0$ from which kriged

estimates are derived (the ordinary cokriging). The second approach is to model it as a combination of known functions of location (simple cokriging with a trend). Here we consider the first option under ordinary kriging version. Let variable 1 be s_u the variable to be predicted from observations $s_u(\mathbf{x}_i)$ and observations of a secondary variable $s_v(\mathbf{x}_i)$ although this can be extended to a number of secondary variables. With a similar rearrangement of Eq. 10.44) to that done above for univariate ordinary kriging, it is clear that constraining the weights so that

$$\sum_{i=1}^{n_{x_0,u}} \lambda_{i,u} = 1$$

and

$$\sum_{i=1}^{n_{x_0,v}} \lambda_{i,v} = 0$$

will filter the (unknown) means from the estimator, while the resulting estimator is unbiased. These are two independent constraints, so the kriging equations need two Lagrange parameters. By a similar derivation to univariate ordinary kriging, we obtain our cokriging equations which may be solved for the weights to be substituted into the cokriging estimator

$$\widehat{s}_u(\mathbf{x}_0) = \sum_{i=1}^{n_{x_0,u}} \lambda_{i,u} s_u(\mathbf{x}_i) + \sum_{i=1}^{n_{x_0,v}} \lambda_{i,v} s_v(\mathbf{x}_i). \tag{10.46}$$

There are three general instances in which we use cokriging. The classical case is where the secondary variable is more densely sampled than the target variable, and so the cross-covariance information allows us to obtain better predictions of the target variable than we could using the data on that variable alone. An interesting example of this approach is given by Leenaers et al. (1990). In general counting for the cross-covariance structure will give improved predictions when the cross-covariance structure includes information on scale-dependent relationships between the variables, i.e. the co-dispersion coefficient changes with lag. If the co-dispersion coefficient is constant, then the cross-covariance model adds no information.

When the primary and secondary variables are equally sampled, then there is generally little advantage from cokriging in terms of improved precision of estimates, but we do achieve coherence as discussed above. Cokriging linear combinations of variables directly (such as the change variable where s_u and s_v are measurements of the same variable at time u and time v) is described in more detail by Papritz and Fluhler (1994), and an example of this approach applied to soil monitoring is given by Lark et al. (2006). One special case in which these properties of cokriging are useful is when we work with compositional data, that is to say,

variables which sum to a fixed value (typically 1, or 100%). In soil science we are often interested in the sand, silt and clay content of soils. These are exhaustive particle size classes, and the values sum to one. It is inappropriate to analyse the proportions or percentages separately since separate OK predictions of sand, silt and clay percentage are not guaranteed to sum to 100. An alternative is to apply cokriging to the additive log-ratio of the particle size fractions. This methodology is set out in detail by Pawlowsky-Glahn and Olea (2004), and Lark and Bishop (2007) illustrate the method applied to data on soil particle size fractions.

A third instance is where we require information to support a decision which depends on two or more variables and when we are interested in the uncertainty in the information. In this case the joint prediction error, characterised by a covariance matrix for kriging estimates, can be obtained as output from cokriging. Lark et al. (2014) give an example in a study in the northern counties of the Irish Republic, and we present this as an example of the application of cokriging.

Advice to farmers in Ireland on the risk of cobalt deficiency for grazing livestock is based on the total concentration of cobalt and manganese in the soil. This is because manganese oxides can bind cobalt, reducing its availability. Lark et al. (2014) used data from a large regional survey of the soil. They conducted two separate geostatistical analyses in two geological domains with contrasting behaviour. Figure 10.11 shows the auto-variograms for Co and Mn and the cross-variograms for both domains, using both the standard estimator in Eq. 10.39 and a robust estimator from Lark (2003). The cokriged maps for Co and manganese are shown in Fig. 10.12. From the cokriging estimates and covariance matrix of cokriging errors, it was then possible, assuming jointly normal errors, to estimate the probability that the soil concentrations at any location fell within a range of values where a Co deficiency would be expected. This probability is mapped in Fig. 10.13.

We are most usually interested in multivariate methods of spatial prediction when the soil variable of primary interest is costly or difficult to measure and so is sampled relatively sparsely. However, one or more secondary variables are available which are easy and/or cheap to measure and so can be sampled densely and are also thought to contain information about the primary variable. Such variables may be remotely sensed measurements of the earth's surface, digital elevation models of terrain or data from geophysical surveys. When it is plausible to treat our variables primary and secondary as intrinsic random functions which conform to an LMCR, then the cokriging described above provides the best linear unbiased estimator of the primary variable given the secondary variable. There are other approaches that should be considered, however. In particular, we can use the secondary variable as a fixed effect in the REML–E-BLUP discussed above. This approach is known as kriging with an external drift, and examples of this approach are given by Bourennane et al. (1996).

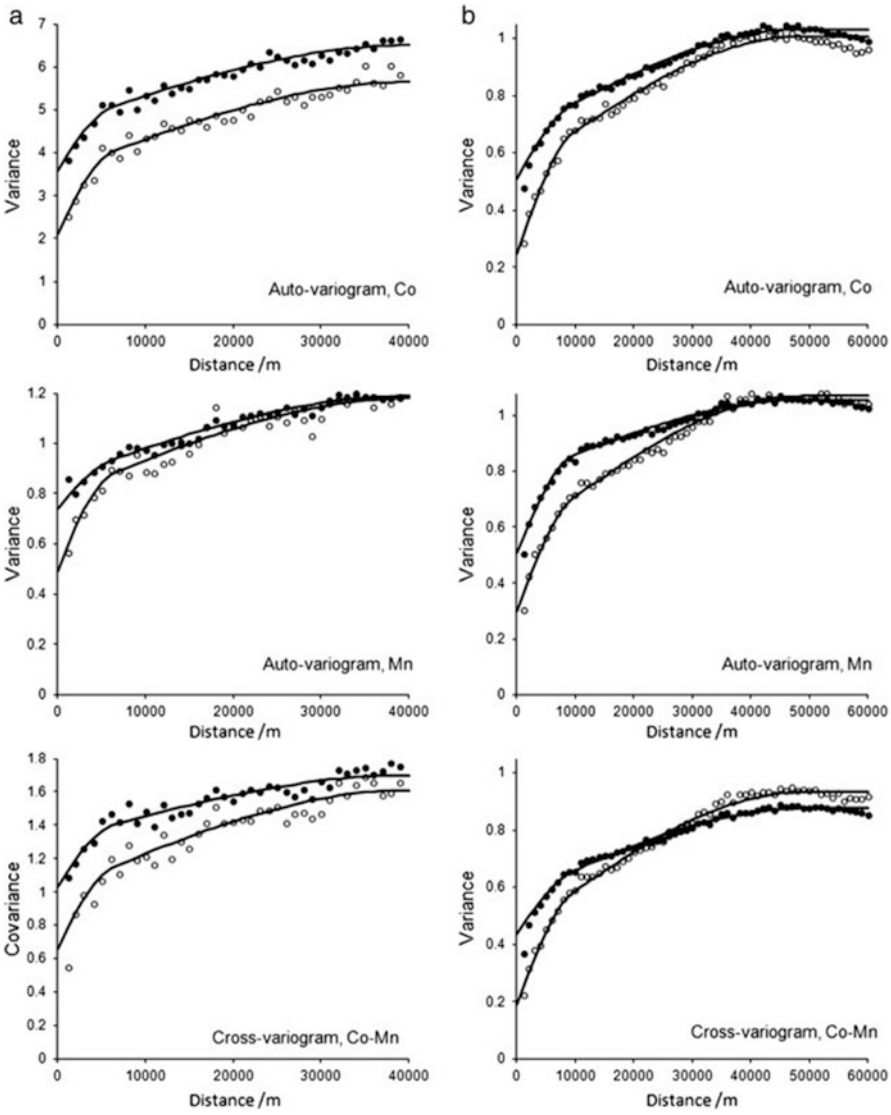


Fig. 10.11 Auto-variograms and cross-variograms for topsoil Co and Mn in two geological domains (**a**, **b**) of the northern counties of Ireland using (*solid symbol*) the standard estimators and (*open symbol*) a robust estimator, with fitted linear models of co-regionalisation (From Lark et al. (2014), under CC-BY licence. https://creativecommons.org/licenses/by/3.0/)

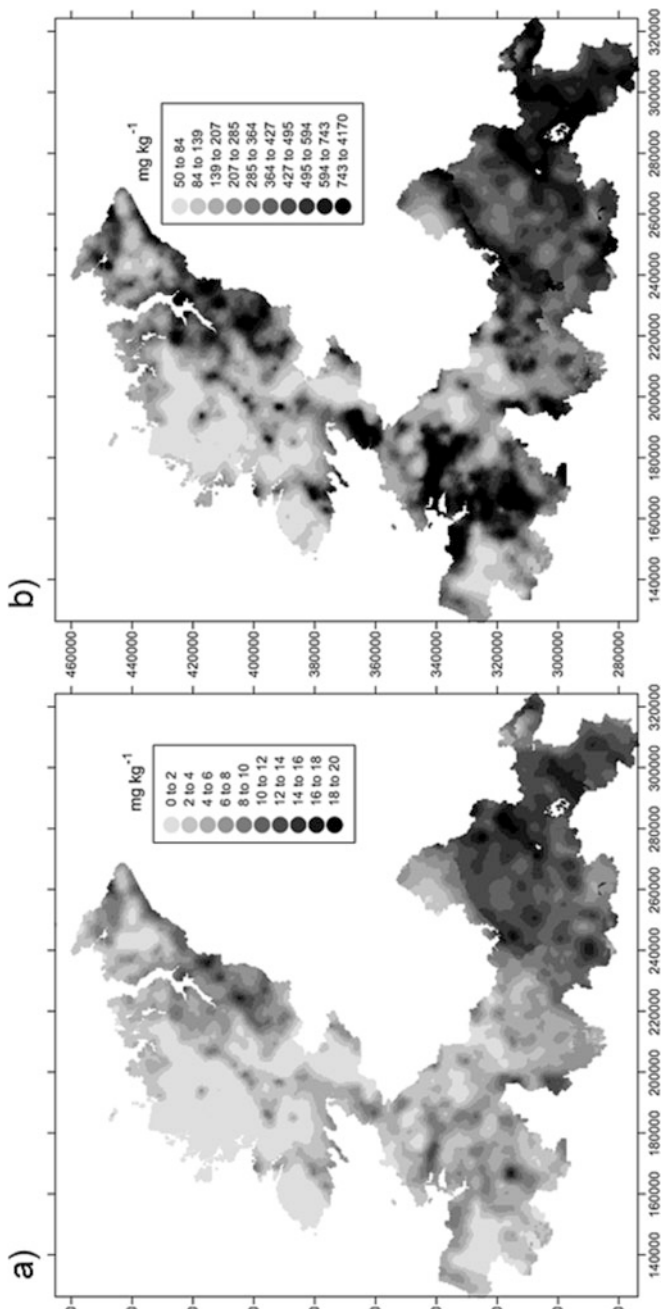


Fig. 10.12 Ordinary cokriging predictions of (**a**) topsoil Co and (**b**) topsoil Mn content across the northern counties of Ireland. Coordinates are the Irish National Grid (From Lark et al. (2014), under CC-BY licence. https://creativecommons.org/licenses/by/3.0/)

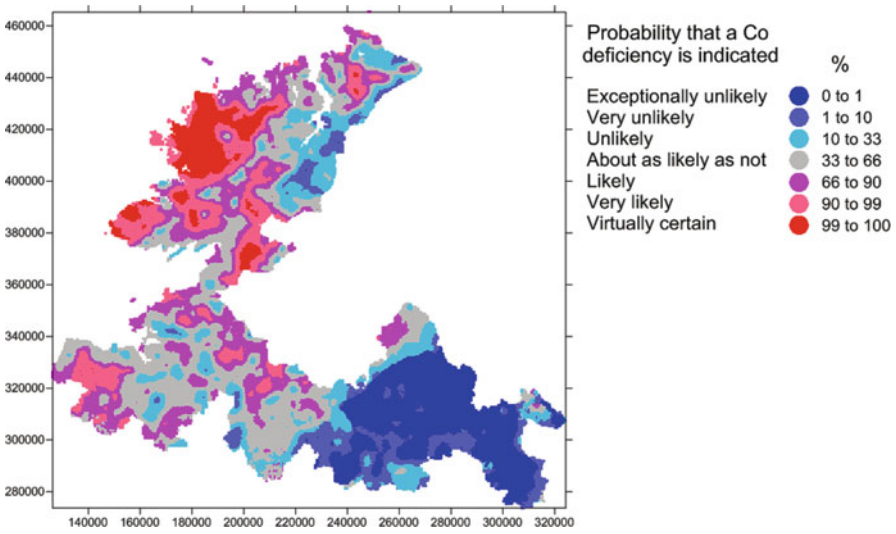


Fig. 10.13 Probability that the local Co and Mn content of soil indicates a risk of Co deficiency to livestock computed from the cokriging predictions and their error covariance matrices (From Lark et al. (2014), under CC-BY licence. https://creativecommons.org/licenses/by/3.0/)

10.6 Spatial Simulation

10.6.1 Simulation vs Prediction

We have discussed at length the problem of estimating a variable at a location from a set of observations on the assumption that our data are a realisation of a random function. Spatial simulation is a different problem. In simulation we draw several different realisations of the random function that we assume to be realised in our data. There are two general types of simulation. In *unconditional simulation* each set of simulated values represents a realisation of a random function with specified statistics (variogram, mean, histogram). In *conditional simulation* the realisation of the random function from which our values are drawn has specified statistics as in unconditional simulation, but is also required to reproduce the observed values at locations where we have sampled. We can think of unconditional simulation as drawing at random a set of values from a set of all possible realisations of a random function with specified statistics, and conditional simulation is drawn at random from a smaller subset of those sets of values which have the observed values at the sample sites.

Why might we want to do simulation? A good deal of simulation is done for pedometrical research or research in allied disciplines for a data with a realistic pattern of spatial variation in order to test the behaviour of pedometrical methods for estimation (e.g. Webster and Oliver 1992; Papritz and Webster 1995). But

simulation is a pedometrical tool in its own right, not simply a tool for research. It is seen in Chap. 13 how simulation can be used to examine the propagation of error in sources of information through models. If we generate multiple realisations of a random function conditional on observations and then look at distribution of simulated values at a particular location, these provide us with an estimate of the conditional cumulative distribution function (ccdf) of the variable at the location (Journel 1994a, b).

Conditional simulation is also used when we need to tackle problems where the variable of interest is obtained by some complex non-linear process model and our data are key model inputs. Consider for example the problem of predicting concentration of pollutant in water from a borehole when a plume of the pollutant has been released at the soil surface some distance away. The movement of the pollutant can be modelled if we know among other things the unsaturated conductivity functions of the soil at intervening points. This can only be measured at a few sites. You might interpolate the conductivities at intervening sites by kriging. However, since the model is non-linear, the fact that the kriged estimates of the conductivity are necessarily much less variable than the true values (although each estimate is unbiased) means that the simulated behaviour of the pollutant is a biased estimate of the true behaviour. An alternative is to generate a conditional simulation of the field of conductivities and to apply the model to these values. Multiples of such realisations can be generated, and from these we may generate a ccdf for an important model output such as the peak concentration of the pollutant in the well water. Other examples could be given where we wish to explore the aggregate impact on some region of some non-linear function of the spatially variable soil property, e.g. Viscarra Rossel et al. (2001) use simulation to estimate the cost function of different strategies for spatially variable application of lime to agricultural fields.

10.6.2 Methods for Simulation

LU Decomposition The most straightforward method for spatial simulation is known as LU decomposition. This is long established as a method to simulate correlated variables; an account of its use in geostatistics is given by Davis (1987). The name is somewhat misleading since it more properly describes the central algebraic step of the process where we perform a factorisation of a matrix into lower (L) and upper (U) triangular matrices. LU decomposition generates realisations of a multivariate normal process. This means that, unless we can assume that our data are from a multivariate normal process, we must transform them to normality. This is done typically by a standard normal score transform where we order our data values and replace each with a corresponding centile of a standard normal variable. The transformed data are stored in a row vector **s**. We then test the plausibility of bivariate normality for pairs of data (see Goovaerts 1997 for details) and then estimate and model a variogram for the normal score transformation

of the data. From this, via Eq. 10.3, we can then compute a covariance matrix for the nodes at which we wish to simulate and the nodes at which we have data on which the simulation is to be conditioned. Element [*i,j*] of covariance matrix $\mathbf{C}$ therefore contains the covariance between node i and node j which, under the stationarity assumption, depends only on the distance in space between the nodes. We order the nodes so that our covariance matrix can be thought of as four component matrices $\mathbf{C}_{D,D}$ (covariance among conditioning data points), $\mathbf{C}_{N,N}$ (covariance among simulation nodes) and $\mathbf{C}_{N,D} = \mathbf{C}^T{}_{D,N}$ (covariance between conditioning data and simulation nodes), so

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{D,D} & \mathbf{C}_{D,N} \\ \mathbf{C}_{N,D} & \mathbf{C}_{N,N} \end{bmatrix}. \tag{10.47}$$

The Cholesky decomposition is a factorisation so that

$$\mathbf{C} = \mathbf{LU} = \begin{bmatrix} \mathbf{L}_{D,D} & \mathbf{0}_{D,N} \\ \mathbf{L}_{N,D} & \mathbf{L}_{N,N} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{D,D} & \mathbf{U}_{D,N} \\ \mathbf{0}_{N,D} & \mathbf{U}_{N,N} \end{bmatrix}, \tag{10.48}$$

where $\mathbf{0}_{D,N}$ ($=\mathbf{0}^T{}_{N,D}$) is a $D \times N$ matrix of zeros. Now, if N standard independent normal random variables are generated in column vector $\mathbf{y}$, then a realisation of our random variable is given in row vector $\breve{\mathbf{s}}$ by

$$\breve{\mathbf{s}} = \mathbf{L}_{N,D}\mathbf{L}_{D,D}\mathbf{s} + \mathbf{L}_{N,N}\mathbf{y}. \tag{10.49}$$

The standard normal variable is then back transformed with the empirical normal score transform originally used on the raw data.

In many ways this is the most theoretically satisfactory method for simulation. Its major drawback in theoretical terms is the assumption of multivariate normality. This assumption can never be disproved or proved for real data; the normal score transform can ensure reasonable representation of the desired histogram only, which is a necessary but not a sufficient condition for the multivariate assumption to be plausible. The multivariate normal assumption can cause problems in applications of simulation to the kinds of problem described above where we wish to use the simulated field in a model. Because the multivariate normal distribution of a set of data has maximum entropy, the extreme values of the distribution tend to be disconnected spatially. Thus in our example of pollutant movement, we find that regions with very large conductivity, which may occur in nature and have a disproportionately large effect on transport through the soil, will not be adequately reproduced by simulation. Another limitation of LU composition is computational. The Cholesky factorisation step is computationally expensive, and commonly it is impractical to simulate more than a few thousand data points. This is not a very large two-dimensional array. Once the decomposition has been done, however, as many realisations as required can be very rapidly computed.

Pebesma and Heuvelink proposed the use of Latin hypercube sampling (LHS) rather than complete random samples (Eq. 10.9) to more efficiently generate the random field. In their simulation study, they found that LHS of size 20 can perform equally as well as a simple random sample using 2,000 realisations.

Sequential methods Sequential simulation has been developed in response to the limitations that the computational load of the Cholesky composition imposes on the size of the fields that may be simulated using LU methods (Journel 1994b). A good account is given by Goovaerts (1997). Consider the simple problem of simulating a realisation of a random function at two locations $\mathbf{x}_1$ and $\mathbf{x}_2$, conditional on n data in the set γ_n. We may characterise the joint distribution of $S(\mathbf{x}_1)$ and $S(\mathbf{x}_2)$ by the bivariate conditional cumulative frequency distribution

$$F\left(\mathbf{x}_1, \mathbf{x}_2; z_1, z_2 \,|\zeta_n\right) \equiv \Pr\left\{s\left(\mathbf{x}_1\right) \leq s_1, s\left(\mathbf{x}_2\right) \leq s_2 \,|\zeta_n\right\}. \tag{10.50}$$

Following Bayes' rule, we can write

$$F\left(\mathbf{x}_1, \mathbf{x}_2; s_1, s_2 \,|\zeta_n\right) = F\left(\mathbf{x}_2; s_2 \,|\zeta_n, s\left(\mathbf{x}_1\right)\right) F\left(\mathbf{x}_1; s_1 \,|\zeta_n\right) \tag{10.51}$$

where the condition of the first ccdf on the right-hand side refers to the n conditioning data and one realisation generated at $\mathbf{x}_1$. This is why the simulated method is sequential because the data of simulated modes are generated as realisations of processes with conditional distribution functions conditioned on a sequence of modes and, ultimately, the conditioning data.

The data are transformed to normality and the plausibility of the bivariate normal assumption is tested. The conditional cumulative distribution function (ccdf) at the first node to be simulated is specified as N (m,v) where m is a kriged estimate at the node and v is the simple kriging variance. Note that simple kriging, ordinary kriging or universal kriging can be used to generate m, but the simple kriging variance must always be specified. Similarly an indicator transform can be used if we wish to simulate indicator variables, but the simple indicator kriging must be used to generate v. The simulation precedes sequentially from one simulation node to another accumulating the simulated values as conditioning ones.

The sequential simulation procedure is efficient. It can be made more so. In kriging there is a "screening effect" whereby the influence of a datum on a krige estimate is masked by an intervening observation. Thus there is little loss of information but a gain in speed if local ccdfs are calculated from all the conditioning data but only the nearest neighbouring simulation modes. In fact the screening effect can cause problems for reproducing long-range correlation structures. For this reason it is sometimes preferred to simulate on a coarse sub-grid of modes and then to "fill in" in intervening locations. Such multiple grids should be visited in a sequence that is randomised between realisations.

In Fig. 10.14 we show five independent conditional simulations of soil copper across our study area in the east of Scotland. These were obtained at the same prediction sites as the kriging predictions in Fig. 10.9 by sequential Gaussian

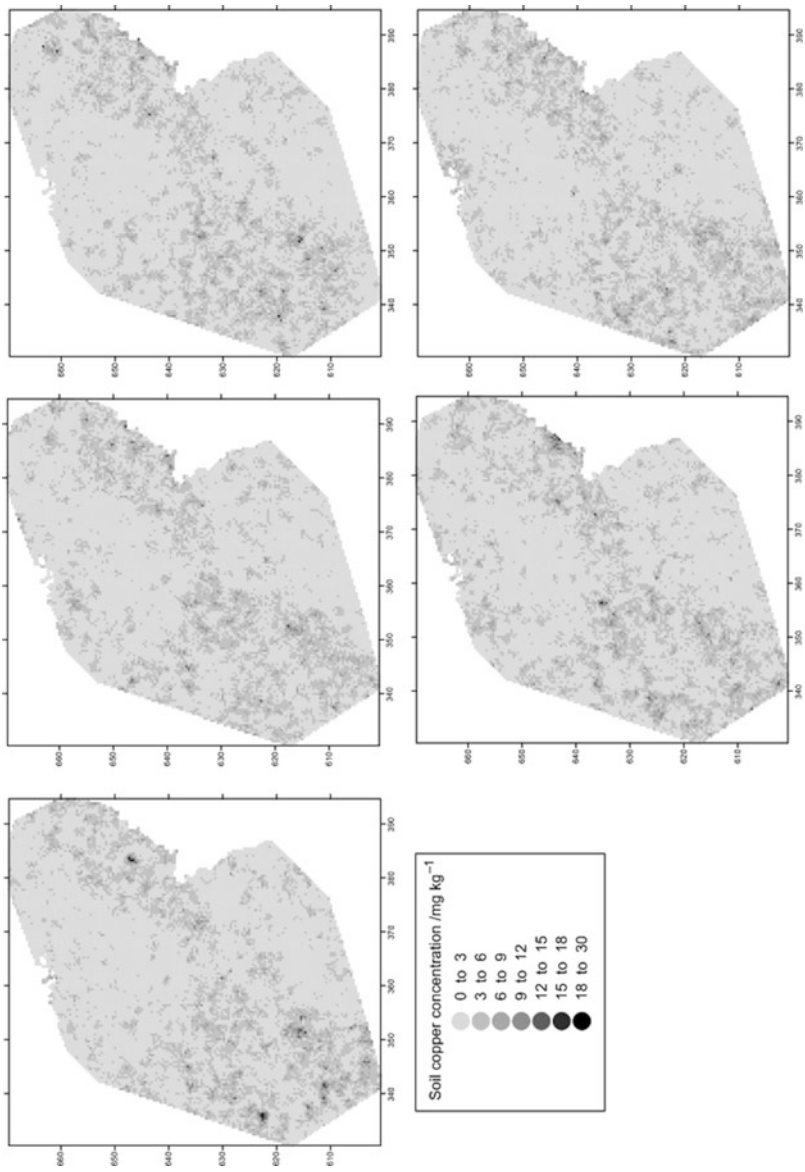


Fig. 10.14 Five mutually independent conditional simulations by sequential Gaussian simulation of lead content in the topsoil of the east of Scotland

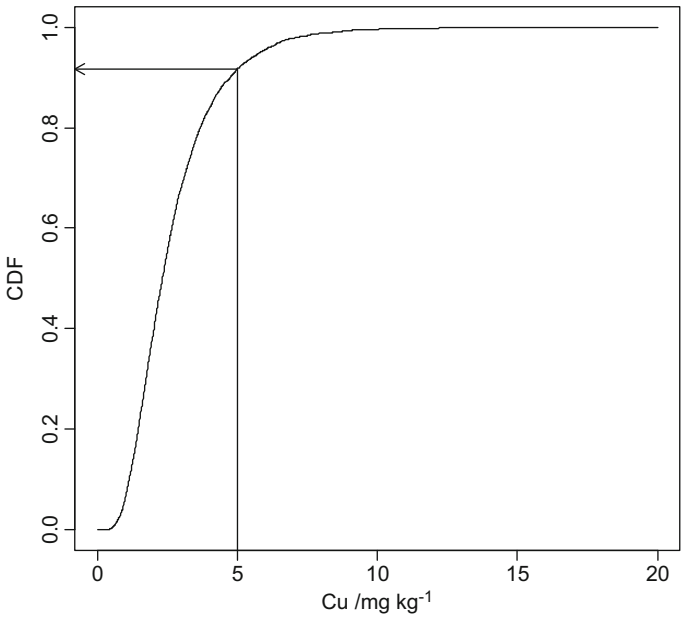


Fig. 10.15 Empirical conditional cumulative distribution function for topsoil lead content at a site (coordinates 340,642) in the east of Scotland, obtained from 5,000 mutually independent conditional simulations by sequential Gaussian simulation

simulation of the log-transformed variable. Each simulated value was then back transformed to the original scale of measurement. In Fig. 10.15 we consider a single location (coordinates 340,642). Here we simulated 5,000 independent realisations, conditional on the data. The figure shows the empirical cumulative distribution function (CDF) of these values, which is the conditional CDF of the variable at that site (conditional on the geostatistical model and the data). We may read off the value of the empirical CDF for some concentration of copper (here 5 mg kg^{-1}) from this graph. The value is 0.92, which allows us to estimate the conditional probability that soil copper content at that location exceeds 5 mg kg^{-1} as 0.08 or 8%.

The usual sequential Gaussian simulation is necessarily tied to assumptions of multivariate normality. These may be questionable and, as we discuss above in LU context, inappropriate assumptions of multivariate normality can cause problems for the simulation of conductivity fields or other variables where the connectedness of large values can have a substantial effect on aggregate behaviour. This is the motivation for recent studies of multiple-point geostatistics in soil science (e.g. Meerschman et al. 2013).

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petov BN, Csaki F (eds) 2nd international symposium on information theory. Akademia Kiado, Budapest, pp 267–281

Barry RP, Kelley Pace R (1997) Kriging with large data sets using sparse matrix techniques. Commun Stat-Simul Comput 26(2):619–629

Beckett PHT, Bie SW (1976) Reconnaissance for soil survey. II. Pre-survey estimates of the intricacy of the soil pattern. J Soil Sci 26:101–110

Beckett PHT, Webster R (1971) Soil variability: a review. Soils Fertil 34:1–15

Bourennane H, King D, Chéry P, Bruand A (1996) Improving the kriging of a soil variable using slope gradient as external drift. Eur J Soil Sci 47

Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138:86–95

Burgess TM, Webster R (1980a) Optimal interpolation and isarithmic mapping of soil properties I the semi-variogram and punctual kriging. J Soil Sci 31:315–331

Burgess TM, Webster R (1980b) Optimal interpolation and isarithmic mapping of soil properties II block kriging. J Soil Sci 31:333–341

Burgess TM, Webster R, McBratney AB (1981) Optimal interpolation and isarithmic mapping of soil properties IV sampling strategy. J Soil Sci 32:643–659

Burrough PA (1983) Multiscale sources of spatial variation in soil. II. A non-Brownian fractal model and its application in soil survey. J Soil Sci 34:599–620

Buttafuoco G, Castrignano A (2005) Study of the spatio-temporal variation of soil moisture under forest using intrinsic random functions of order *k*. Geoderma 128:208–220

Chilès J, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Clark I, Basinger KL, Harper WV (1989) MUCK – a novel approach to cokriging. In: Buxton BE (ed) Proceedings of the conference on geostatistical, sensitivity and uncertainty methods for ground-water flow and radionuclide transport modeling. Battelle Press, Columbus, pp 473–493

Corstanje R, Grunwald S, Lark RM (2008) Inferences from fluctuations in the local variogram about the assumption of stationarity in the variance. Geoderma 143:123–132

Cressie N (1985) Fitting variogram models by weighted least squares. Math Geol 17:563

Cressie N (1993) Statistics for spatial data. Wiley, New York

Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. J R Stat Soc: Ser B (Stat Methodol) 70(1):209–226

Cressie, N., Kang, E.L., 2010. High-resolution digital soil mapping: Kriging for very large datasets. In: Proximal soil sensing. Springer, Dordrecht, pp 49–63

Davis MW (1987) Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Math Geol 19:91–98

Deutsch CV, Journel AG (1998) GSLIB geostatistical software and user's guide, 2nd edn. Oxford University Press, New York

Dowd PA (1984) The variogram and kriging: robust and resistant estimators. In: Verly G, David M, Journel AG, Marechal AŽ (eds) Geostatistics for natural resources characterization. Part 1. Reidel, Dordrecht, pp 91–106

Goovaerts P (1994) Comparative performance of indicator algorithms for modeling conditional probability distribution functions. Math Geol 26:389–411

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math Geol 24:269–286

Haas TM (1990) Lognormal and moving window methods of estimating acid deposition. J Am Stat Assoc 85:950–963

Halvorson JJ, Smith JL, Bolton H, Rossi RE (1995) Evaluating shrub-associated patterns of soil properties in a shrub steppe ecosystem using multiple-variable geostatistics. Soil Sci Soc Am J 59:11476–11487

Isaaks E, Srivastava RM (1989) An introduction to applied Geostatistics. Oxford University Press, New York
Journel AG (1983) Non-parametric estimation of spatial distributions. Math Geol 15:445–468
Journel AG (1994a) Resampling from stochastic simulations. Environ Ecol Stat 1:63–91
Journel AG (1994b) Modelling uncertainty: some conceptual thoughts. In: Dimitrakopoulos R (ed) Geostotistics for the next century. Kluwer Academic, Dordrecht, pp 30–43
Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London
Kaufman CG, Schervish MJ, Nychka DW (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. J Am Stat Assoc 103(484):1545–1555
Künsch HR, Papritz A, Bassi F (1997) Generalized cross-covariances and their estimation. Math Geol 29:779–799
Lark RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. Eur J Soil Sci 51:137–157
Lark RM (2002) Robust estimation of the pseudo cross-variogram for cokriging soil properties. Eur J Soil Sci 53:253–270
Lark RM (2003) Two robust estimators of the cross-variogram for multivariate geostatistical analysis of soil properties. Eur J Soil Sci 54:187–201
Lark RM, Bishop TFA (2007) Cokriging particle size fractions of the soil. Eur J Soil Sci 58: 763–774
Lark RM, Ferguson RB (2004) Mapping risk of soil nutrient deficiency or excess by disjunctive and indicator kriging. Geoderma 118:39–53
Lark RM, Papritz A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115:245–260
Lark RM, Bellamy PH, Rawlins BG (2006) Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. Geoderma 133:363–379
Lark RM, Ander EL, Cave MR, Knights KV, Glennon MM, Scanlon RP (2014) Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. Geoderma 226–227:64–78
Leenaers H, Okx JP, Burrough PA (1990) Comparison of spatial prediction methods for mapping floodplain soil pollution. Catena 17:535–550
Marchant BP, Lark RM (2006) Adaptive sampling for reconnaissance surveys for geostatistical mapping of the soil. Eur J Soil Sci 57:831–845
Marchant BP, Lark RM (2007) Optimized sample schemes for geostatistical surveys. Math Geol 39:113–134
Matérn B (1986) Spatial variation. Springer, Berlin
Matheron G (1962) Traité de géostatistique appliquée, Tome I: Mémoires du Bureau de Recherch-esGéologiques et Minières, vol 14. Editions Technip, Paris
Matheron G (1976) A simple substitute for conditional expectation: the disjunctive kriging. In: Guarascio M, David M, Huijbregts C (eds) Advanced geostatistics in the mining industry. D. Reidel, Dordrecht, pp 221–236
McBratney AB, Webster R (1983) Optimal interpolation and isarithmic mapping of soil properties V. Co-regionalization and multiple sampling strategy. J Soil Sci 34:137–162
McBratney AB, Webster R (1986) Choosing functions for semivariances of soil properties and fitting them to sample estimates. J Soil Sci 37:617–639
McBratney AB, Webster R, Burgess TM (1981) The design of optimal sampling schemes for local estimation and mapping of regionalised variables. I. Theory and method. Comput Geosci 7:331–334
McBratney AB, Webster R, McLaren RG, Spiers RB (1982) Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland. Agronomie 2:969–982
Meerschman E, Van Meirvenne M, Van De Vijver E, De Smedt P, Islam MM, Saey T (2013) Mapping complex soil patterns with multiple-point geostatistics. Eur J Soil Sci 64:183–191
Minasny B, McBratney AB (2005) The Matérn function as a general model for soil variograms. Geoderma 128:192–207

Minasny B, McBratney AB, Whelan BM (2006) VESPER version 1.62. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney. NSW
Myers DE (1991) Pseudo-cross variograms, positive definiteness, cokriging. Math Geol 23: 805–816
Odeh IOA, McBratney AB, Chittleborough DJ (1994) Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63:197–214
Papritz A, Dubois J-P (1999) Mapping heavy metals in soil by (non-)linear kriging: an empirical validation. In: Gomez-Hernandez J, Soares A, Froidevaux R (eds) geoENV II: Geostatistics for environmental applications, Quantitative geology and geostatistics, vol 10. Kluwer Academic Publishing, Dordrecht, pp 429–440
Papritz A, Flühler H (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. Geoderma 62:29–43
Papritz A, Webster R (1995) Estimating temporal change in soil monitoring. 1. Statistical theory. Eur J Soil Sci 46:1–12
Papritz A, Kunsch HR, Webster R (1993) On the pseudo cross-variogram. Math Geol 25: 1015–1026
Pawlowsky-Glahn V, Olea RA (2004) Geostatistical analysis of compositional data. Oxford University Press, New York
Ploner A (1999) The use of the variogram cloud in geostatistical modelling. Environmetrics 10:413–437
Rivoirard J (1994) Introduction to disjunctive kriging and non-linear geostatistics. Oxford University Press, Oxford
Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York
Sun W, Minasny B, McBratney A (2012) Analysis and prediction of soil properties using local regression-kriging. Geoderma 171:16–23
Van Meirvenne M, Goovaerts P (2001) Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102:75–100
Viscarra Rossel RA, Goovaerts P, McBratney AB (2001) Assessment of the production and economic risks of site-specific liming using geostatistical uncertainty modelling. Environmetrics 12:699–711
von Steiger B, Webster R, Schulin R, Lehmann R (1996) Mapping heavy metals in polluted soil by disjunctive kriging. Environ Pollut 94:205–215
Webster R (2015) Daniel Krige's influence on soil science. J South Afr Inst Min Metall 11:165–117
Webster R, Burgess TM (1980) Optimal interpolation and isarithmic mapping of soil properties III changing drift and universal kriging. J Soil Sci 31:505–524
Webster R, Burgess TM (1984) Sampling and bulking strategies for estimating soil properties in small regions. J Soil Sci 35:127–140
Webster R, Butler BE (1976) Soil survey and classification studies at Ginninderra. Aust J Soil Res 14:1–24
Webster R, Lark RM (2013) Field sampling for environmental science and management. Routledge, London. pp viii+192
Webster R, Oliver MA (1992) Sample adequately to estimate variograms of soil properties. J Soil Sci 43:177–192
Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester. pp. x+328
Wood G, Oliver MA, Webster R (1990) Estimating soil salinity by disjunctive kriging. Soil Use Manag 6:97–104
Yates F (1948) Systematic sampling. Phil Trans R Soc A 241:345–377
Youden WJ, Mehlich A (1937) Selection of efficient methods for soil sampling. Contrib Boyce Thompson Inst Plant Res 9:59–70

Chapter 11
Model-Based Soil Geostatistics

Ben P. Marchant

"A model's just an imitation of the real thing".

Mae West

11.1 Introduction

In Chap. 10, we described how classical geostatistical methods can be used to interpolate measurements of soil properties at locations where they have not been observed and to calculate the uncertainty associated with these predictions. The idea that soil properties can be treated as realizations of regionalized random functions in this manner has perhaps been the most significant ever in pedometrics (Webster 1994). The approach has been applied in thousands of studies for every imaginable soil property at scales varying from the microscopic to the global and has greatly enhanced our understanding of the spatial variability of soil properties.

However, despite its popularity amongst pedometricians, the classical geostatistical methodology has met with some criticism (e.g. Stein 1999). This has primarily been because of the subjective decisions that are required when calculating the empirical semivariogram and then fitting the variogram model. When calculating the empirical semivariogram, the practitioner must decide on the directions of the lag vectors that will be considered, the lag distances at which the point estimates are to be calculated and the tolerance that is permitted for each lag distance bin (see Sect. 10.2). Furthermore, when fitting a model to the empirical variogram, the practitioner must decide which of the many authorized models should be used, what criterion should be applied to select the best fitting parameter values and the weights that should be applied when calculating this criterion for the different point estimates. Researchers have given some thought to how these selections should be made. For example, the Akaike information criterion (Akaike 1973) is often used

B.P. Marchant (✉)
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
e-mail: benmarch@bgs.ac.uk

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_11

to select the authorized model that best fits the data. However, the formula used to calculate the AIC when using the method of moments variogram estimator is only an approximation and is very much dependent on how the fitting criterion is weighted for different lag distances (McBratney and Webster 1986). The selection of these weighting functions is a particular challenge since the uncertainty of the point estimates and hence the most appropriate choice of weights depend on the actual variogram which the user is trying to estimate. Further complications arise because the same observations feature in multiple point estimates of the experimental variogram, and some observations can feature more often than others. Therefore, these point estimates are not independent (Stein 1999). Strictly this correlation between the point estimates should be accounted for when fitting the model parameters. The correlation can result in artefacts or spikes in the experimental variogram that might be mistaken for additional variance structures (Marchant et al. 2013b).

In all of the choices listed above, different subjective decisions (or indeed carefully manipulated choices) can lead to quite different estimated variograms and in turn to quite different conclusions about the spatial variation of the soil property. Therefore, there is a need for a single objective function that quantifies how appropriately a proposed variogram model represents the spatial correlation observed in a dataset without requiring the user to make subjective decisions. Model-based geostatistics (Stein 1999; Diggle and Ribeiro 2007) uses the likelihood that the observations would arise from the proposed random function as such an objective criterion. The likelihood function is calculated using the observed data without the need for an empirical semivariogram. The variogram parameter values that maximize the likelihood function for a particular set of observations correspond to the best fitting variogram model. These parameters are referred to as the maximum likelihood estimate. Once a model has been estimated, it can be substituted into the best linear unbiased predictor (BLUP; Lark et al. 2006; Minasny and McBratney 2007) to predict the expected value of the soil property at unobserved locations and to determine the uncertainty of these predictions. The relative suitability of two proposed variograms can be assessed by the ratio of their likelihoods or the exact AIC. Hence, it is possible to compare objectively fitted models which use different authorized variogram functions. Rather than selecting a single best fitting variogram, it is also possible to identify a set of plausible variograms according to the closeness of their likelihoods to the maximized likelihood. Thus, the uncertainty in the estimate of the variogram can be accounted for by averaging predictions across this set of plausible variograms (Minasny et al. 2011).

The linear mixed model (LMM) is often used in model-based geostatistical studies. The LMM divides the variation of the observations into fixed and random effects. The fixed effects are a linear function of environmental covariates that describe the variation of the expectation of the random function across the study region. The covariates can be any property that is known exhaustively, such as the eastings or northings, the elevation and derived properties such as slope or a remotely sensed property. The random effects have zero mean everywhere, and they describe the spatially correlated fluctuations in the soil property that cannot be explained by the fixed effects.

The model-based approach does have its disadvantages. The formula for the likelihood function for n observations includes the inverse of an $n \times n$ square matrix. This takes considerably longer to compute than the weighted difference between an empirical semivariogram and a proposed variogram function. Mardia and Marshall (1984) suggested that maximum likelihood estimation was impractical for $n > 150$. Modern computers can calculate the likelihood for thousands of observations (e.g. Marchant et al. 2011), but it is still impractical to calculate a standard likelihood function for the tens of thousands of observations that might be produced by some sensors. The use of the likelihood function also requires strong assumptions about the random function. The most common assumption being that the random effects are realized from a second-order stationary multivariate Gaussian random function. Such a restrictive set of assumptions is rarely completely appropriate for an environmental property and therefore the reliability of the resultant predictions is questionable. One active area of research is the development of model-based methodologies where these assumptions are relaxed.

The software required to perform model-based geostatistics has been made available through several R packages such as geoR (Ribeiro and Diggle 2001) and gstat (Pebesma 2004). The software used in this chapter has been coded in MATLAB. It forms the basis of the Geostatistical Toolbox for Earth Sciences (GaTES) that before the end of 2017 will be available at http://www.bgs.ac.uk/StatisticalSoftware. The Bayesian analyses require the DREAM package (Vrugt 2016).

11.2 The Scottish Borders Dataset

We illustrate some of the model-based geostatistics approaches that have been adopted by pedometricians by applying them to a set of measurements of the concentrations of copper and cobalt in soils from the south east of Scotland (Fig. 11.1). This dataset was analysed using classical geostatistical methods in Chap. 10. The measurements were made between 1964 and 1982. At that time, there were concerns that livestock grazing in the area were deficient in copper and/or cobalt. Therefore, staff from the East of Scotland College of Agriculture measured the field-mean extractable concentrations of these elements in more than 3500 fields. Full details of the sampling protocol and laboratory methods are provided by McBratney et al. (1982). The dataset has been extensively studied using a selection of geostatistical techniques (McBratney et al. 1982; Goovaerts and Webster 1994; Webster and Oliver 2007). These authors have mapped the probabilities that the concentrations of copper and cobalt are less than acceptable thresholds (1.0 and 0.25 mg kg^{-1}, respectively) and related their results to the previously mapped soil associations.

We consider a subset of the data. We randomly select 400 copper measurements and 500 cobalt measurements from this dataset. The remaining copper observations are kept for model validation (Fig. 11.2). We use soil information from the

Fig. 11.1 Location of the Scottish Borders study region (*shaded region*) within Great Britain

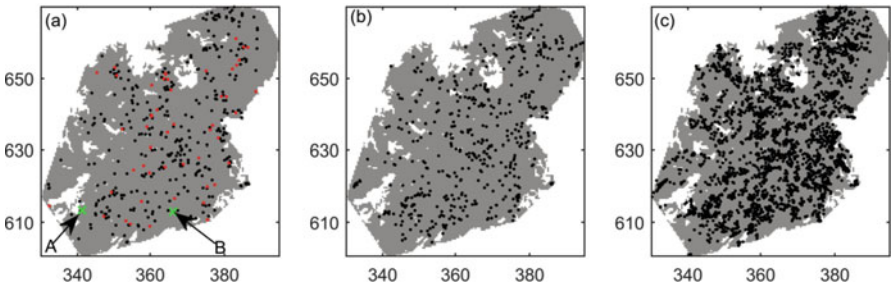


Fig. 11.2 Locations of (**a**) 400 copper observations used for calibration, (**b**) 500 cobalt observations used for LMCR calibration and (**c**) 2481 copper observations used for model validation. The locations coloured *red* in (**a**) are the 50-point subset of these data, and the *green* crosses are the locations of the predictions shown in Figs. 11.8 and 11.9. Coordinates are km from the origin of the British National Grid

1:250,000 National Soil Map of Scotland (Soil Survey of Scotland Staff 1984). These data are available under licence from http://www.soils-scotland.gov.uk/data/soils. The study area contains eight soil types. We only consider four of these, namely, mineral gleys, peaty podzols, brown earths and alluvial soils (Fig. 11.3), within which more than 95% of the soil measurements were made. Our primary

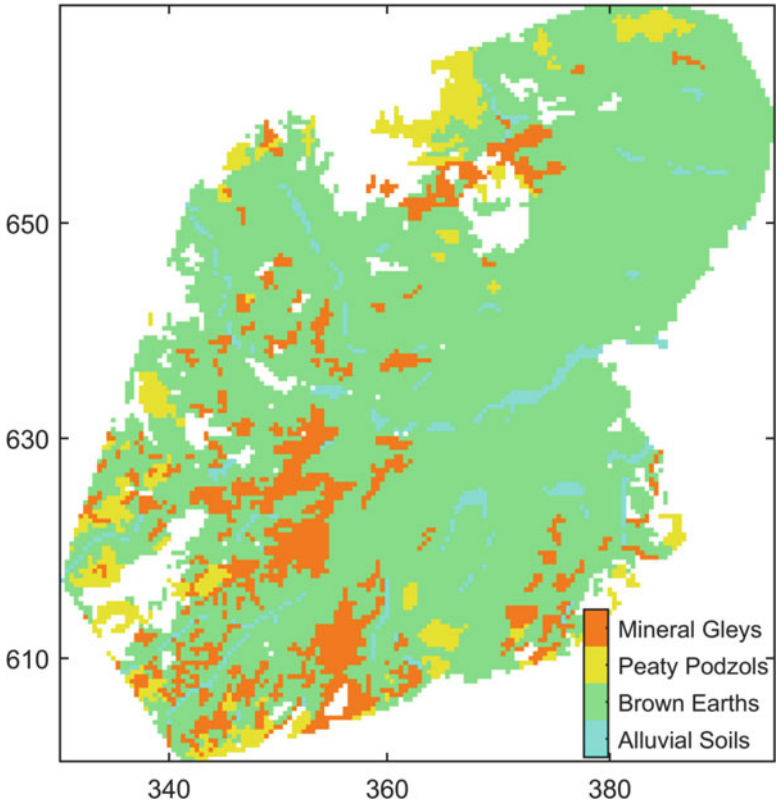


Fig. 11.3 Map of the soil types within the study region according to the 1:250,000 National Soil Map of Scotland (Used with permission from the James Hutton Institute, © Crown copyright). Coordinates are km from the origin of the British National Grid

objective is to use these data to map the concentration of copper and the probability that it is less than 1.0 mg kg^{-1}. We explore the extent to which the soil type information and the observed cobalt concentrations can explain the variation of copper concentrations and hence improve the accuracy of our maps. We validate our maps using the remaining 2481 observations of copper. We explore the implications of using fewer observations by repeating our analyses on a 50 sample subset of the 400 copper measurements (Fig. 11.2a). At the two sites marked by green crosses in Fig. 11.2a, we describe the prediction of the copper concentration in more detail. Site 'A' is 0.6 km from the nearest observation, whereas site 'B' is 1.3 km from an observation.

11.3 Maximum Likelihood Estimation

11.3.1 The Linear Mixed Model

We denote an observation of the soil property, in our case the concentration of copper, at location x_i by $s_i = s(x_i)$ and the set of observations by $\mathbf{s} = \{s_1, s_2, \ldots, s_n\}^{\mathrm{T}}$ where T denotes the transpose of the vector. We assume that $\mathbf{s}$ is a realization of an LMM:

$$\mathbf{s} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{11.1}$$

Here, the $\mathbf{M}\boldsymbol{\beta}$ are the fixed effects and the $\boldsymbol{\varepsilon}$ are the random effects. The design matrix $\mathbf{M}$ is of size $n \times q$ where q is the number of covariates that are included in the random effects model. The vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)^{\mathrm{T}}$ contains the fixed effects parameters or regression coefficients. Each column of $\mathbf{M}$ contains the value of a covariate at the n observed sites. If the random effects include a constant term, then each entry of the corresponding column of $\mathbf{M}$ is equal to one. A column of $\mathbf{M}$ could consist of the values of a continuous covariate such as elevation or the output from a remote sensor (e.g. Rawlins et al. 2009). Thus, the fixed effects include a term that is proportional to the covariate. If the fixed effects differ according to a categorical covariate such as soil type, then the columns of $\mathbf{M}$ include c binary covariates which indicate the presence or absence of each of the c classes at each site. Note that if the c classes account for all of the observations, then a constant term in the fixed effects would be redundant.

The vector $\boldsymbol{\varepsilon}$ contains the values of the random effects at each of the n sites. The elements of $\boldsymbol{\varepsilon}$ are realized from a zero-mean random function with a specified distribution function. The $n \times n$ covariance matrix of the random effects is denoted $\mathbf{C}(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ is a vector of covariance function parameters. The elements of this matrix can be determined from any of the authorized and bounded variogram functions described in Chap. 10 since

$$C(h) = \gamma(\infty) - \gamma(h), \tag{11.2}$$

where $C(h)$ is a covariance function for lag h, $\gamma(h)$ is a bounded and second-order stationary variogram and $\gamma(\infty)$ is the upper bound or total sill of the variogram. We will focus on the nested nugget and Matérn variogram (Minasny and McBratney 2005) so that

$$C(h) = \begin{cases} c_0 + c_1 \text{ if } h = 0 \\ c_1 G(h) \text{ for } h > 0' \end{cases} \tag{11.3}$$

where

$$G(h) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}h}{a} \right)^{\nu} K_{\nu} \left(\frac{2\sqrt{\nu}h}{a} \right). \tag{11.4}$$

Here, Γ is the gamma function and K_ν is a modified Bessel function of the second kind of order ν. The random effects model parameters are c_0 the nugget, c_1 the partial sill, a the distance parameter and ν the smoothness parameter. Jointly, there are four random effect parameters, i.e. $\boldsymbol{\alpha} = (c_0, c_1, a, \nu)^{\mathrm{T}}$.

11.3.2 Coregionalized Soil Properties

In Sect. 10.4, we saw that it was possible to extend the classical geostatistical model of a single soil property to consider the spatial correlation between observations of two or more properties by using a linear model of coregionalization (LMCR). Marchant and Lark (2007a) demonstrated how the LMM could also be extended to include coregionalized variables. The LMCR consists of a variogram for each soil property and a series of cross-variograms describing the spatial correlation between each pair of properties. Each variogram or cross-variogram must have the same variogram structure. This means they are based on the same authorized models, and they have common spatial and, in the case of the Matérn model, smoothness parameters. The nugget and sill parameters can differ for the different soil properties. We denote these parameters by $c_0^{d,e}$ and $c_1^{d,e}$, respectively, where the d and e refer to the different soil properties. So, if $d = e$, these are parameters of a variogram, whereas if $d \neq e$, they are parameters of a cross-variogram. There are further constraints to ensure that the LMCR leads to positive definite covariance matrices. If we define matrices $\mathbf{B}_0$ and $\mathbf{B}_1$ by

$$\mathbf{B}_0 = \begin{bmatrix} c_0^{1,1} & \cdots & c_0^{1,v} \\ \vdots & \ddots & \vdots \\ c_0^{v,1} & \cdots & c_0^{v,v} \end{bmatrix} \text{ and } \mathbf{B}_1 = \begin{bmatrix} c_1^{1,1} & \cdots & c_1^{1,v} \\ \vdots & \ddots & \vdots \\ c_1^{v,1} & \cdots & c_1^{v,v} \end{bmatrix} \tag{11.5}$$

where v is the number of soil properties; then, both of these matrices must have a positive determinant.

The LMCR can be incorporated into the LMM (Eq. 11.1) by altering the random effects covariance matrix to accommodate the different variograms and cross-variograms. In this circumstance, the observation vector $\mathbf{s}$ will include observations of each of the v soil properties. It is likely that the random effects matrix $\mathbf{M}$ will require sufficient columns to accommodate different fixed effects for each soil property. If each variogram and cross-variogram consists of a nested nugget and Matérn model, then element i, j of the random effects covariance matrix would be:

$$C_{i,j} = \begin{cases} c_0^{v_i,v_j} + c_1^{v_i,v_j} & \text{if } h_{i,j} = 0 \\ c_1^{v_i,v_j} G\left(h_{i,j}\right) & \text{if } h_{i,j} > 0 \end{cases}, \tag{11.6}$$

where the ith element of $\mathbf{s}$ is an observation of property v_i, the jth element of $\mathbf{s}$ is an observation of property v_j and $h_{i,j}$ is the lag separating the two observations.

In common with the LMCR from classical geostatistics, the cross-variogram nuggets and sills must be constrained to ensure that the determinants of $\mathbf{B}_0$ and $\mathbf{B}_1$ are positive. For $v_i \neq v_j$, the $c_0^{v_i,v_j}$ parameter only influences the covariance function if there are some locations where both v_i and v_j are observed. If this is not the case, then the parameter cannot be fitted. If $\nu = 2$, and both properties are observed at some sites, then $\boldsymbol{\alpha} = \left(c_0^{1,1}, c_0^{1,2}, c_0^{2,2}, c_1^{1,1}, c_1^{1,2}, c_1^{2,2}, a, \nu\right)^{\mathrm{T}}$.

11.3.3 The Likelihood Function

If the distribution function of the random effects is assumed to be multivariate Gaussian, then the log of the likelihood function is equal to

$$\ln\, l\,(\mathbf{s}|\boldsymbol{\alpha},\boldsymbol{\beta}) = L\,(\mathbf{s}|\boldsymbol{\alpha},\boldsymbol{\beta}) = \text{Constant} - \frac{1}{2}\ln|\mathbf{C}\,(\boldsymbol{\alpha})\,| - \frac{1}{2}(\mathbf{s}-\mathbf{M}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\,(\mathbf{s}-\mathbf{M}\boldsymbol{\beta})\,, \tag{11.7}$$

where $|\ |$ denotes the determinant of a matrix. The log-likelihood is the objective function which we use to test the suitability of an LMM to represent the spatial variation of a soil property.

The assumption of Gaussian random effects is restrictive and often implausible for soil properties. For example, the histograms of observed copper and cobalt concentrations in the Scottish Borders region are highly skewed (Fig. 11.4a, b). A transformation $s^* = H(s)$ can be applied to the data so they more closely approximate a Gaussian distribution. Figure 11.4c, d show the more symmetric histograms that result when the natural log-transform, $s^* = \ln(s)$, is applied to the copper and cobalt concentrations. The Box-Cox transform might also be applied to skewed data. It generalizes the natural log-transform via a parameter λ:

$$H(s) = \begin{cases} \ln(s) & \text{if } \lambda = 0, \\ \frac{s^{\lambda}-1}{\lambda} & \text{otherwise.} \end{cases} \tag{11.8}$$

If we assume that the observed data are multivariate Gaussian after the application of a transformation, then the formula for the log-likelihood becomes

$$\begin{aligned} L\,(\mathbf{s}\,|\boldsymbol{\alpha},\boldsymbol{\beta}\,) = \text{Constant} - \frac{1}{2}\ln|\mathbf{C}\,(\boldsymbol{\alpha})| - \frac{1}{2}\left(\mathbf{s}^*-\mathbf{M}\boldsymbol{\beta}\right)^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\left(\mathbf{s}^*-\mathbf{M}\boldsymbol{\beta}\right) \\ + \sum_{i=1}^{n}\ln\left(J\,\{s_i\}\right) \end{aligned} \tag{11.9}$$

where $J\,\{s_i\} = \frac{\mathrm{dH}}{\mathrm{ds}}$ is the derivative of the transformation function evaluated at $s = s_i$. For the Box-Cox transform,

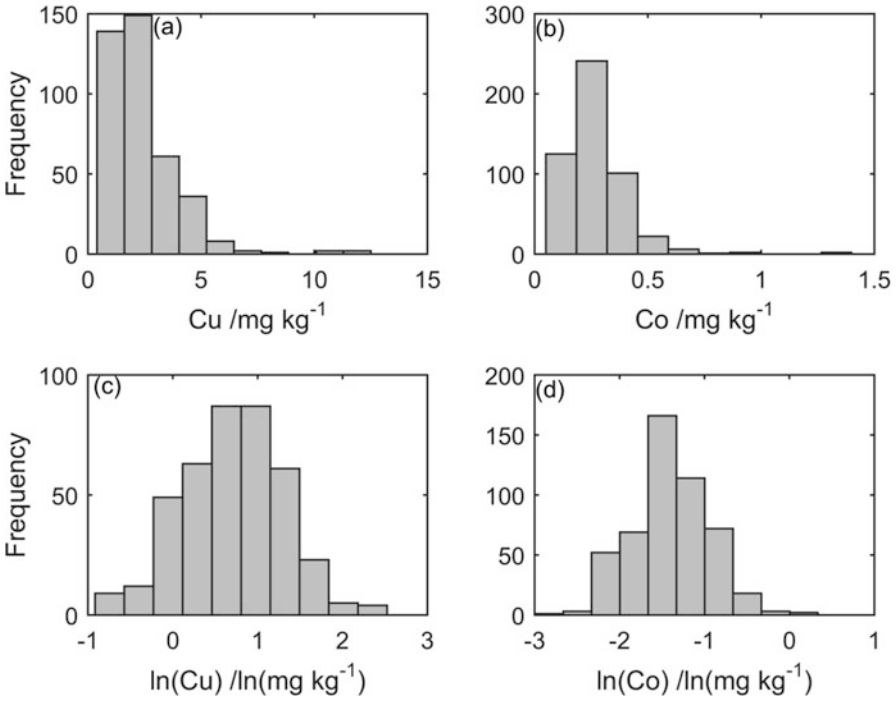


Fig. 11.4 Histograms of the observations of (**a**) 400 soil copper concentrations, (**b**) 500 soil cobalt concentrations, (**c**) the natural logarithm of the copper concentrations and (**d**) the natural logarithm of the cobalt concentrations

$$\ln\left(J\left\{s_i\right\}\right) = (\lambda - 1)\ln\left(s_i\right), \tag{11.10}$$

and the corresponding function for the natural log-transform is found by setting $\lambda = 0$.

Once a fixed-effect design matrix, a covariance function and any transformation have been proposed, the problem of maximum likelihood estimation is reduced to finding the elements of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ vectors which lead to the largest value of the log-likelihood and hence of the likelihood. This can be achieved using a numerical optimization algorithm to search the parameter space for these parameter values. In the examples presented in this chapter, we use the standard MATLAB optimization algorithm which is the Nelder-Mead method (Nelder and Mead 1965). This is a deterministic optimizer in the sense that if it is run twice from the same starting point, the same solution will result each time. It is prone to identifying local rather than global maxima. Therefore, Lark et al. (2006) recommends the use of stochastic optimizers such as simulated annealing which permit the solution to jump away from a local maximum. In our implementation of the Nelder-Mead method, we run

the algorithm from ten different starting points in an attempt to avoid the selection of local maxima. The parameter space is constrained to ensure that negative variance parameters or negative definite covariance matrices cannot result.

The optimization problem can be simplified by noting that the log-likelihood function is maximized when

$$\boldsymbol{\beta} = \left(\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{s}. \tag{11.11}$$

Hence, there is only a need to search the $\boldsymbol{\alpha}$ parameter space, and the corresponding optimal value of $\boldsymbol{\beta}$ can be found by using the above formula. However, it is still a computational challenge to find an optimal $\boldsymbol{\alpha}$ vector within a four- or higher-dimensional parameter space. Diggle and Ribeiro (2007) suggest reducing the problem further to a series of optimizations in a lower-dimensional parameter space. They search for the optimal values of c_0, c_1 and a for a series of fixed v values. Then they plot the maximum log-likelihood achieved (or equivalently the minimum negative log-likelihood) against the fixed v and extract best of these estimates. This plot is referred to as a profile-likelihood plot.

It is possible to fit LMMs of varying degrees of complexity by adding more terms to the fixed and random effects. If we add an extra term to a model (e.g. an extra column to the $\mathbf{M}$ matrix), then the maximized likelihood of the more complex model will be at least as large as the likelihood of the simpler model. We need a test to decide whether the improvement that is achieved by adding extra terms is worthwhile. If an LMM is too complex, there is a danger of overfitting. This means that the model is too well suited to the intricacies of the calibration data, but it performs poorly on validation data that were not used in the fitting process.

If two LMMs are nested their suitability to represent the observed data can be compared by using a likelihood ratio test (Lark 2009). By nested, we mean that by placing constraints on its parameters it is possible to transform the more complex model to the simpler model. For example, if one LMM included a Box-Cox transform of the data and another model was identical except that the natural log-transform was applied, then these two models would be nested. The more complex model could be transformed to the simpler model by setting $\lambda = 0$. We denote the parameters of the complex model by $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$ and the parameters of the simpler model by $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$. Under the null hypothesis that the additional parameters in the more complex model do not improve the fit, the test statistic:

$$D = 2L\left(\mathbf{s}|\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1\right) - 2L\left(\mathbf{s}|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\right) \tag{11.12}$$

will be asymptotically distributed as a chi-squared distribution with r degrees of freedom. Here, r is the number of additional parameters in the more complex model. Therefore, it is possible to conduct a formal statistical test to decide whether the more complex model has a sufficiently larger likelihood than the simpler one. However, the likelihood ratio test does not always meet our needs. An LMM with a nugget and exponential variogram and an identical model except that the variogram

is pure nugget would not be properly nested. This is because the more complex model has two additional parameters, c_1 and a, but only the first of these needs to be constrained to $c_1 = 0$ to yield the simpler model. Therefore, it is unclear what the degrees of freedom should be in the formal test. Lark (2009) used simulation approaches to explore this issue.

In this chapter, we calculate the AIC (Akaike 1973) for each estimated model:

$$\mathrm{AIC} = 2k - 2L, \tag{11.13}$$

where k is the number of parameters in the model. The preferred model is the one with the smallest AIC value. This model is thought to be the best compromise between quality of fit (i.e. the likelihood) and complexity (the number of parameters). The AIC does not require the different models to be nested.

11.3.4 The Residual Maximum Likelihood Estimator

Patterson and Thompson (1971) observed that there was a bias in variance parameters estimated by maximum likelihood. This bias occurs because the $\boldsymbol{\beta}$ parameters are estimated from the data and are therefore uncertain, whereas in the log-likelihood formula, they are treated as if they are known exactly. This problem is well known when considering the variance of independent observations. The standard formula to estimate the variance of a population with unknown mean that has been sampled at random is:

$$\mathrm{Var}\,(\mathrm{S}) = \frac{1}{n-1}\sum_{i=1}^{n}(s_i - \bar{s})^2 \tag{11.14}$$

where n is the size of the sample and $\bar{s}$ is the sample mean. Since the variance is defined as $\mathrm{E}[\{S - \mathrm{E}(S)\}^2]$, one might initially be surprised that the denominator of Eq. 11.14 is $n - 1$ rather than n. However, it can be easily shown that if the denominator is replaced by n because the mean of the population is estimated, the expectation of the expression would be a factor of $n/(n-1)$ times the population variance.

Patterson and Thompson (1971) devised a method for correcting the analogous bias in the maximum likelihood estimates of random effects parameters. They transformed the data into stationary increments prior to calculating the likelihood. The likelihood of these increments was independent of the fixed effects and hence the bias did not occur. The expression for the residual log-likelihood that resulted is:

$$L_R\,(\mathbf{s}|\boldsymbol{\alpha}) = \mathrm{Constant} + \frac{1}{2}\ln|\mathbf{W}| + \frac{1}{2}\ln|\mathbf{C}\,(\boldsymbol{\alpha})| + \frac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{Q}\mathbf{s}, \tag{11.15}$$

where $\mathbf{W} = \mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{M}$ and $\mathbf{Q} = \mathbf{I} - \mathbf{M}\mathbf{W}^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}$.

11.3.5 Estimating Linear Mixed Models for the Scottish Borders Data

We illustrate maximum likelihood estimation of an LMM using the 400 copper concentration measurements from the Scottish Borders. We initially assume that the fixed effects are constant and that a log-transform is sufficient to normalize the data. Thus, we assume that the Box-Cox parameter λ is zero. In fact, when the Box-Cox parameter was unconstrained, the increase to the log-likelihood was negligible and did not improve the AIC. We estimated the c_0, c_1 and a parameters for fixed ν equal to 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0 and 2.5. The profile-likelihood plot of the minimized negative log-likelihoods that resulted is shown in Fig. 11.5a. The smallest negative log-likelihood occurred for $\nu = 0.1$. When ν was unfixed, a smaller negative log-likelihood resulted with $\nu = 0.12$. The best fitting variogram is plotted in Fig. 11.5b. It sharply increases as the lag is increased from zero, reflecting the small value of the smoothness parameter. This maximum likelihood estimate of the variogram is reasonably consistent with the empirical variogram (see Fig. 11.5b).

Figure 11.6 shows the maximum likelihood estimates of the variograms and cross-variograms when the **s** vector contained the 400 observations of copper and the 500 observations of cobalt. The natural log-transform was applied to each soil property, and the fixed effects consisted of a different constant for each property. Again, there is reasonable agreement between the maximum likelihood estimate and the empirical variograms. However, some small discrepancies are evident. The maximum likelihood estimate for cobalt has a longer range than might be

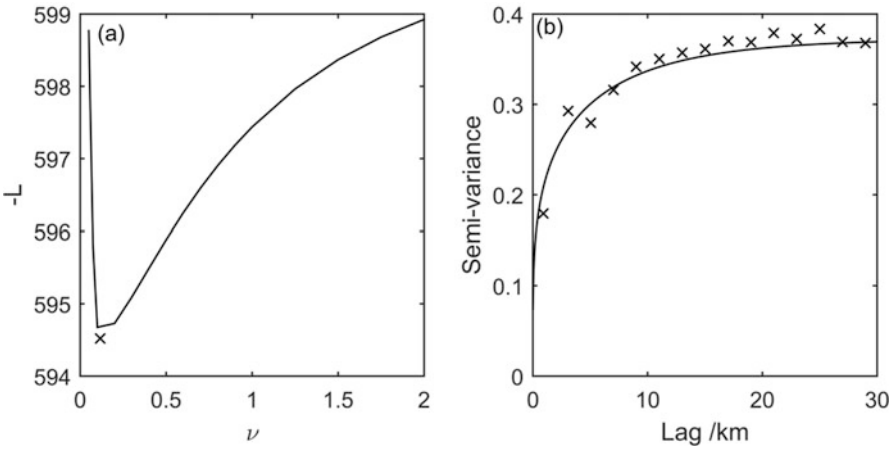


Fig. 11.5 (**a**) Plot of minimized negative log-likelihood of 400 ln copper concentrations for different fixed values of ν. The *black* cross denotes the minimized negative log-likelihood when ν is unfixed, (**b**) maximum likelihood estimate of the variogram (*continuous line*) of ln copper concentrations and the corresponding method of moments point estimates (*black crosses*)

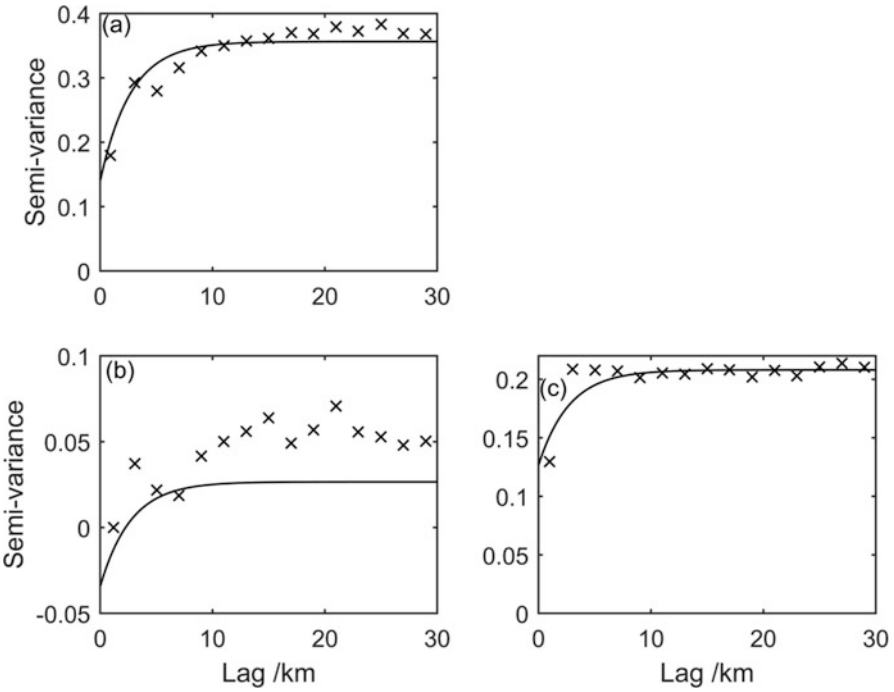


Fig. 11.6 Maximum likelihood estimate of (**a**) the LMCR ln copper auto-variogram, (**b**) the LMCR ln copper and ln cobalt cross-variogram and (**c**) the LMCR ln cobalt auto-variogram. The *black crosses* are the corresponding method of moments point estimates

fitted to the empirical variogram, and the maximum likelihood estimate of the cross-variogram appears to be consistently less than the corresponding empirical variogram. This was probably caused by the constraints placed on the parameters, such as the requirement that all the variograms had the same range.

In Table 11.1, we show the negative log-likelihood and AIC values that result when different LMMs are estimated for the Scottish Borders data. The simplest models only consider copper observations and assume constant fixed effects and a pure nugget variogram. Then, the pure nugget variogram is replaced by a nugget and Matérn model. The third model also replaces the constant fixed effects by ones that vary according to the soil types displayed in Fig. 11.3. The final model considers observations of both copper and cobalt and assumes that each of these properties has constant fixed effects. The models are estimated for both the 50 and 400 observation samples of copper. Both coregionalized models include 500 cobalt observations.

For both sample sizes, the negative log-likelihood decreases upon inclusion of the Matérn variogram function and the nonconstant fixed effects. In the case of the 400 observation samples, the AIC also decreases in the same manner. However, for the 50-point sample, the addition of these extra terms to the model causes the AIC to increase. This indicates that there is insufficient evidence in the 50-point sample

Table 11.1 Number of parameters (p), minimized negative log-likelihood ($-L$) and AIC value for models with different fixed effects (**M**), variogram models, number of copper observations (n_{Cu}) and number of cobalt observations (n_{Co})

				$n_{Cu} = 50$		$n_{Cu} = 400$	
M	Variogram	n_{Co}	p	$-L$	AIC	$-L$	AIC
Constant	Nugget	0	2	−82.25	168.50	−640.26	1284.53
Constant	Matérn	0	5	−81.36	172.72	−594.52	1199.04
Soil type	Matérn	0	8	−78.20	172.39	−590.68	1197.37
Constant	Matérn	500	9	−34.38	86.76	−225.26	468.52

to indicate that copper is spatially correlated or that it varies according to soil type. For both sample sizes, the lowest AIC occurs when the 500 cobalt observations are included in the model. Note that the models of coregionalized variables were estimated by minimizing the negative log-likelihood function which included both properties. However, the negative log-likelihood that is quoted in Table 11.1 is the likelihood of the copper observations given both the estimated parameters and the cobalt observations. This means that the corresponding AIC value is comparable to those from the other three models.

11.4 Bayesian Methods and Variogram Parameter Uncertainty

The model-based methods described in this chapter are compatible with Bayesian methodologies which can be used to quantify the uncertainty of the random effects parameters (Handcock and Stein 1993). Classical statistical methodologies assume that model parameters are fixed. Generally, when applying classical or model-based geostatistics, we look for a single best fitting variogram model and take no account of variogram uncertainty. In Bayesian analyses, model parameters are treated as probabilistic variables. Our knowledge of the parameter values prior to collecting any data is expressed as a prior distribution. Then the observations of the soil property are used to update these priors and to form a posterior distribution which combines our prior knowledge with the information that could be inferred from the observations.

Minasny et al. (2011) demonstrated how Bayesian approaches could be applied to the spatial prediction of soil properties. They placed uniform priors on all of the random effects parameters and then used a Markov chain Monte Carlo (MCMC) simulation approach to sample the multivariate posterior distribution of these parameters. Rather than a single best fitting estimate, this approach led to a series of parameter vectors that were consistent with both the prior distributions and the observed data. The MCMC approach generates a chain of parameter vectors which follow a random walk through the parameter space. The chain starts at some

value of parameters, and the log-likelihood is calculated. Then the parameters are perturbed and the log-likelihood is recalculated. The new parameter values are accepted or rejected based on the difference between the likelihoods before and after the perturbation according to Metropolis-Hastings algorithm (Hastings 1970). If the log-likelihood increases, then the new parameter vector is always accepted. If the likelihood decreases, then the proposal might be accepted. The probability of acceptance decreases as the difference in likelihood increases.

Under some regularity conditions, the set of parameter vectors that result from the Metropolis-Hastings algorithm are known to converge to the posterior distribution of the parameters. However, the algorithm requires careful tuning of some internal settings within the algorithm. These settings particularly relate to the distribution from which a proposed parameter vector is sampled. If this distribution is too wide, then too many of the proposed parameter vectors will be rejected and the chain will remain at its starting point. If the proposal distribution is too narrow, then nearly all of the parameter vectors will be accepted, but the perturbations will be small, and it will take a considerable amount of time to consider the entire parameter space. Therefore Vrugt et al. (2009) developed a DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm to automate the selection of these internal settings and to produce Markov chains that converge efficiently. The DREAM algorithm simultaneously generates multiple Markov chains. The information inferred from acceptances and rejections within each chain is pooled to select efficient proposal distributions. MATLAB and R implementations of the DREAM algorithm are freely available (Vrugt 2016; Guillaume and Andrews 2012).

We used the DREAM algorithm to sample parameters of the nested nugget and Matérn model for both the 50 and 400 observations of ln copper. In each case, we used four chains and sampled a total of 101,000 parameter vectors. The bounds on the uniform prior distributions were zero and one $\{\ln(\text{mg kg}^{-1})\}^2$ for c_0 and c_1, zero and 40 km for a and 0.01 and 2.5 for ν. For comparison, the variance of the ln copper observations was 0.35. The first 1000 of the sampled vectors were discarded since the MCMC was converging to the portion of parameter space that was consistent with the observed data. This is referred to as the burn-in period. We used the R statistic (see Vrugt et al. 2009) to confirm that the chain had converged. Successive entries of the series of parameter vectors that remained were correlated because the parameter vector was either unchanged or only perturbed a short distance. Therefore every 100th entry of this series was retained. The final series contained 1000 vectors which were treated as independent samples from the posterior distribution of the parameter vector. Figure 11.7 shows the 90% confidence intervals for the variogram of each dataset. These confidence intervals stretch between the 5th and 95th percentiles of the semi-variances for each lag. It is evident that the variogram from the 50 observation sample is uncertain across all lag distances. The uncertainty is greatly reduced for the 400 observation sample.

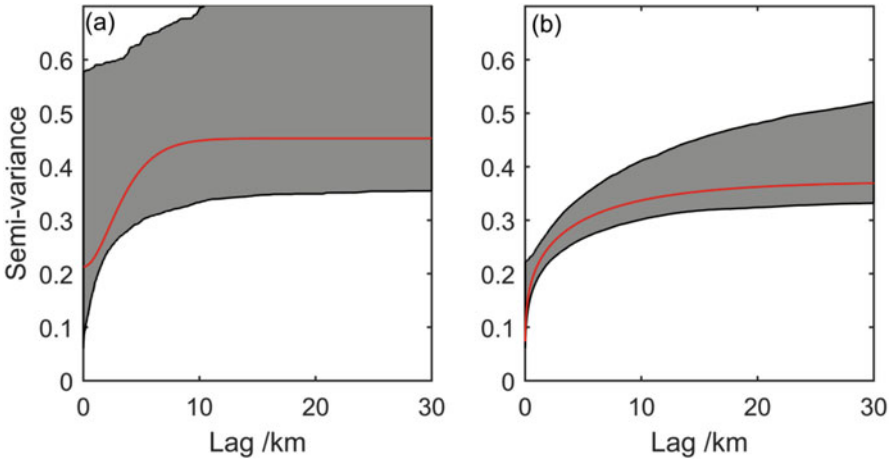


Fig. 11.7 Maximum likelihood estimates of the variogram for (**a**) 50 ln copper concentrations and (**b**) 400 ln Cu concentrations. The *grey* shading indicates the 90% confidence interval for the variogram function according to the MCMC sample of variogram parameters

11.5 Spatial Prediction and Validation of Linear Mixed Models

11.5.1 The Best Linear Unbiased Predictor

Having estimated the fixed and random effects parameters $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, we can use the LMM and the observations of the soil property to predict the expected value and uncertainty of the possibly transformed soil property at a set of locations $\mathbf{x}_\mathrm{p}$ where it has not been observed. We denote the fixed effects design matrix at these locations by $\mathbf{M}_\mathrm{p}$, the matrix of covariances between the random effects of the soil property at the observation and prediction locations by $\mathbf{C}_\mathrm{po}$ and the random effects covariance matrix at the prediction locations by $\mathbf{C}_\mathrm{pp}$. These matrices are calculated using the estimated $\widehat{\boldsymbol{\alpha}}$ parameters. The best linear unbiased predictor (BLUP; Lark et al. 2006; Minasny and McBratney 2007) of the expected value of the possibly transformed soil property at the unobserved locations is:

$$\widehat{\mathbf{S}}_\mathrm{p}^{*} = \mathrm{E}\left[\mathbf{s}^{*}\left(\mathbf{x}_\mathrm{p}\right)\right] = \mathbf{M}_\mathrm{p}\widehat{\boldsymbol{\beta}} + \mathbf{C}_\mathrm{po}\mathbf{C}^{-1}\left(\mathbf{s}^{*} - \mathbf{M}\widehat{\boldsymbol{\beta}}\right), \tag{11.16}$$

and the corresponding prediction covariance matrix is:

$$\mathbf{V} = \left(\mathbf{M}_\mathrm{p} - \mathbf{C}_\mathrm{po}\mathbf{C}^{-1}\mathbf{M}\right)\left(\mathbf{M}^\mathrm{T}\mathbf{C}^{-1}\mathbf{M}\right)^{-1}\left(\mathbf{M}_\mathrm{p} - \mathbf{C}_\mathrm{po}\mathbf{C}^{-1}\mathbf{M}\right)^\mathrm{T} + \left(\mathbf{C}_\mathrm{pp} - \mathbf{C}_\mathrm{po}\mathbf{C}^{-1}\mathbf{C}_\mathrm{po}^\mathrm{T}\right). \tag{11.17}$$

The first term in Eq. 11.17 accounts for the uncertainty in predicting the fixed effects, whereas the second term accounts for the uncertainty in predicting the random effects. The elements of the main diagonal of $\mathbf{V}$, which we denote V_{ii}, are the total prediction variances for each site. Since we have assumed that $\mathbf{s}^*$ is Gaussian, we have sufficient information (i.e. the mean and the variance) to calculate the probability density function (pdf) or cumulative density function (cdf) for $\mathbf{s}^*$ at each of the sites. Density functions are discussed in more detail in Sect. 14.2.2. We might calculate the pdf for N equally spaced values of the variable with spacing Δy (i.e. $y_j = j\Delta y$ for $j = 1, \ldots, N$). The formula for the Gaussian pdf is:

$$f_j \equiv f\left(y_j | \widehat{\mathrm{S}}_i^*, V_{ii}\right) = \frac{1}{\sqrt{2\pi V_{ii}}} \exp\left\{-\frac{1}{2}\left(\frac{y_j - \widehat{\mathrm{S}}_i^*}{\sqrt{V_{ii}}}\right)^2\right\} \tag{11.18}$$

If the density is zero (to numerical precision) for $y_i < y_i$ and $y_i > y_N$, then the area under the curve f will be one and the f_j will sum to $1/\Delta y$. The cdf can then be deduced from the f_j:

$$F_j \equiv \mathrm{Prob}\left(s_i < y_j\right) = \Delta y \sum_{k=1}^{j} f_k, \tag{11.19}$$

In Fig. 11.8, we show these pdf and cdf for ln copper concentration at site 'A' based on the maximum likelihood estimate of the LMM for 400 copper observations. The area of the grey-shaded region is equal to the probability that ln copper concentration is negative (i.e. that the concentration of copper is less than 1 mg kg^{-1}). This probability can be more easily extracted from the value of the cdf when ln copper is equal to zero (Fig. 11.8b).

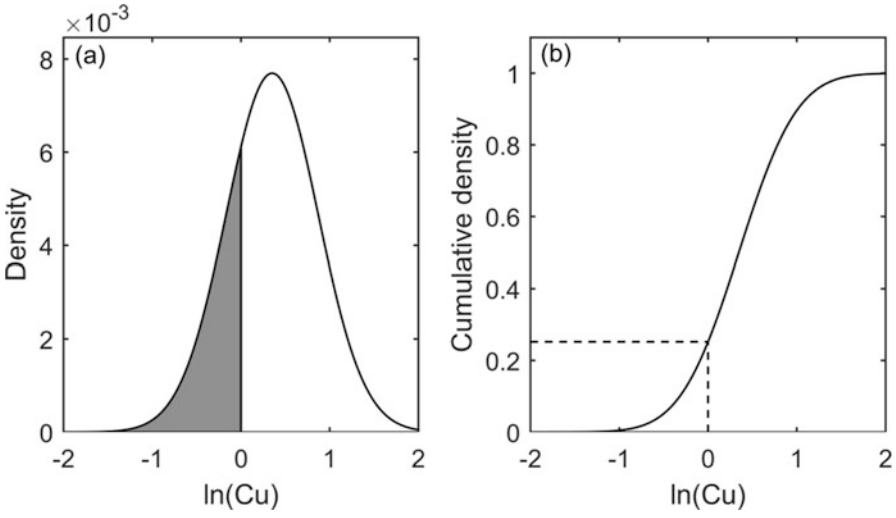


Fig. 11.8 Predicted (**a**) pdf and (**b**) cdf for ln copper at location 'A'

If we have a MCMC sample of variogram parameter vectors rather than a single estimate, then we might calculate the pdf for each of these variograms. Then we could calculate a pdf that accounted for variogram uncertainty by averaging these individual pdfs and calculate the cdf using Eq. 11.19. This is an example of the Monte Carlo error propagation method described in Sect. 14.4.2. In Fig. 11.9b, we show the cdf of ln copper at site 'A' based on the 400 observation sample. The grey-shaded region is the 90% confidence interval for this cdf using the MCMC sample of variogram parameters to account for variogram uncertainty. It is apparent that variogram uncertainty does not have a large effect on the cdf. However, when the pdfs are based on the MCMC for 50 observations, a larger effect of variogram uncertainty is evident (Fig. 11.9a). Recall that site 'A' is only 0.6 km from the nearest observation. We will see in Sect. 11.6 of this chapter that such a prediction is sensitive to uncertainty in the variogram parameters. When we repeat the exercise at site 'B' which is 1.3 km from an observation, the effects of variogram uncertainty are small using the MCMC samples based on both 50 and 400 observations.

The BLUP encompasses many of the kriging algorithms described in Chap. 10. For example, when the fixed effects are constant, it performs the role of the ordinary kriging estimator; when covariates are included in the fixed effects, it performs the role of the regression or universal kriging predictor; and when multiple soil properties are included in the observation vector, it performs the role of the co-kriging estimator. Equations 11.16 and 11.17 lead to predictions on the same support as each observation. If we wish to predict the soil property across a block that has a larger support than each observation, then $\mathbf{C}_{\text{po}}$ and $\mathbf{C}_{\text{pp}}$ should be replaced by $\overline{\mathbf{C}}_{\text{po}}$, the covariances between the observations and the block averages, and $\overline{\mathbf{C}}_{\text{pp}}$ the covariances between the block averages.

We previously noted that when multiple properties are included in the observation vector of an LMM, the nugget parameters for the cross-variograms can only be estimated if there are co-located observations of the two properties. This parameter will be required in the BLUP if we wish to predict one soil property at the exact location where another one was observed or if we wish to know the covariance between the predictions of the two properties at the same site. However, this parameter is not required to produce maps of each soil property on a regular grid that does intersect the observation locations or to calculate the prediction variances for each property.

If a transformation has been applied to the observations then it will be necessary to back-transform the predictions before they can be interpreted. If we simply calculate the inverse of the transformation for $\widehat{S}^{*}_{\text{p}(i)}$, the prediction of the mean of the transformed property at the ith prediction location, the result is the prediction of the median of the untransformed property. The 0.5 quantile of the transformed property has been converted to the 0.5 quantile of the untransformed property. It is possible to back-transform every quantile of the cdf in this manner. It is generally more difficult to determine the mean of a back-transformed prediction. Instead, the back-transformed mean can be approximated through simulation of the transformed variable. If the mean and variance of the transformed prediction for a

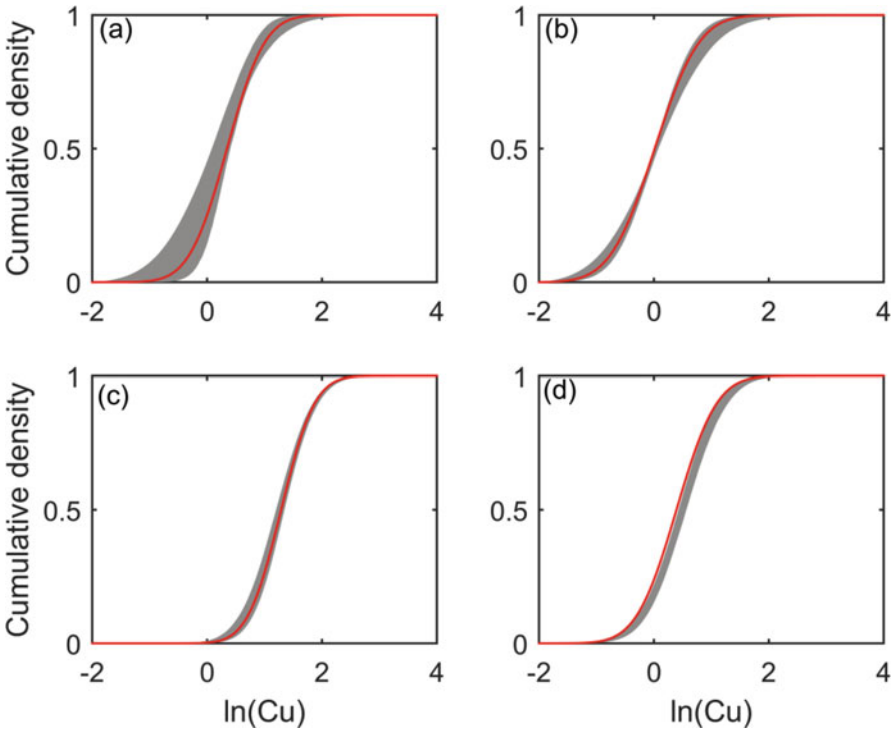


Fig. 11.9 Predicted cdf (*red line*) and 90% confidence interval accounting for variogram uncertainty (*shaded area*) of ln copper concentration at (**a**) site 'A' conditional on the 400 observation samples and the MCMC sample of variogram parameters calibrated on the 50 observation subsamples, (**b**) site 'A' conditional on the 400 observation sample and the MCMC sample of variogram parameters calibrated on the same 400 observations, (**c**) site 'B' conditional on the 400 observation sample and the MCMC sample of variogram parameters calibrated on the 50 observation subsample and (**d**) site 'A' conditional on the 400 observation sample and the MCMC sample of variogram parameters calibrated on the same 400 observations

site are $\widehat{S}^*_{p(i)}$ and $V^*_{p(i)}$, respectively, then one might simulate 1000 realizations of the Gaussian random variable with this mean and variance, apply the inverse transform to each realization and then calculate the mean (or other statistics) of these back-transformed predictions.

11.5.2 Validation of the LMM

It is important to validate an LMM to confirm that the predictions are as accurate as we believe them to be. Close inspection of validation results might also reveal patterns in the model errors that indicate that an additional covariate is required in

the fixed effects or a further generalization is required in the random effects. Ideally, validation should be conducted using a set of data that were not used to calibrate the LMM. However, in some instances, data are sparse, and then there is little choice but to carry out cross-validation. In leave-one-out cross-validation, the model is fitted to all of the measurements, and then one datum, s_i say, is removed, and the remaining data and the BLUP are used to predict the removed observation. The process is repeated for all n observations.

When conducting either cross-validation or validation, we wish to look at both the accuracy of the predictions and the appropriateness of the prediction variances. We can assess the accuracy of predictions by looking at quantities such as the mean error (ME),

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}\left\{s_i - \widehat{S}_i\right\}, \tag{11.20}$$

and the root mean squared error (RMSE),

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left\{s_i - \widehat{S}_i\right\}^2\right]^{\frac{1}{2}}. \tag{11.21}$$

The appropriateness of the prediction variances are often explored by calculating the standardized squared prediction errors at each site:

$$\theta_i = \frac{\left\{s_i - \widehat{S}_i\right\}^2}{V_i}, \tag{11.22}$$

where V_i is the prediction variance for $\widehat{S}_i$. If, as we expect, the errors follow a Gaussian distribution, then the θ_i will be realized from a chi-squared distribution with one degree of freedom. Pedometricians often calculate the mean $\overline{\boldsymbol{\theta}}$ and median $\tilde{\boldsymbol{\theta}}$ of the θ_i and compare them to their expected values of 1.0 and 0.45 (e.g. Minasny and McBratney 2007; Marchant et al. 2009). If they are properly applied, the ML and REML estimators tend to ensure that $\overline{\boldsymbol{\theta}}$ is close to 1.0. However, although the average standardized error is close to its expected values, the set of θ_i values might not be consistent with the chi-squared distribution. Deviations from this distribution are often indicated by values of $\tilde{\boldsymbol{\theta}}$ that are far from 0.45. If $\tilde{\boldsymbol{\theta}}$ is considerably less than 0.45, this might indicate that the LMM should have a more highly skewed distribution function.

It can also be useful to consider the entire cdf of the standardized errors. If the errors are Gaussian, then the

$$p_i = \Phi_{0,1}^{-1}\left(\frac{s_i - \widehat{S}_i}{\sqrt{V_i}}\right), \tag{11.23}$$

should be uniformly distributed between zero and one where $\Phi_{0,1}^{-1}$ is the inverse cdf for a Gaussian distribution of zero mean and unit variance. We want to confirm that the p_i is consistent with such a uniform distribution. This can be achieved through the inspection of predictive QQ plots (Thyer et al. 2009). These are plots of the n theoretical quantiles of a uniform distribution against the sorted p_i values. If the standardized errors are distributed as we expect, then the QQ plot should be a straight line between the origin and (1,1). If all the points lie above (or, alternatively, below) the 1:1 line, then the soil property is consistently under-/over-predicted. If the points lie below (above) the 1:1 line for small theoretical values of p and above (below) the 1:1 line for large theoretical values of p, then the predictive uncertainty of the MM is under-/overestimated. Alternatively, Goovaerts (2001) uses accuracy plots rather than QQ plots. In these plots, the [0,1] interval is divided into a series of bins bounded by the $(1-p)/2$ and $(1+p)/2$ quantiles for p between 0 and 1. The two plots differ in that each bin of the accuracy plot is symmetric about 0.5. Therefore it is not possible to consider the upper and lower tail of the distribution separately. In contrast, with QQ plots, we can see how well the left-hand tail of the distribution is approximated by looking close to the origin of the plot, and we can examine how the right-hand tail is approximated by looking close to (1,1).

11.5.3 Predicting Copper Concentrations in the Scottish Borders

Figure 11.10 shows maps of the expectation of ln copper across the study region for models listed in Table 11.1. The model with a pure nugget variogram is not included since the predictions are constant. The map based on 50 copper observations and with constant fixed effects is much less variable than the other predictions. More features of the copper variation are evident when the soil type is added to the fixed effects. However, there are obvious and possibly unrealistic discontinuities in the predictions at the boundaries between different soil types. The model including cobalt observations retains the detail of the variable fixed-effect model, but there are no discontinuities. Further detail is added to all of the maps when 400 rather than 50 copper observations are used. The hotspots of copper that are evident occur close to urban centres.

The most striking feature in the validation results (Tables 11.2 and 11.3) is the difference in performance for the models based on 50 observations of copper and those based on 400 observations. When 400 observations are used, the cross-validation and validation RMSEs are very similar and smaller than those for the 50-point sample. For the 50-observation sample, there is also a greater difference between the cross-validation and validation results indicating that these models might well have been overfitted. For both sample sizes, the largest RMSEs occur for the pure nugget model, but there is little difference in the RMSEs for the other three LMMs. The prediction variances for the LMMs calibrated on 400 observations

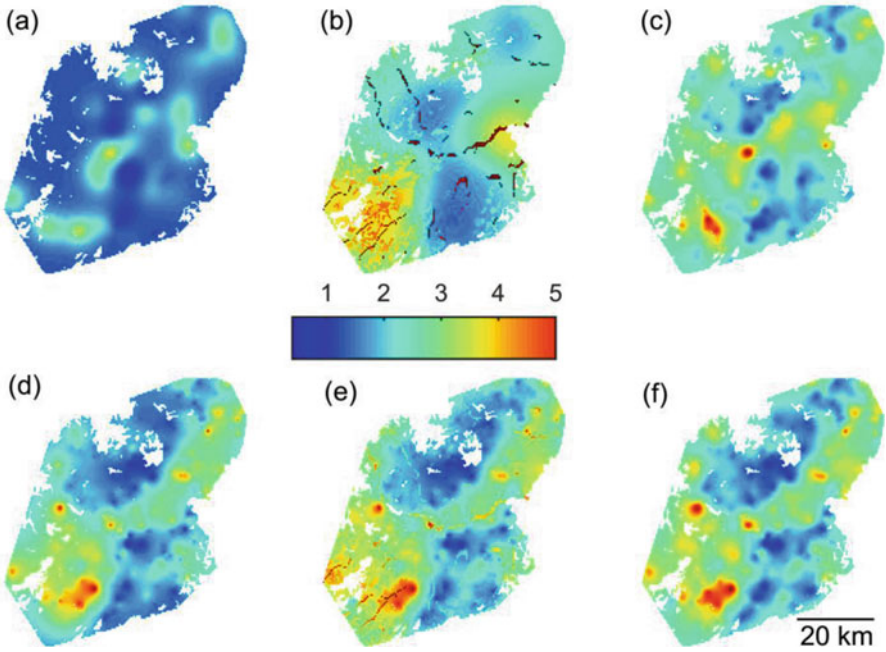


Fig. 11.10 Median prediction of copper concentration (mg/kg) across the study region from (**a**) BLUP using 50 Cu calibration data with constant fixed effects, (**b**) BLUP using 50 Cu calibration data with fixed effects varying according to soil type, (**c**) BLUP using 50 Cu calibration data and 500 Co calibration data represented by an LMCR with constant fixed effects, (**d**) BLUP using 400 Cu calibration data with constant fixed effects, (**e**) BLUP using 400 Cu calibration data with fixed effects varying according to soil type and (**f**) BLUP using 400 Cu observations and 500 Co observations represented by an LMCR with constant fixed effects

also appear to be more accurate than those based on the 50 observation sample. The mean standardized prediction errors upon cross-validation and validation for all models calibrated on 400 copper observations are between 0.98 and 1.00, and the median standardized prediction errors range between 0.38 and 0.47. Hence both quantities are close to their expected values of 1.0 and 0.45. In the case of the 50 observation samples, the ranges of these statistics are wider, stretching between 0.78 and 1.34 for the mean and 0.32 and 0.67 for the median. The cross-validation standardized prediction errors are closer to their expected values than the corresponding validation values. Again, this is an indication of overfitting when only using 50 observations. There is also evidence of overfitting in the QQ plot for the LMM with constant fixed effects and calibrated on 50 observations (Fig. 11.11a). The validation plot deviates a large distance from the 1:1 line. All of the other QQ plots more closely follow the 1:1 line although some improvement in using the larger sample size is evident.

Table 11.2 Mean error (ME) and root mean squared error (RMSE) upon leave-one-out cross-validation and validation of models for ln copper with different fixed effects (**M**), variogram models, number of copper observations (n_{Cu}) and number of cobalt observations (n_{Co})

				Calibration		Validation	
M	Variogram	n_{Cu}	n_{Co}	ME	RMSE	ME	RMSE
Constant	Nugget	50	0	0.000	0.68	0.092	0.60
Constant	Matérn	50	0	0.009	0.66	0.045	0.74
Soil type	Matérn	50	0	0.011	0.64	0.023	0.58
Constant	Matérn	50	500	−0.001	0.65	0.048	0.54
Constant	Nugget	400	0	0.000	0.60	0.028	0.59
Constant	Matérn	400	0	0.002	0.50	0.063	0.51
Soil type	Matérn	400	0	0.002	0.50	0.011	0.50
Constant	Matérn	400	500	0.002	0.50	0.013	0.50

Table 11.3 Mean standardized prediction errors $(\overline{\theta})$, median standardized prediction errors $(\widetilde{\theta})$ and mean error in QQ plot $(\overline{Q_e})$ upon leave-one-out cross-validation and validation of models for ln copper with different fixed effects (**M**), variogram models, number of copper observations (n_{Cu}) and number of cobalt observations (n_{Co})

				Calibration			Validation		
M	Variogram	n_{Cu}	n_{Co}	$(\overline{\theta})$	$(\widetilde{\theta})$	$\overline{Q_e}$	$(\overline{\theta})$	$(\widetilde{\theta})$	$\overline{Q_e}$
Constant	Nugget	50	0	1.02	0.61	0.000	0.80	0.38	0.040
Constant	Matérn	50	0	1.03	0.49	−0.007	1.34	0.67	0.197
Soil type	Matérn	50	0	1.08	0.62	−0.022	0.90	0.38	0.012
Constant	Matérn	50	500	1.03	0.33	−0.003	0.78	0.32	0.019
Constant	Nugget	400	0	1.00	0.47	−0.002	0.99	0.45	0.012
Constant	Matérn	400	0	1.00	0.42	0.003	0.99	0.39	0.028
Soil type	Matérn	400	0	1.00	0.41	0.002	1.00	0.38	−0.001
Constant	Matérn	400	500	1.00	0.44	0.003	0.98	0.39	0.000

11.6 Optimal Sample Design

It is often costly to obtain soil samples from a study area and then analyse them in the laboratory to determine the properties of interest such as the concentrations of cobalt and copper in the Scottish Borders survey. Therefore, it can be important to optimize the locations where the samples are extracted from so that an adequate spatial model or map can be produced for the minimum cost. Many spatial surveys have been conducted using a regular grid design (see examples in Webster and Oliver 2007) since this ensures that the observations are fairly evenly distributed across the study region and that the kriging variances are not unnecessarily large at any particular locations. Also, it is relatively easy to apply the method of moments variogram estimator to a grid-based sample since the variogram lag bins can be selected according to the grid spacing. Some authors (e.g. Cattle et al. 2002) have

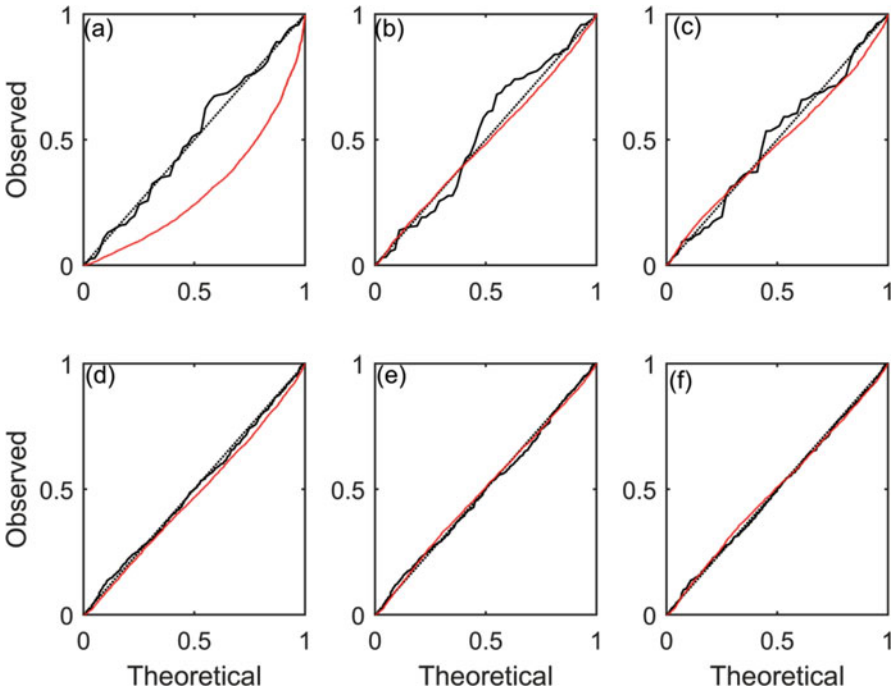


Fig. 11.11 QQ plots resulting from leave-one-out cross-validation of the calibration copper data (*black line*) and validation of the validation copper data (*red line*). The predictors and observations for plots (**a**–**f**) are identical to those in Fig. 11.10

included close pairs of observations in their survey designs since these ensure that the variogram can be reliably estimated for short lag distances. However, the decision as to how much of the sampling effort should be allocated to even coverage of the study region and how much to estimating the spatial model over short distances is often made in a subjective manner.

The kriging or prediction variance (Eq. 11.17) can form the basis of a more objective criterion for the efficient design of spatial surveys. If the random effects are second-order stationary and Gaussian, then the prediction variance does not depend on the observed data. If the variogram model is assumed to be known, then prior to making any measurements, it is possible to use Eq. 11.17 to assess how effective a survey with a specified design will be. Alternatively, we can select the configuration of a specified number of sampling locations that lead to the smallest prediction variance.

Van Groenigen et al. (1999) suggested an optimization algorithm known as spatial simulated annealing (SSA) to perform this task. The n samples are initially positioned randomly across the study region and the corresponding kriging variance (or any other suitable objective function) is calculated. Then the position of one of these samples is perturbed a random distance in a random direction. The kriging

variance is recalculated. If it has decreased, then the perturbation is accepted. If the kriging variance has increased, then the perturbation might be accepted. In common with the DREAM algorithm, described in Sect. 11.4 of this chapter, the probability of acceptance decreases with the magnitude of the increase in the objective function according to the Metropolis-Hastings algorithm (Hastings 1970). However, in contrast to the DREAM algorithm, the probability of acceptances also decreases as the optimization algorithm proceeds. The potential to accept perturbations that increase the kriging variance is included to ensure that the optimizer does not converge to a local rather than global minimum. The gradual decrease in the probability of such an increase ensures that a minimum is eventually reached. If a perturbation is rejected, then the sample is returned to its previous location. The SSA algorithm continues, perturbing each point in turn until the objective function settles to a minimum.

We illustrate the application of the SSA algorithm by optimizing 50 location sample schemes for the Scottish Borders study area in Fig. 11.12. In all of the plots within this figure, we consider a soil property where the random effects are realized from a Gaussian random function with a nested nugget and Matérn covariance function with $c_0 = 0.5$, $c_1 = 0.5$, $a = 10$ km and $\nu = 0.5$. The kriging variance is the objective function for the designs shown in Fig. 11.12a, b. In plot (a), the fixed effects are assumed to be constant. The optimized sample locations are spread evenly across the region. The locations are randomly allocated to the four different soil types (Fig. 11.12d) with the majority of samples being situated in the most prevalent brown earth class. When the fixed effects are assumed to vary according to soil type, the optimized samples are still fairly evenly distributed across the region. However, the number of locations that are situated in the less prevalent soil classes increases (Fig 11.12e). This ensures that a reasonably accurate estimate of the fixed effects can be calculated for each soil type. When Brus and Heuvelink (2007) optimized sample schemes for universal kriging of a soil property with an underlying trend that was proportional to a continuous covariate, they found that the soil was more likely to be sampled at sites where this covariate was particularly large or particularly small. This ensured that the gradient of the trend function could be reasonably accurately estimated.

In addition to spatial prediction, the set of observed samples should also be suitable for estimating the spatial model. Therefore, Marchant and Lark (2007b) and Zhu and Stein (2006) expanded the objective function to account for uncertainty in estimating the spatial model or variogram. These authors noted that for a linear predictor such as the BLUP,

$$\widehat{S}_{\mathrm{p}(k)} = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{s}, \tag{11.24}$$

where $\widehat{S}_{\mathrm{p}(k)}$ is the prediction of the expectation of S at the ith prediction site, $\boldsymbol{\lambda}$ is a length n vector of weights and $\mathbf{s}$ is a length n vector of observations; the extra contribution to the prediction variance resulting from variogram uncertainty could be approximated by a Taylor series:

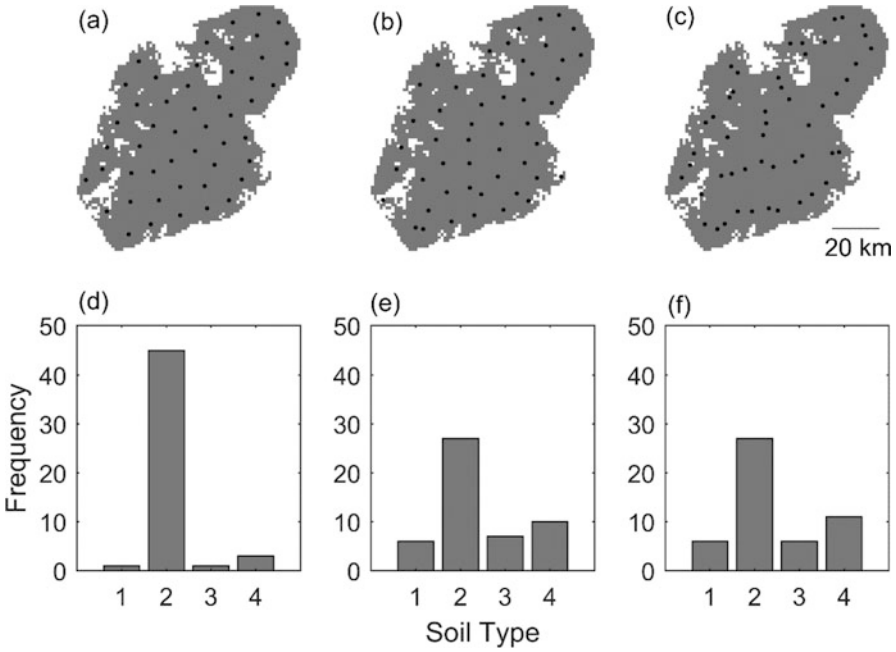


Fig. 11.12 Optimized sampling locations where the objective function is (**a**) the kriging variance with no fixed effects, (**b**) the kriging variance with fixed effects varying according to soil type and (**c**) the kriging variance plus the prediction variance due to variogram uncertainty with fixed effects varying according to soil type. Plots (**d**–**f**) show the distribution of sampling locations amongst soil types for the design above. The soil types are (1) alluvial soils, (2) brown earths, (3) peaty podzols and (4) mineral gleys

$$\tau^2 = \mathrm{E}\left[\left\{\widehat{S}_{\mathrm{p}(k)}\left(\widehat{\boldsymbol{\alpha}}\right) - \widehat{S}_{\mathrm{p}(k)}\left(\boldsymbol{\alpha}\right)\right\}^2\right] = \sum_{i=1}^{r}\sum_{j=1}^{r}\mathrm{Cov}\left(\alpha_i, \alpha_j\right)\frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial \alpha_i}\mathbf{C}\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_j}, \tag{11.25}$$

where τ^2 is the expected squared difference between the prediction at the ith site using the actual variogram parameters $\boldsymbol{\alpha}$, and the prediction at this site using the estimated parameters, $\widehat{\boldsymbol{\alpha}}$; r is the number of variogram parameters and $\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_i}$ is the length n vector containing partial derivatives of the $\boldsymbol{\lambda}$ weights matrix with respect to the ith variogram parameter. The covariances between the variogram parameters can be approximated using the Fisher information matrix $\mathbf{F}$ (Marchant and Lark 2004):

$$\mathrm{Cov}\left(\alpha_i, \alpha_j\right) \approx \left[\mathbf{F}^{-1}\left(\boldsymbol{\alpha}\right)\right]_{ij};\ \left[\mathbf{F}\left(\boldsymbol{\alpha}\right)\right]_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha_j}\right] \tag{11.26}$$

Here, $[]_{ij}$ denotes element i, j of the matrix inside the brackets, $\frac{\partial \mathbf{C}}{\partial \alpha_i}$ is the $n \times n$ matrix of partial derivatives of the covariance matrix $\mathbf{C}$ with respect to α_i and Tr

denotes the trace or sum of elements on the main diagonal of the matrix that follows. The $\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_i}$ can be calculated using a numerical approximation (e.g. Zhu and Stein 2006) although Marchant and Lark (2007b) noted the standard linear algebra relationship that if

$$\mathbf{L} = \mathbf{A}^{-1}\mathbf{b}, \tag{11.27}$$

then

$$\frac{\partial \mathbf{L}}{\partial \alpha_i} = \mathbf{A}^{-1}\left(\frac{\partial \mathbf{b}}{\partial \alpha_i} - \frac{\partial \mathbf{A}}{\partial \alpha_i}\mathbf{A}^{-1}\mathbf{b}\right). \tag{11.28}$$

Thus, this equation can be used to exactly calculate the $\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_i}$ matrices for the universal kriging predictor formulation of the BLUP which is written in the form of Eq. 11.27 with:

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\varphi} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{M} \\ \mathbf{M}^{\mathrm{T}} & \mathbf{0}_{q,q} \end{bmatrix} \text{and } \mathbf{b} = \begin{bmatrix} \mathbf{C}^{\mathrm{T}}_{\mathrm{p}(k)\mathbf{o}} \\ \mathbf{M}_{\mathrm{p}(k)} \end{bmatrix} \tag{11.29}$$

where φ is the length q vector of Lagrange multipliers, $\mathbf{0}_{q,q}$ is a $q \times q$ matrix of zeros and $\mathbf{C}_{\mathrm{p}(i)\mathrm{o}}$ is the ith row of matrix $\mathbf{C}_{\mathrm{po}}$. For most authorized covariance functions, the elements of the $\frac{\partial \mathbf{C}}{\partial \alpha_i}$ that are required to calculate Eq. 11.26 and Eq. 11.28 can be determined exactly. However, in the case of the Matérn function, numerical differentiation remains the most practical method to calculate $\frac{\partial \mathbf{C}}{\partial \nu}$.

Figure 11.12c shows an optimized 50-location design that results when the objective function is the prediction variance (assuming that the fixed effects vary according to soil type) plus the τ^2. In this case, the sampling locations are less evenly spread across the study region. Short transects of close locations are evident which are suitable for estimating the spatial covariance function over small lags. The prediction variance and τ^2 for this design are mapped in Fig. 11.13a, b, respectively. The prediction variance is smallest close to sampling locations and increases for locations where there are no nearby samples and that are at the boundary of the region. The additional component of uncertainty because of the estimation of the spatial model is largest close to sample locations and decreases at locations where there are no nearby samples.

As previously discussed in Sect. 10.2, there is one obvious flaw in this strategy to optimize sample designs. In reality, the spatial model is not known prior to sampling, and therefore, it is not possible to calculate the prediction variance. A spatial model must be assumed, perhaps using the results of previous surveys of the soil property at similar locations or based on a reconnaissance survey (e.g. Marchant and Lark 2006). Alternatively, a prior distribution might be assumed for each covariance function parameter and the objective function averaged across these distributions (Diggle and Lophaven 2006). Also, the prediction variance cannot be calculated

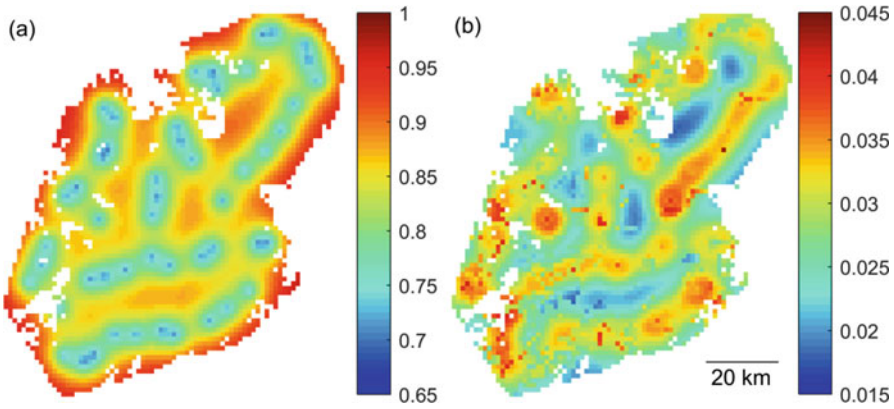


Fig. 11.13 (**a**) The kriging variance with fixed effects varying according to soil type for the optimized sample design in Fig. 11.12c. (**b**) The contribution to the prediction variance due to variogram uncertainty for the same sample design

prior to sampling if the property of interest has a skewed distribution, since the prediction variances vary according to the observed data. Marchant et al. (2013a) demonstrated how simulation could be used to optimise a multi-phase survey in these circumstances.

11.7 Conclusions

The use of model-based rather than classical geostatistical methods removes many of the subjective decisions that are required when performing geostatistical analyses. The LMM is flexible enough to incorporate linear relationships between the soil property of interest and available covariates and to simultaneously represent the spatial variation of multiple coregionalized soil properties. The log-likelihood is an objective function that can be used to compare proposed model structures and parameters. Since a multivariate distribution function is specified in the model, it is possible to predict the complete pdf or cdf of the soil property at an unsampled location. From these predicted functions, one can easily determine the probability that the soil property exceeds a critical threshold (e.g. Li et al. 2016).

The primary disadvantages of the model-based methods are the time required to compute the log-likelihood and the requirement to specify the multivariate distribution function from which the soil observations were realized. Maximum likelihood estimation of an LMM for more than 1000 observations is likely to take several hours rather than the seconds or minutes required to estimate a variogram by the method of moments. However, we have seen that the uncertainty in LMM parameters estimated from a 400 observation sample have little effect on the uncertainty of the final predictions. Therefore, when estimating an LMM for a large number of observations, it is reasonable to subsample the data. The complete dataset

should be used when using the BLUP. Stein et al. (2004) suggested an approximate maximum likelihood estimator. This uses all of the data but is faster to compute because it ignores some of the covariances between observations. We have described how the standard assumptions that the random effects of an LMM are realized from a multivariate Gaussian random function can be relaxed. A Box-Cox or natural log-transform can be applied to skewed observations so that their histogram more closely approximates that of a Gaussian distribution. Alternatively, it is possible to write the log-likelihood function in a different form that is compatible with any marginal distribution function (Marchant et al. 2011). The assumption that the expected value of the random variable is constant can be relaxed via the fixed effects of the LMM. Some authors have also explored strategies to permit the variability of the random variable to be related to a covariate (e.g. Lark 2009; Marchant et al. 2009; Haskard et al. 2010).

The model-based methods described in this chapter are also compatible with the modelling of space-time variation (e.g. Heuvelink and van Egmond 2010). However, space-time models do require more flexible models for the covariance function because the pattern of temporal variation is likely to be quite distinct from the temporal variation. De Cesare et al. (2001) review the covariance functions that are commonly used for this purpose. Such models might also be used to represent the spatial variation of soil properties in three dimensions when the vertical variation is quite different to the horizontal variation (e.g. Li et al. 2016). The same models could be estimated by classical methods. However, these models contain multiple variogram structures, and it is challenging to appropriately select the different lag bins and fitting weights required for each of these.

Acknowledgements This chapter is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council). We are grateful to the James Hutton Institute for giving permission to use data extracted from the 1:250,000 National Soil Map of Scotland. Copyright and database rights of the 1:250,000 National Soil Map of Scotland are owned by the James Hutton Institute (October, 2013). All rights reserved. Any public sector information contained in the data is licensed under the Open Government Licence v.2.0. We thank the former East of Scotland College of Agriculture, especially Mr. H.M. Gray, Dr. R.G. McLaren and Dr. R.B. Speirs, for the data from their survey of the Scottish Borders region.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akadémiai Kiadó, Budapest, pp 267–281

Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138:86–95

Cattle JA, McBratney AB, Minasny B (2002) Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. J Environ Qual 31:1576–1588

De Cesare L, Myers DE, Posa D (2001) Estimating and modelling space-time correlation structures. Stat Probab Lett 51:9–14

Diggle PJ, Lophaven S (2006) Bayesian geostatistical design. Scand J Stat 33:53–64

Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics. Springer, New York
Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. Geoderma 103:3–26
Goovaerts P, Webster R (1994) Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur J Soil Sci 45:79–95
Guillaume J, Andrews F (2012) Dream: DiffeRential evolution adaptive metropolis. R package version 0.4–2. http://CRAN.R-project.org/package=dream
Handcock MS, Stein ML (1993) A Bayesian-analysis of kriging. Technometrics 35:403–410
Haskard KA, Welham SJ, Lark RM (2010) Spectral tempering to model non-stationary covariance of nitrous oxide emissions from soil using continuous or categorical explanatory variables at a landscape scale. Geoderma 159:358–370
Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109
Heuvelink GBM, van Egmond FM (2010) Space-time geostatistics for precision agriculture: a case study of NDVI mapping for a Dutch potato field. In: Oliver MA (ed) Geostatistical applications for precision agriculture. Springer, Dordrecht, pp 117–137
Lark RM (2009) Kriging a soil variable with a simple non-stationary variance model. J Agric Biol Environ Stat 14:301–321
Lark RM, Cullis BR, Welham SJ (2006) On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Eur J Soil Sci 57:787–799
Li HY, Marchant BP, Webster R (2016) Modelling the electrical conductivity of soil in the Yangtze delta in three dimensions. Geoderma 269:119–125
Marchant BP, Lark RM (2004) Estimating variogram uncertainty. Math Geol 36:867–898
Marchant BP, Lark RM (2006) Adaptive sampling and reconnaissance surveys for geostatistical mapping of the soil. Eur J Soil Sci 57:831–845
Marchant BP, Lark RM (2007a) Estimation of linear models of coregionalization by residual maximum likelihood. Eur J Soil Sci 58:1506–1513
Marchant BP, Lark RM (2007b) Optimized sample schemes for geostatistical surveys. Math Geol 39:113–134
Marchant BP, Newman S, Corstanje R, Reddy KR, Osborne TZ, Lark RM (2009) Spatial monitoring of a non-stationary soil property: phosphorus in a Florida water conservation area. Eur J Soil Sci 60:757–769
Marchant BP, Saby NPA, Jolivet CC, Arrouays D, Lark RM (2011) Spatial prediction of soil properties with copulas. Geoderma 162:327–334
Marchant BP, McBratney AB, Lark RM, Minasny B (2013a) Optimized multi-phase sampling for soil remediation surveys. Spat Stat 4:1–13
Marchant BP, Rossel V, Webster R (2013b) Fluctuations in method-of-moments variograms caused by clustered sampling and their elimination by declustering and residual maximum likelihood estimation. Eur J Soil Sci 64:401–409
Mardia KV, Marshall RJ (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 72:135–146
McBratney AB, Webster R (1986) Choosing functions for semivariances of soil properties and fitting them to sample estimates. J Soil Sci 37:617–639
McBratney AB, Webster R, McLaren RG, Spiers RB (1982) Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland. Agronomie 2:969–982
Minasny B, McBratney AB (2005) The Matérn function as a general model for soil variograms. Geoderma 128:192–207
Minasny B, McBratney AB (2007) Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140:324–336
Minasny B, Vrugt JA, McBratney AB (2011) Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation. Geoderma 163:150–162
Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30:683–691
Rawlins BG, Marchant BP, Smyth D, Scheib C, Lark RM, Jordan C (2009) Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. Eur J Soil Sci 60:44–54
Ribeiro Jr PJ, Diggle PJ (2001) GeoR: a package for geostatistical analysis. R-News 1(2), ISSN 1609-3631
Soil Survey of Scotland Staff (1984) Organisation and methods of the 1:250 000 soil survey of Scotland, Handbook 8. The Macaulay Institute for Soil Research, Aberdeen
Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York
Stein ML, Chi Z, Welty LJ (2004) Approximating likelihoods for large spatial data sets. J R Stat Soc B 66:275–296
Thyer M, Renard B, Kavetski D, Kuczera G, Franks S, Srikanthan S (2009) Critical evaluation of parameter consistency and predictive modelling in hydrological modelling: a case study using Bayesian total error analysis. Water Resour Res 45:W00B14
Van Groenigen JW, Siderius W, Stein A (1999) Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87:239–259
Vrugt JA (2016) Markov chain Monte Carlo simulation using DREAM software package: theory, concepts, and Matlab implementation. Environ Model Softw 75:273–316
Vrugt JA, ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D (2009) Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int J Nonlinear Sci Numer Simul 10:273–290
Webster R (1994) The development of pedometrics. Geoderma 62:1–15
Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester
Zhu Z, Stein ML (2006) Spatial sampling design for prediction with estimated parameters. J Agric Biol Environ Stat 11:24–44

Chapter 12
Digital Mapping of Soil Classes and Continuous Soil Properties

Brendan P. Malone, Nathan P. Odgers, Uta Stockmann, Budiman Minasny, and Alex. B. McBratney

"Soils have shape and area, breadth and width, as well as depth".

Charles E. Kellog, 1949

12.1 General Soil Mapping Concepts

12.1.1 Conventional Soil Mapping

Soil is often described as mantling the land more or less continuously with the exception being where there is bare rock and ice (Webster and Oliver 2006). Our understanding of soil variation in any region is usually based on only a small number of observations made in the field. Across the spatial domain of the region of interest, predictions of the spatial distribution of soil properties are made at unobserved locations based on the properties of the small number of soil observations. There are two principal approaches for making predictions of soil at unobserved locations. The first approach subdivides the soil coverage into discrete spatial units within which the soils conform to the characteristics of a class in some soil classification (Heuvelink and Webster 2001). The second approach treats soils as a suite of continuous variables and attempts to describe the way these variables vary across

B.P. Malone (✉) • U. Stockmann • B. Minasny • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: brendan.malone@sydney.edu.au; uta.stockmann@sydney.edu.au;
budiman.minasny@sydney.edu.au; alex.mcbratney@sydney.edu.au

N.P. Odgers
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia

Soils and Landscapes Team, Manaaki Whenua – Landcare Research, PO Box 69040,
Lincoln 7640, New Zealand
e-mail: OdgersN@landcareresearch.co.nz

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_12

the landscape (Heuvelink and Webster 2001). The second approach is necessarily quantitative, as it requires numerical methods for interpolation between the locations of actual soil observations.

Because of its roots in geological survey and biological taxonomy, virtually all of the national soil survey programmes carried out across the world during the twentieth century mapped the distribution of soils using the first approach. Researchers including Simonson (1989), Arnold (2006) and Hewitt et al. (2008) have written exhaustively about the techniques and procedures required to make a soil map with this approach. In the first step, a soil surveyor makes a detailed observation of the physical landscape in order to relate characteristics of the soil with landscape features. Often ancillary information, including aerial photos, geology, vegetation and topographic maps, are useful aids in this process. In so doing the surveyor develops a detailed conceptual model that can be used to infer soil characteristics at a given site in the landscape. This conceptual model relies heavily on the tacit knowledge of the soil surveyor since the depth of his soil-landscape knowledge is dependent on the degree of his skill and prior experience (Hudson 1992). The third step involves applying this conceptual model across the survey area to predict at unobserved locations. Generally less than 0.001% of the survey region is actually observed (Burrough et al. 1971). The conceptual model is transferred to a cartographic model—usually on an aerial photo base—by delineating areas of the landscape that have the same soil-landscape relationships. These areas belong to map units, which are comprised of assemblages of soil classes belonging to some classification system.

The maps that result from this process are commonly known as choropleth or area-class maps because each delineation, or polygon, in the map belongs to one and only one category (Wright 1944). The category is usually a soil map unit. The mapping assumes, albeit implicitly, that soil variation within delineations is unimportant and that abrupt changes occur at the polygon boundaries (Burrough 1989; Heuvelink and Huisman 2000). Burrough et al. (1997) labelled this kind of mapping double crisp because the identified soil groups are supposed to be crisply delineated in both taxonomic space (the space defined by the soil properties) and geographic space. Unfortunately, this soil mapping concept is at odds with the continuous nature of soil variability and has been questioned by numerous researchers (e.g. Webster and Cuanalo de la Cerda 1975; Nortcliff 1978; Nettleton et al. 1991). The creation of conventional soil maps draws upon the experience and skills of the soil surveyor and the tacit knowledge that is acquired and nuanced over a number of years. A criticism levelled at this heuristic approach to soil mapping is that the learned concepts are generally unfalsifiable and therefore unverifiable in any objective sense (Hewitt 1993; Lagacherie et al. 1995).

Because of the limitations inherent in soil maps made using traditional approaches, their application to the resolution of sophisticated and quantitative problems is limited. For example, much legacy soil information is qualitative in nature. The cartographic scale often limits the application of legacy mapping to resolve particular questions (Hartemink et al. 2008). The cost and time required to compile a soil map using traditional methods is expensive, so it is usually prohibitive

to compile specific maps to suit specific purposes. Finally, while polygon mapping can be rasterised in order to be integrated into modern raster-based analytical procedures, the inherent scale mismatch between the relatively coarse-detailed soil polygons and the relatively fine-detailed terrain or remotely sensed rasters means that the fidelity of information that traditional soil mapping brings to quantitative analyses is often relatively low.

12.1.2 Quantitative Representation of Soil Variability

Electronic computers were introduced to soil science in the early 1960s, and this made the application of intensive mathematical operations on soil data a practical proposition for the first time (Heuvelink and Webster 2001).

12.1.2.1 Fuzzy Sets

A major advance was the introduction of a logic that enabled a relaxation of the traditional double-crisp model of soil variation (Burrough et al. 1997). This logic, called fuzzy logic, was first introduced by Zadeh (1965) as part of his fuzzy set theory. Fuzzy logic is a generalisation of traditional Boolean logic and enables the expression of truth on a continuous scale (e.g. 'degree of truth') rather than a binary scale ('true' or 'false'). In the same way, fuzzy sets are generalisations of crisp sets. In both cases an individual, x, has a membership, μ, in a set or class $c = 1,2,\ldots k$. In a crisp set, membership is binary, and μ is a discrete variable. Hence x is a member of class c if its membership, μ, is 1 only; i.e. $\mu_c(x) = 1$. Conversely it is not a member of class c if $\mu_c(x) = 0$. Since the sum of memberships across all classes is 1, i.e. $\sum_{c=1}^{k} \mu_c(x) = 1$, it implies that in a crisp system an individual can only be a member of one and only one class at once.

On the other hand, fuzzy logic allows for a partial degree membership in any class; μ is a continuous variable. In this case $\mu_c(x) = [0, 1]$, and since the memberships sum to 1, it implies that x can have partial membership in several classes simultaneously.

McBratney et al. (1992), McBratney and de Gruijter (1992), Odeh et al. (1992), Lagacherie et al. (1997) and Grunwald et al. (2001) provide comprehensive discussions of the application of fuzzy sets in soil science. Fuzzy sets have found particular application in soil classification (see Chap. 8) but have also been applied in land evaluation, modelling and simulation of soil processes, fuzzy soil geostatistics, soil quality indices and fuzzy measurement of imprecisely defined soil phenomena. Though useful, fuzzy set theory has not been widely adopted for soil mapping (Grunwald and Lamsal 2006). One reason is possibly the difficulty in interpreting the outputs. For example, instead of one map of a given soil attribute, there are potentially many for a single region of interest.

12.1.2.2 Spatial Modelling

Soil classification systems and the practice of mapping soils have developed side by side since soil survey programmes began in the early and mid-twentieth century across the world (Brown 2006). The language instilled in soil classifications and descriptions thereof has been suitable for communicating soil information among people with knowledge of soils. Communicating soil information to non-soil science-minded people (but who have an interest and need for soil information) has proven difficult because (1) of abstract technical jargon and (2) they have a preference for soil information that describes the variability of soil properties across a survey region (Sanchez et al. 2009). Mapping the spatial variability of soil properties requires the use of geospatial prediction models as well as incorporating gridded or raster-based data models to make predictions onto and display maps (thus a trend away from the polygon data model).

The basic premise behind all digital soil mapping is that given a set of samples of the soil (classes or attributes) and a suitable model we can make predictions of the soil at unsampled locations. When we refer to soil classes we often mean those that are part of an established soil classification system, although we can make spatial predictions of bespoke classes too (e.g. Odgers et al. 2011; Triantafilis et al. 2001). When we refer to soil attributes we usually mean measurable properties like its pH, electrical conductivity or particle size fractions among many others. It is possible to distinguish between purely spatial models and environmental correlation models that relate observations to other characteristics of the landscape (McBratney et al. 2003). Purely spatial approaches appeared first because rasters of environmental characteristics did not become widely and readily available until the late 1970s.

12.1.2.3 Purely Spatial Approaches

The first purely spatial modelling approaches applied to soil attribute prediction were based on trend surface analysis. In simple terms, trend surfaces are low-order polynomials of spatial coordinates. In other words, they predict soil attributes solely as a function of the components of spatial position. Davies and Gamm (1970) were some of the first researchers to apply this technique in soil science. They made predictions of soil pH in Kent in England.

Some other purely spatial approaches focused on modelling local variation in soil attributes. Examples include nearest-neighbour interpolation, inverse-distance-weighted interpolation and smoothing splines. These approaches have been discussed by Laslett et al. (1987) and Burrough and McDonnell (1998).

By far the most well-known and widely applied purely spatial modelling approaches are based on geostatistics—the so-called regionalised variable theory developed in the 1960s and 1970s by French mathematician and geologist Georges Matheron. By treating soil attributes as regionalised variables, kriging interpolation could be used to model more complex spatial patterns than was previously possible

(McBratney et al. 2003). Kriging was introduced to soil science in the 1980s (Burgess and Webster 1980a, 1980b; Webster and Burgess 1980). Chapters 10 and 11 describe geostatistics in greater detail.

12.1.2.4 Environmental Correlation

Researchers have used statistics to quantify relationships between soil and landscape parameters for several decades. Initial work focussed on understanding the variability in soil attributes within and between soil map units (e.g. Wilding et al. 1965; Protz et al. 1968). Other work focussed on establishing numerical relationships between soil attributes and position along a hillslope (e.g. Ruhe and Walker 1968). Later work sought to understand soil distribution in a more multivariate way by including information on vegetation, geology and topography (e.g. Shovic and Montagne 1985), but such models were not put into a framework for spatial prediction until the early 1990s once computing power and raster technology were sufficiently developed. Bell et al. (1994) were one of the first to do so. They created maps of soil drainage class across a study area in Pennsylvania in the United States using a soil-landscape model they had developed in previous work (Bell et al. 1992).

Bell et al.'s work, and others like it (e.g. Moore et al. 1993), represented an alternative strategy for making spatial predictions of soil attributes and could be seen as a demonstration of how Jenny's fundamental concepts could be quantified in a multivariate way. About the same time, other researchers anticipated that such work would lead to better understanding of soil formation (McSweeney et al. 1994). Efforts by other researchers soon followed. McKenzie and Austin (1993) realised that multivariate soil-landscape models could lead to a new type of soil survey and coined the term environmental correlation to describe the process. Skidmore et al. (1991), Odeh et al. (1994), Lagacherie and Holmes (1997) and McKenzie and Ryan (1999) are other early examples.

The fundamental idea is that in order to make a map describing the distribution of a soil attribute across a given spatial domain, soil observations at points are intersected with layers of environmental data, and a model of some structure (frequently a regression model) is fitted to describe the relationship between the soil observations and the environmental data. The model is then used to make predictions of the soil attribute at unvisited sites, which are frequently the nodes of a raster grid across the area of interest. This approach is necessarily quantitative and relies on the assumption that the soil observations are correlated with the environmental data to some degree.

During the mid-1990s researchers realised that the classical geostatistical approach to soil spatial prediction could be merged with the environmental correlation approach. The combined approach became known as regression-kriging (Odeh et al. 1995). Regression-kriging can be summarised in three steps:

1. Calibrate an appropriate environmental correlation model using soil-landscape data from observed sites, and make predictions of the target soil attribute at unvisited sites.

2. Compute the statistical residuals of the environmental correlation model at the calibration sites, and interpolate the residuals at the unvisited sites,
3. If there is evidence of a spatial trend in the residuals, sum the residuals and the environmental correlation prediction to arrive at a final prediction of the target soil attribute at the unvisited sites.

Regression-kriging, when the regression component is a multiple linear regression model, is equivalent to universal kriging (Stein and Corsten 1991), or kriging with external drift (Wackernagel 1998) or even kriging after de-trending (Goovaerts 1999). In such cases the soil attribute is some linear function of the predictors. Regression-kriging differs from universal kriging when more complex models like regression trees are used in practice.

12.2 Digital Soil Mapping

These various threads in soil spatial prediction research began to converge in the early 2000s. Two seminal papers were published in 2003 that summarised historical developments, albeit from different perspectives. Scull et al. (2003) wrote about past research in soil spatial prediction from a physical geographic perspective and coined the term predictive soil mapping to describe the general approach. About the same time, McBratney et al. (2003) generalised the diversity of quantitative mapping approaches into a framework they called scorpan, which uses a clorpt-like framework not for mere explanation but rather empirical quantitative description of soil-landscape relationships for the purpose of spatial prediction. They called these activities digital soil mapping (DSM).

The scorpan framework is more formally known as scorpan-SSPFe, which includes soil spatial prediction functions (SSPF) and autocorrelated errors (e; McBratney et al. 2003). The scorpan factors are:

s: Soil, its classes or properties
c: Climate, including precipitation and temperature
o: Organisms, including vegetation, fauna and other factors of the biotic environment
r: Relief, or topography
p: Parent material, including lithology
a: Age
n: Space, or spatial position

The scorpan model is written as

$$S_c = f\,(s, c, o, r, p, a, n) \text{ or } \; S_p = f\,(s, c, o, r, p, a, n)$$

Table 12.1 Possible sources of information to represent the scorpan factors

scorpan factor	Possible representatives
s	Legacy soil maps, point observations, expert knowledge
c	Temperature and precipitation records
o	Vegetation maps, species abundance maps, yield maps, land use maps
r	Digital elevation model, terrain attributes
p	Legacy geology maps, gamma radiometric information
a	Weathering indices, geology maps
n	Latitude and longitude or easting and northing, distance from landscape features, distance from roads, distance from point sources of pollution

where *Sc* is soil classes and *Sp* is soil properties or attributes. Considering soils are sampled at spatial coordinates x, y at an approximate point in time, t, the model can be expressed explicitly as

$$S[x, y, t] = f(s[x, y, t], c[x, y, t], o[x, y, t], r[x, y, t], p[x, y, t], a[x, y, t], n[x, y, t],)$$

Soil (s) is included as a factor because soil can be predicted from its properties, soil properties from its class or other properties (McBratney et al. 2003), as earlier research has shown (e.g. McBratney and Webster 1983). This additional soil information could be gathered from a prior soil map or from either remote or proximal soil sensing or even expert knowledge. The n factor means that soil can be predicted as a function of spatial position alone, as in the case of kriging, but it may also be predicted as a function of the distance from some landscape feature such as streams, hilltops, roads or point sources of pollution, etc. Information that may represent the other scorpan factors is described in Table 12.1.

In the last 15 years, there has been a proliferation of high-resolution environmental spatial data, and many of these can be used to represent various scorpan factors. Digital elevation models (DEMs; Minasny et al. 2008) and remotely and proximally sensed data are two prominent examples (Mulder et al. 2011). Subsequently, DSM techniques have been employed to build or populate spatial soil information systems from relatively sparse datasets from the ground up (Lagacherie 2008). DSM techniques have also been used to update or renew existing soil mapping (e.g. Kempen et al. 2009, 2015; Nauman and Thompson 2014; Odgers et al. 2014). Generally speaking, DSM is seen as a practicable framework for fulfilling the current and future demand for relevant soil information (Sanchez et al. 2009). We discuss the growth and operationalisation of digital soil mapping later in this chapter.

In practice, the creation of soil maps using the DSM approach was discussed previously in the description of the environmental correlation and regression-kriging approaches. Within some mapping domain, a set of new or existing soil observations m are taken from explicit locations [x, y]. This is followed by fitting some kind of mathematical or statistical function to a set of pedologically meaningful environmental layers which are generally gridded raster layers of a given spatial resolution. Once the model is fitted at the m observation points, the model is extended to all grid cell nodes of the raster layers, giving a digital soil map. This

three-component process is the hallmark of DSM (Minasny and McBratney 2016), which entails (1) the input, (2) the modelling process and (3) the output.

This DSM approach is quite distinct from earlier notions of digital soil mapping that simply involved the digitisation of conventional soil mapping via electronic scanning. A more appropriate term for this would be digitised soil mapping.

12.2.1 Soil Spatial Prediction Functions

The modelling step is a crucial step in the digital soil mapping process and fundamentally distinguishes it from digitised soil mapping. The form of the spatial soil prediction function $f()$ is usually determined at the outset of a digital soil mapping project. When deciding which mathematical model is the most appropriate for a given application, several factors are usually taken into account, including:

1. The operator's familiarity with the model
2. The model's ease of application in the context of the project, the availability of covariates and the idiosyncrasies of the available soil information
3. The model's complexity and its power to capture potentially complex soil-landscape relationships within the target mapping domain

Many mathematical models are available to represent $f()$, and their development continues with advances in statistical theory; McBratney et al. (2000, 2003) discuss some important factors, and Hastie et al. (2009) review them in more detail. Generally, some models, such as multiple linear regression or regression trees, are suited to modelling continuous variables, whereas others, such as logistic regression or classification trees, are suited to modelling ordinal or nominal categorical data.

Some of the simplest models are simple linear models with either ordinary or generalised least-squares fitting. More complex models include generalised linear and additive models (Hastie and Tibshirani 1990). Logistic regression models are a type of generalised linear model suited to modelling categorical variables (Kempen et al. 2009). Recursive partition models such as classification and regression tree models are particularly favourable because of their non-parametric structure and their capacity to deal with non-linear relationships between soil data and environmental covariates (Breiman et al. 1984). For similar reasons neural networks (Hastie et al. 2009) are often considered, as are advanced data mining or machine learning algorithms such as Cubist models (Quinlan 1993b; http://www.rulequest.com/), random forests (Breiman 2001), quantile regression forests (Meinshausen 2006) and support vector machines (Smola and Scholkopf 2004).

In terms of handling any spatial autocorrelation in the residual (e) that is likely to result from fitting a scorpan model, a regression-kriging (alternatively scorpan kriging; McBratney et al. 2003) or universal kriging methodology could be used. Coupling a machine learning model with geostatistical modelling of the residuals is also a generalised regression-kriging approach that is often used in digital soil mapping studies (e.g. Malone et al. 2009).

Lark et al. (2006) described a residual maximum likelihood-empirical best linear unbiased predictor (REML-EBLUP) model. REML-EBLUP is intrinsically similar to regression-kriging in that both are mixed models where the observed data are modelled as the additive combination of fixed effects (the secondary environmental data), random effects (the spatially correlated residuals e) and independent random error. The difference is that REML estimates the parameters of the trend and covariance functions without bias. These parameters are then used in the EBLUP, i.e. a general linear mixed model. The statistical theory of REML-EBLUP is discussed in Lark et al. (2006) and arose out of a need to rectify issues associated with normal regression-kriging (the method proposed by Odeh et al. 1995) where estimation of the variogram of residuals is theoretically biased (Cressie 1993). While regression-kriging is an ad hoc method and may be theoretically suboptimal, the improvement in prediction accuracy from REML-EBLUP has been demonstrated to be only small when comparative analyses have been performed (e.g. Minasny and McBratney 2007).

The application of formal geostatistical modelling approaches to categorical variables, such as soil classes, is relatively limited. Kempen et al. (2012) reminded us that the popular methods for categorical prediction of soil types for the ultimate creation of a digital soil type map—multinomial logistic regression, classification tress, etc.—are actually nonspatial models, i.e. we do not consider spatial properties of the target variable as we would do for continuous variables. Subsequently, they explored a generalised linear geostatistical model framework that addressed this issue. While the method they explored was geostatistically appealing, it was also computationally cumbersome and ultimately did not yield significant accuracy gains when compared to a nonspatial multinomial logistic regression model.

Overall, it is clear there are many potential models available to use for DSM, and they can range from the very simple to the highly complex. Importantly, most models are now freely accessible through open source software packages. An example is the R computing software (R Core Team 2015) and the associated caret package (Kuhn et al. 2016) where, as of March 2017, there are 448 types of modelling algorithms available for both continuous and categorical target variables. Although many of these will not be suited for DSM, a large number will be, and algorithms will continue to be added to the caret package into the future.

12.2.1.1 Comparative Studies

A number of researchers have compared and contrasted suites of models for the spatial prediction of soil classes and attributes with a view to identifying the ‘optimal’ model for a particular situation. For example, Taghizadeh-Mehrjardi et al. (2016) evaluated the ability of six different models for mapping soil organic carbon across a region in Iran. Brungard et al. (2015) performed a similar analysis for mapping categorical variables in the south west of the United States, and Heung et al. (2016) compared a number of machine learning algorithms for prediction of soil classes in British Columbia in Canada. Generally speaking, the more complex

models performed better than the simpler models. Reasons for this include that soil data are often ill-suited for linear regression modelling because of non-normal or unbalanced distributions, whereas this does not appear to be an issue for recursive partitioning or machine learning-type algorithms. Secondly, the more complex models are better at discovering patterns and non-linearity in the data that cannot be easily discovered by simple linear modelling.

Despite this, there is a downside to this gain in predictive precision. First, the more complex models require more parameters compared to the simpler models. In turn, these parameters need to be optimised appropriately. Even for non-parametric models such as regression trees and the related boosted regression algorithms, the size and complexity of the fitted models can make them ultimately difficult to interpret and cumbersome to apply spatially. Such complex models are often referred to as black boxes, such that the practitioner is often not fully aware of their inner workings. In situations like these, the model just becomes a processing object where data is fed into it, processes it, and then outputs a computed result. The job of the operator then is to interpret the computed result and describe the predicted target variable spatial distribution in terms of the environmental variables that were used as model predictors. Often such complex models will provide output that gives the operator some idea of the importance of the predictor variables in the same way that significance tests are used in linear models to distinguish good predictor variables from poor ones. This information helps the operator to understand the strengths and weaknesses of digital soil maps produced using the model.

12.2.1.2 Knowledge-Based Inference

An alternative to purely empirical modelling is knowledge-based inference (Zhu and Band 1994). Knowledge-based inference enables the digital soil mapper to integrate expert knowledge into the mapping process. One of the most well-known knowledge-based tools is the soil-land inference model or SoLIM (e.g. Zhu et al. 1996, 1997, 2001). SoLIM allows an expert to manually build membership functions that describe the presumed relationship between specific soil types and a range of environmental and topographic variables. With these membership functions, the expert is able to make predictions of soil type or properties at unobserved locations via a weighted estimate that is based on an environmental similarity score to each soil member. Subsequently, the appealing concept of fuzziness between soil objects is maintained, as well as some quantitative means for assessing the uncertainty of mapped predictions.

In similar work, Bui (2004) pointed out that the soil map legend, which is a representation of the distilled knowledge of the soil surveyors' mental model of soil variation across a mapping domain, contains valuable and important structured language that can be used for automating soil mapping if spatial coverage of environmental predictors are available. Probably the closest empirical relative to this approach is the decision tree type models, where data are recursively partitioned to minimise some predictive variance.

The idea of utilising existing soil mapping in new, quantitative, ways has also found application in the spatial disaggregation of soil map units, where the information contained in these map units can be downscaled in order to make spatial predictions of the map units' constituent soil types (e.g. Nauman and Thompson 2014; Odgers et al. 2014). Bayesian networks, described by Taalab et al. (2015), also enable the operator to explicitly include expert knowledge. They are also able to be tuned empirically. The methodology requires the need to first express prior knowledge, in the form of probabilities, about the interaction between a target variable and a set of environmental covariates. Using these prior probabilities together with Bayesian inference, one is able to estimate either continuous or categorical variables and then use the derived posterior probabilities as a quantitative measure of uncertainty.

12.2.1.3 Uncertainty

It is relatively easy to quantify the accuracy of a digital soil map using a set of soil observations held aside from the spatial modelling process for this purpose. A range of statistical tests are available, and some of these are explored further in Chap. 14. These statistics provide a global appraisal of model performance: that is, they typically cannot tell us anything about the performance of a model at a specific location in the prediction area.

On the other hand, the local appraisal of a model's performance can be done through examination of uncertainties, if they are available. Such uncertainties can be quantified at each grid cell across the prediction area and are often expressed in the form of a prediction variance or prediction interval for soil attributes (Malone et al. 2011) or even probability estimate for soil classes (e.g. Odgers et al. 2014) or exceedance thresholds (e.g. Brus et al. 2002; von Steiger et al. 1996).

Uncertainties may be computed using a range of methods. For example, it is relatively straightforward to compute the kriging variance using geostatistics. Minasny et al. (2011) demonstrated a model-based Bayesian approach. Bayesian networks (Taalab et al. 2015) may offer a more expert-driven approach.

Machine-learning methods, and particularly those based on iterative resampling and boosted model fitting, can provide empirical estimates of uncertainty. An example here is the quantile regression forest algorithm that was used in France by Vaysse and Lagacherie (2017).

Methods based on model perturbation through data resampling or bootstrapping have been shown to be effective for quantifying uncertainties, particularly where the number of model parameters is high, or for very large mapping extents the use of model-based geostatistical approaches is computationally prohibitive. For example, Viscarra Rossel et al. (2015) fitted complex Cubist models coupled with residual kriging for spatial modelling of a number of soil properties across continental Australia. Such an approach necessitated the use of bootstrapping to compute 90% prediction intervals about the mapped predictions.

Malone et al. (2011) demonstrated another empirical approach which is based on data resampling and fuzzy k-means with extragrades. In their approach, the environmental data space is partitioned into clusters which share similar model errors. A prediction interval (PI) is constructed for each cluster on the basis of the empirical distribution of residual observations that belong to each cluster. A PI is then computed for each prediction grid cell according to the grade of its membership to each cluster in the environmental data space. The approach is flexible because the user is able to incorporate a complex model, such as a data mining model with associated residual kriging, into the framework and also efficiently compute prediction uncertainties across large spatial extents.

In quantifying prediction uncertainties, we explicitly acknowledge that a digital soil map is not free from error. A major source of error is the sparseness of soil data, both in the landscape space and the attribute space. Often the uncertainty created by this error is of a magnitude that would preclude the use of a digital soil map in many situations where fine precision is a requirement. Consequently, decisions or policies developed on the basis of the mapping need to be made with a certain amount of risk, although this risk can often be quantified. In order to reduce this risk, the uncertainties may be used to prioritise data collection resources or direct the application of alternative modelling approaches to improve the digital soil map and reduce its uncertainty.

12.2.1.4 Case Study: Digital Soil Property Mapping

In this section we provide an example that demonstrates a typical digital soil mapping workflow. We illustrate this workflow by fitting several different spatial prediction functions that exemplify just a few of the models that may be considered for DSM. The target variable in these examples is subsoil pH. The soil dataset comprises of 506 observations collected from a small part of the lower Hunter Valley region of New South Wales, Australia. For simplicity, the mapping is performed across a small part (approximately 220 km^2) of the area from which the samples were collected. The spatial prediction functions we use in this example are (i) multiple linear regression, (ii) Cubist models, (iii) Cubist models with kriging of model residuals and (iv) random forest models.

A number of spatial datasets were assembled for this region; all were co-registered from their original resolutions to a 25 m grid cell spacing; and all were used as predictors or environmental covariates. In this example, these data principally represent the r and o factors of the scorpan framework. The variables representing r were derived from a digital elevation model, while those representing o were derived from remotely sensed imagery captured by the Landsat 7 ETM+ satellite platform. These data are summarised in Table 12.2.

Some initial checks of the observed data are always performed to evaluate their statistical properties and check their distribution (see Chaps. 2 and 3). Data may need to be transformed so that their distribution becomes approximately

Table 12.2 Environmental covariates used for digital soil mapping example. Covariates are all co-registered to the same spatial resolution

scorpan factor	Source data	Variable
r	Digital elevation model	Elevation, mid-slope position, multi-resolution valley bottom flatness, terrain wetness index, slope gradient, estimated incoming solar radiation
o	Landsat 7 ETM+ satellite data (retrieved Jan 2012)	Landsat bands 1–5 and 7, normalised difference vegetation index (band 4 – band 3)/(band 4 + band 3))

normal. It is often necessary to remove outliers or obviously erroneous data. When the covariate dataset is acceptable, a spatial intersection between it and the soil observations is performed that retrieves the values of the covariate layers where the soil observations are positioned.

Quantitative assessment of the performance of spatial prediction functions is a distinguishing characteristic of DSM. According to Hastie et al. (2009), popular model goodness-of-fit statistics include root mean square error (RMSE), mean error (model bias), coefficient of determination (R2) and concordance correlation coefficient (CCC; Lin 1989). Importantly, it is necessary to compute these statistics using an independent test dataset in order to ensure that they are unbiased, to assess the generalisation performance of a model and to provide guidance in regard to selection of the optimal model. In a data-rich situation as recommended in Hastie et al. (2009), it is required to divide a dataset into three parts: a training set, a validation set and a test set. A typical split might be 50%, 25% and 25%, respectively, for each of the datasets. The training set is used for model fitting, while validation and test sets are used to evaluate to prediction and generalisation error, respectively. The test set is ideally only used right at the end of an analysis once the ideal model has been selected. In situations of data limitation, it is often recommended to implement data reuse procedures for the validation set, where cross-validation and bootstrapping are common examples of these (Hastie et al. 2009). In the proceeding examples, we split the data 75% and 25% for training and validation sets, respectively. We evaluate the goodness-of-fit functions of RMSE and CCC for the validation data for each of the considered soil spatial prediction functions.

With respect to the considered models, the simplest, multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. A stepwise regression is often implemented to filter out variables that don't add to the performance of the model. The Cubist model is a data mining algorithm which allows one to explore non-linear relationships in observed data. It is similar to a typical regression tree model in terms of it being a data partitioning algorithm. The Cubist model is based on the M5 algorithm of Quinlan (1993b). The Cubist model recursively partitions the data into subsets which are more internally homogeneous with respect to the

target variable and covariates than the dataset as a whole. A series of rules defines the partitions, and these rules are arranged in a hierarchy. Each rule takes the form:

if [condition is true]
then [regress]
else [apply next rule]

Each condition is based on a threshold for one or more covariates. For example, the Cubist model of subsoil pH used regression on the covariates within that partition. If the condition returns false, then the rule identifies the next node in the tree to move to, and the sequence of if, then-else is repeated. The result is that a separate regression equation is fit within each a threshold of <= 0.005 for MRVBF for one of the rules in partitioning the input dataset. If the condition returns true for this threshold, then the next step is the prediction of the target variable by OLS partition, and the errors are smaller than they would be if a single regression was fit to the entire dataset (Quinlan 1993b). Cubist models with residual kriging are an enhancement to using just Cubist modelling alone. Here, after the Cubist model has been trained, the model residuals (difference between observations and model predictions) are retained in order to assess whether they display a spatial pattern of their own (autocorrelation). This is assessed via variogram fitting to the residual data, and then once a candidate model is selected, kriging ensues. Variogram fitting could be performed locally for small neighbourhoods of data, or globally to all available data, which is generally dependent on the amount of available data (see Chap. 10). Ultimately, the regression model output and the kriged residuals are then added together to result in a final prediction. Lastly, random forests are a boosted decision tree model. Boosting is brought about via an ensemble learning method for that operates by constructing a multitude of decision trees which are later aggregated to give one single prediction for each observation in a dataset. One is able to establish a reasonable idea of goodness of fit of the training model via what is called 'out-of-bag' samples, which are data that are internally withheld during tree building. Further information on the popular model can be found in Breiman (2001) and Grimm et al. (2008) as an example of its application in DSM studies.

The relationship between target variable and spatial data is learned through the model fitting stage. Once this has been realised, the parameters are then used to predict the target variable across the mapping extent. Figure 12.1 shows the digital maps that resulted from applying each of the models across the mapping extent.

From Fig. 12.1, subtle differences between the maps can be observed, but all more or less show the same general spatial pattern. This may not always be the case however for other DSM exercises. The information contained in Table 12.3 summarises the RMSE and CCC goodness-of-fit statistics for each of the models both in terms of the training and validation datasets. It is generally the case that goodness of fit to appear better for the training relative to validation. For the MLR and Cubist model, this difference is minimal, indicating these models are not overfitting, or, in other words, they appear to be generally applicable. Cubist models with residual kriging results in the best predictions in terms of the validation. The poorest model was the random forest, yet this model was the best in terms of

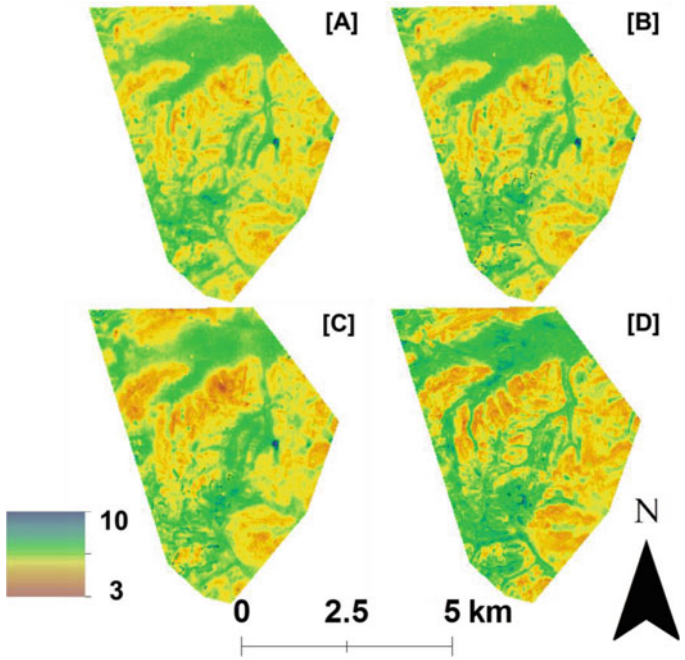


Fig. 12.1 Subsoil pH mapping as a result of (**a**) multiple linear regression, (**b**) Cubist modelling, (**c**) Cubist modelling with residual kriging and (**d**) random forest modelling

Table 12.3 Goodness of fit statistics for subsoil pH spatial models based on training and validation datasets. (CCC) concordance correlation coefficient (RMSE) root mean square error

	Training		Validation	
	CCC	RMSE	CCC	RMSE
Multiple linear regression	0.40	1.17	0.40	1.19
Cubist model	0.47	1.13	0.41	1.19
Cubist model with residual kriging	0.85	0.66	0.60	1.05
Random forest model	0.89	0.53	0.32	1.22

the training data. From this example, it is clear of the importance to assess model goodness of fit via validation and not on the training data alone. The random forest model gives the impression of being very accurate but is actually quite susceptible to overfitting in this case.

12.2.1.5 Case Study: Digital Soil Class Mapping

The following example is used to demonstrate how soil classes may also be mapped using digital soil mapping techniques. We illustrate two approaches using profile

observations from across the Hunter Wine Country Private Irrigation District in Pokolbin in the lower Hunter Valley, New South Wales, Australia (hereafter the PID dataset), where the previous example was located. As in the previous example, environmental conditions in the study area were represented by a range of terrain attributes derived from a digital elevation model with a cell size of 25×25 m and by spectral reflectance values derived from several bands of a Landsat 5 image captured in October 2004, around the time the profiles were collected.

The PID dataset consists of 243 profiles allocated to the suborder level of the Australian Soil Classification (ASC; Isbell 1996) although the predictions presented here are made to order level. The dataset is unbalanced with respect to profile class since soils belonging to the Dermosol order (55% of observations) dominate the study area, whereas soils such as Calcarosols (1 observation) occur much more sparsely. The data were randomly split into a calibration dataset (75% of observations) and a validation dataset (25% of observations). We illustrate both approaches here using the same calibration and validation datasets.

Classification trees may be used to calibrate a set of observations of soil profile class to a set of environmental covariates. The C5.0 (or See5) algorithm (based on Quinlan 1993a) is one of the more popular classification tree algorithms and operates in a similar manner to the Cubist algorithm, a major difference being the leaves of classification trees are predictions of a single class, not a regression equation. Application of a C5.0 classification tree to the PID dataset yielded the predictions of ASC order in Fig. 12.2a.

Multinomial logistic regression (MNLR) is used to model relationships between a categorical variable and a set of numerical predictors. More specifically, the log odds of the outcomes (classes) are modelled as a linear combination of the predictors. The end result is not only a prediction of the most probable class but also the underlying probabilities of occurrence of all the classes being modelled. Figure 12.2b contains the most probable ASC orders, and Fig. 12.3 contains predictions of the probabilities of occurrence of several ASC orders.

The maps in Fig. 12.3 are generally similar to each other. Three features are readily apparent: (i) Dermosols and, to a much lesser extent, Chromosols dominate the slopes under viticultural and grazing use throughout the central part of the private irrigation district; (ii) Kurosols are predicted under native woodland on the periphery of the district; and (iii) less structured and less developed soils such as Kandosols, Rudosols and Tenosols are predominant along drainage lines and rugged slopes particularly in the southwest of the private irrigation district. In broad terms the spatial distribution of soil orders as represented by these predictions fits well with what we have experienced in the field although neither map is a perfect depiction of reality. Dermosols are probably overpredicted in both maps, and the diversity of patterns along the southwestern escarpment between both maps warrants further investigation. In addition each map has its own pros and cons: for example, while the multinomial logistic regression predictions better captured the soil variability along drainage lines, Chromosols were possibly better predicted in the classification tree map. Furthermore, it is known that soils like Sodosols and

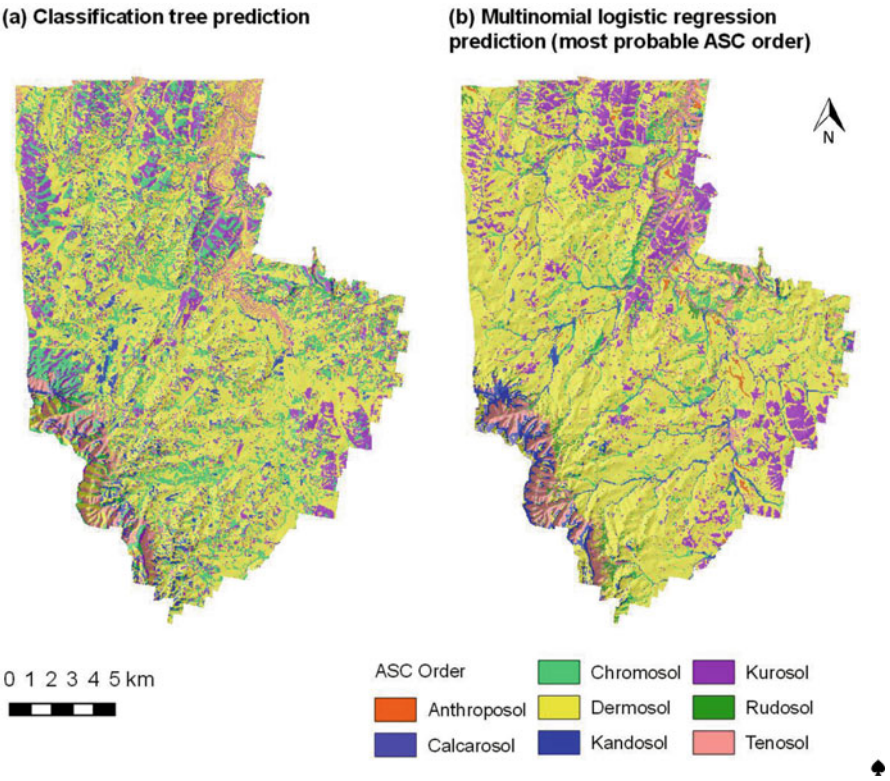


Fig. 12.2 Soil class predictions for the Hunter Wine Country Private Irrigation District using (**a**) a classification tree and (**b**) a multinomial logistic regression model

Hydrosols occur in relatively minor extent in some marginal lands in low-lying parts of the private irrigation district. These soils were not sampled in the PID dataset and so were unable to be predicted by either model.

The predictions in Fig. 12.2 were validated using the validation dataset of 25% of the PID profiles. It is convenient to summarise the results of a validation in a confusion matrix. The confusion matrix for the validation of the classification tree predictions is presented in Table 12.4 and that for the multinomial logistic regression is presented in Table 12.5. The overall accuracy—that is, the proportion of validation profiles whose ASC order was predicted correctly—is computed as the sum of the elements on the major diagonal divided by the total number of validation profiles. Thus the overall accuracy, A_o, of the classification tree predictions is

$$A_O = \left(\frac{0 + 0 + 1 + 26 + 0 + 2 + 0 + 0}{61} \right) \times 100 = 47.5\%$$

The overall accuracy of the multinomial logistic regression predictions is 50.1%.

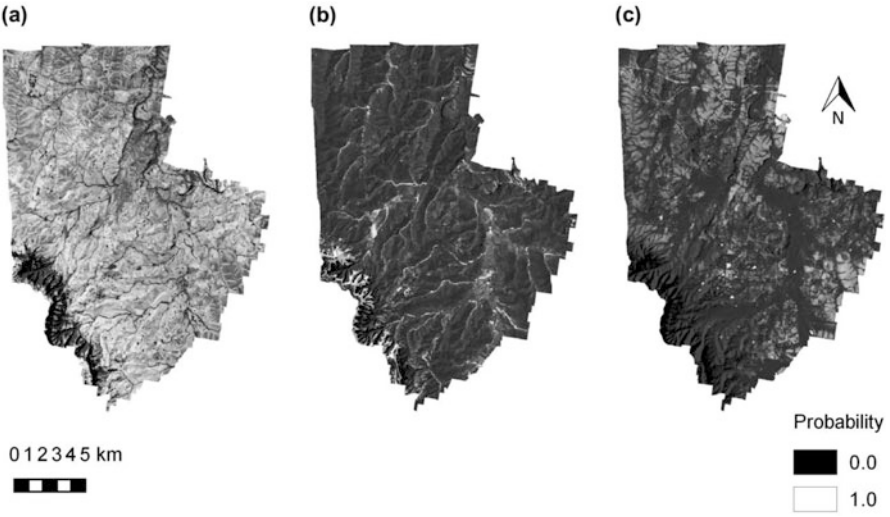


Fig. 12.3 Probabilities of occurrence from a multinomial logistic regression model for (**a**) Dermosols, (**b**) Kandosols and (**c**) Kurosols

Table 12.4 Confusion matrix for external validation of classification tree predictions of ASC order

		Observed ASC order								Row sum
		AN	CA	CH	DE	KA	KU	RU	TE	
Predicted ASC order	AN	**0**	0	0	0	0	0	0	0	0
	CA	0	**0**	0	0	0	0	0	0	0
	CH	0	0	**1**	2	0	0	1	0	4
	DE	0	0	12	**26**	3	4	1	0	46
	KA	0	0	0	0	**0**	0	0	0	0
	KU	0	0	3	6	0	**2**	0	0	11
	RU	0	0	0	0	0	0	**0**	0	0
	TE	0	0	0	0	0	0	0	**0**	0
Column sum		0	0	16	34	3	6	2	0	**61**

AN Anthroposol, *CA* Calcarosol, *CH* Chromosol, *DE* Dermosol, *KA* Kandosol, *KU* Kurosol, *RU* Rudosol, *TE* Tenosol

The overall accuracy does not indicate how the accuracy is distributed across the individual soil orders. There are two ways to look at this: (i) from the perspective of what we would expect the map to tell us based on what we observe in the field and (ii) from the perspective of what we would expect to see at a site in reality based on examination of the map. A measure that quantifies accuracy in the former case is frequently known as the producer's accuracy and in the latter case the user's accuracy (Story and Congalton 1986; Congalton 1991). For a given soil order, the producer's accuracy is computed by dividing its value on the major diagonal of

Table 12.5 Confusion matrix for external validation of multinomial logistic regression predictions of ASC order

		Observed ASC order								Row sum
		AN	CA	CH	DE	KA	KU	RU	TE	
Predicted ASC order	AN	**0**	0	0	0	0	0	0	0	0
	CA	0	**0**	0	0	0	0	0	0	0
	CH	0	0	**1**	1	0	0	0	0	2
	DE	0	0	13	**28**	2	5	1	0	49
	KA	0	0	1	0	**1**	0	0	0	2
	KU	0	0	0	2	0	**1**	1	0	4
	RU	0	0	0	2	0	0	**0**	0	2
	TE	0	0	1	1	0	0	0	**0**	2
Column sum		0	0	16	34	3	6	2	0	**61**

AN Anthroposol, *CA* Calcarosol, *CH* Chromosol, *DE* Dermosol, *KA* Kandosol, *KU* Kurosol, *RU* Rudosol, *TE* Tenosol

Table 12.6 User and producer's accuracies for predictions of Australian Soil Classification order

		AN	CA	CH	DE	KA	KU	RU	TE
User's accuracy (%)	C5.0	*NA*	*NA*	25.0	56.5	*NA*	18.2	*NA*	*NA*
	MNLR	*NA*	*NA*	50	57.1	50.0	25.0	0.0	0.0
Producer's accuracy (%)	C5.0	*NA*	*NA*	6.3	76.5	0.0	33.3	0.0	*NA*
	MNLR	*NA*	*NA*	6.3	82.4	33.3	16.7	0.0	*NA*

AN Anthroposol, *CA* Calcarosol, *CH* Chromosol, *DE* Dermosol, *KA* Kandosol, *KU* Kurosol, *RU* Rudosol, *TE* Tenosol

the confusion matrix by the corresponding column sum and multiplying by 100. Likewise, its user's accuracy is computed by dividing its major diagonal value by the corresponding row sum and multiplying by 100.

Values of the user's and producer's accuracies are presented in Table 12.6. They may be interpreted as follows (after Congalton 1991): 33.3% of the time, a site observed to be a Kurosol was predicted as such by the classification tree map; on the other hand, 18.2% Kurosol classification tree map predictions turn out to be Kurosols in reality. According to Table 12.6, the map of multinomial logistic regression predictions is generally more accurate than the map of classification tree predictions. Several features of Table 12.6 are notable. Some user's accuracies are reported as NA because the relevant ASC orders were not predicted in the given map. The number of user's accuracy NAs varies between the C5.0 and MNLR validations because the models do not make identical predictions. Some producer's accuracies are reported as NA because the relevant ASC orders were present in the calibration dataset but not in the validation dataset. The number of producer's accuracy NA values is the same for both maps because they share the same calibration dataset.

12.2.2 Soil Depth Functions and Digital Soil Mapping

Soil survey, soil classification and conventional mapping of soils consider soil as a three-dimensional (3D) entity or body (Hole 1953). In the words (or to that effect) of American pioneering soil scientist Charles E. Kellogg: 'soils have shape and area, breadth and width, as well as depth' (Kellogg 1949). Most initial DSM research tended to focus only on the 2D sense of this general concept where predictions of soil property variation were made for single depth intervals or horizons (and predominantly only from the top soil) (Grunwald 2009). In fact, Grunwald (2009), who reviewed 90 journal articles from high impact soil science journals, found that 28% of reviewed studies performed scorpan modelling for multiple soil layers or horizons. Since the review by Grunwald (2009), however, an active area of research in DSM has been the exploration of approaches for mapping of soil depth functions of continuous variables which is discussed in more detail further on. As an aside, the DSM of soil classes is an exception for 3D soil mapping and has much utility because a number of soil properties can be inferred from one classification which are usually derived from the modal profile of each class. However, the problem with mapping soil classes is that the soil properties vary discretely in the taxonomical space which may be problematic as soil variability across a mapping domain will appear as a stepped function rather than a fully continuous function (Webster and Oliver 2006).

It is quite a sensible undertaking to attempt to map the variation of soil properties in both the lateral and vertical dimensions. Understanding the carbon sequestration potential of soil and for carbon accounting (Bajtes 1996; Lal 2004), determining the amount of water soils can hold across a field or even a watershed, determining the depth to an impeding layer for crop growth across a farm and investigating soil acidity (just to detail a few examples) will most likely require some understanding of how soil properties vary with depth. Ponce-Hernandez et al. (1986) describe that soil properties vary more or less continuously with depth. The variation is often anisotropic for certain properties such as carbon (Hiederer 2009) and soil texture (Myers et al. 2011), which may be the result of land use activity or the gravitational vector of profile weathering and development or both (Hole 1961). Exceptions to continuous soil property variation with depth is where there is strong anthropogenic (cultivation, removal and replacement of soils), geologic (contrasting parent materials) and pedological (the development of clear and abrupt soil horizons) forcing for which sharp discontinuities in the depth distribution of soil properties will occur.

Empirical functions describing the depth distribution of soil properties include linear and polynomial functions (Colwell 1970; Moore et al. 1972) and exponential and logarithmic functions (Russell and Moore 1968). Myers et al. (2011) introduced an asymmetric peak function for modelling complex and anisotropic soil property depth profiles with horizons of weathered loess. Smoothing splines for soil property variation were introduced by Erh (1972) which was followed by work from Ponce-Hernandez et al. (1986) and mathematical derivation by Bishop et al. (1999) of

pycnophylactic (mass preserving) smoothing splines. The mass-preserving splines model the continuous variation of soil properties with depth while maintaining the average of the observed property through the observed horizons or layers. Bishop et al. (1999) demonstrated successfully the application of the mass-preserving spline for a number of soil properties with much success. The proviso for a 'good fit' is that a sufficient number of observations at regular depths are required. Many of these depth functions are described in Chap. 9.

The coupling of soil depth functions with DSM seems an intuitive advance towards understanding soil variation in all its spatial dimensions. A study by Minasny et al. (2006) used the negative exponential depth function to describe soil carbon concentration variation with depth in the Edgeroi area, Australia. The authors modelled the parameters of the exponential function using a modified neural network approach, then predicted parameters of the exponential function over the whole area, which enabled them to calculate the carbon distribution over the profile and also the storage of carbon at any depth. Mishra et al. (2009) fitted an exponential function to soil profile data from Indiana, USA, and then interpolated the parameters independently using ordinary kriging. Meersmans et al. (2009) performed something similar and developed empirical functions which predicted the parameters of the exponential depth function for the area of Flanders in Belgium. The functions were stratified based on land use, and the parameters were related to particle size distribution and height of groundwater. Veronesi et al. (2012) used a polynomial soil depth function for predicting soil penetration resistance. The coefficients of the depth function were modelled spatially via a REML-EBLUP geostatistical model.

Malone et al. (2009) described a regression-kriging approach using neural networks coupled with the Bishop et al. (1999) mass-preserving spline for mapping available water capacity and soil carbon stocks in Australia. The approach is performed in two steps with the first being the standardisation of depth intervals given a collection of soil profile data with the spline. Integrating the spline for the specified depth intervals, for example, 0–5 cm, 5–15 cm, 30–60 etc., an average value was defined for each depth for each soil profile. Not only are these values the predicted means for a specified depth interval but are also parameters of the spline function that can be used for subsequent refitting and soil information interrogation. After the splines are fitted, for each specified depth interval, a spatial model is fitted which is then ultimately used for mapping. With such a coupling of depth function and DSM, Malone et al. (2009) were able to generate scenarios such as the depth at which the cumulative total of soil carbon was equal to 5 kg m^{-2}. The mass-preserving spline in concert with DSM is widely used throughout the world with salient examples being Adhikari et al. (2013) for national scale mapping in Denmark and both Kidd et al. (2015b) and Viscarra Rossel et al. (2015) where spatial extents were the state of Tasmania, Australia and continental Australia, respectively.

Poggio and Gimona (2014) describe a one-step approach to 3D digital soil mapping using a hybrid GAM geostatistical model for carbon stocks in Scotland. One-step approaches are naturally appealing as it reduces the complexity of workflow. With the Poggio and Gimona (2014) approach, they are able to refrain

from standardising soil prediction depth intervals, i.e. use the actual observed values, and then model the 3D soil variability trend with a GAM and an associated 3D smoother with related covariates. The one-step approach was also appealing to Orton et al. (2016), who introduced an area-to-point kriging methodology. This model was able to include all the data in one statistical analysis and importantly maintained the integrity of the soil profile data support as well as providing an explicit methodology for quantifying the uncertainty of the fitted soil depth function. Hengl et al. (2014) described a similar 3D geostatistical model except information regarding the support of the soil profile observation was not considered. In their global study, they assumed a point support soil depth function by fitting splines to the values at mid-depths of observed soil horizons or depth intervals.

Kempen et al. (2011) developed a depth function that combines general pedological knowledge with geostatistical modelling. They modelled the distribution of soil organic matter content based on typical horizons from ten soil types. Five depth function building blocks were defined, and for each soil type, the depth function structure was obtained by stacking a subset of modelled horizons. The parameters of the depth function for each of the horizons were interpolated using a geostatistical procedure combining environmental information.

One of the limitations for 3D DSM, particularly of using equivalents of the negative exponential depth function, is that the function is only useful for certain soil properties, such as soil organic carbon, that naturally have that type of anisotropic variation down the soil profile. Polynomial soil depth functions also need to be considered carefully because the value at one depth will affect the fit of the curve at other depths too. Additionally, while pedologically the most appealing approach, the horizon depth function method from Kempen et al. (2011) could be limiting because its application is limited to the spatial extent in which it was developed or otherwise in similar landscape contexts or soil types.

A general method, i.e. one that can be extended to a variety of soil properties and can cope to some extent with discontinuous variation in soil properties, is the mass-preserving spline method described in Bishop et al. (1999). The mass-preserving features of this function are also appealing, particularly in instances where auditing of soil variables (e.g. carbon stocks) is required. In contrast, specifying a depth function based on midpoints observations as Hengl et al. (2014) proposed in their one-step 3D geostatistical modelling approach would be of limited use where the context for spatial soil mapping is for auditing purposes.

12.2.3 The Many Faces of Digital Soil Mapping

While most application of DSM is based in the using of observed soil point data coupled with environmental covariates and a spatial predictive model function, it is certainly not restricted to this nature of available data. For example, some

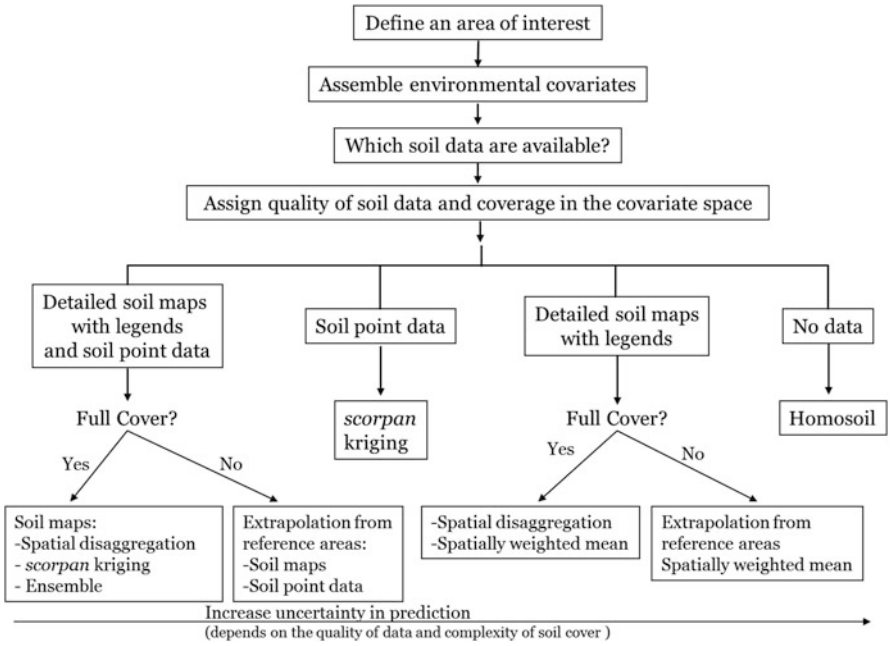


Fig. 12.4 Decision tree of digital soil mapping approaches (Adapted from Minasny and McBratney 2010)

previous discussion of soil map disaggregation has been made where mapping units are disaggregated into their constituent classes or series whose spatial pattern is determined by a spatial prediction function.

Before exploring that concept in further detail, it is worth pointing out the various possible scenarios that could be encountered for DSM. These scenarios are framed nicely by Minasny and McBratney (2010), who presented a decision tree for DSM methodology on the basis of the nature of available legacy soil data (Fig. 12.4). This tree was proposed as a general framework to aid in the delivery of a digital global soil map product. It can be viewed though as guidance to practitioners in 'what do I do?' situations. Once the practitioner has defined an area of interest, and assembled a suite of environmental covariates for that area, depending on what available data there is to use, there are suggested approaches that could be implemented for delivering digital soil maps. The most common approach obviously is scorpan kriging and is performed exclusively when there is only point data but can be used also when there is both point and map data available. The idea here is that a fusion between scorpan kriging and soil map disaggregation is made via a combinatorial approach such as model averaging. Malone et al. (2014) exemplified such a procedure in an area of Queensland, Australia. The options are quite different when there is only soil map information available. Bear in mind that the quality of the soil maps depends on the scale and subsequently variation of soil cover,

such that smaller-scaled maps, e.g. 1:100,000, would be considered better and more detailed than large-scaled maps, e.g. 1:500,000. The elemental basis for extracting soil properties from legacy soil maps comes from the central and distributional concepts of soil mapping units. For example, modal soil profile data of soil classes can be used to quickly build soil property maps. An early example of this is the multilayer soil characteristics dataset for the conterminous United States (CONUS-SOIL), which has been and continues to be used in many climate, hydrology and land surface models (Miller and White 1998). Where mapping units consist of more than one component, we can use a spatially weighted means type method, i.e. estimation of the soil properties is based on the modal profile of the components and the proportional area of the mapping unit each component covers, e.g. Odgers et al. (2012). As a pre-processing step prior to creating soil attribute maps, it may be necessary to harmonise soil mapping units (in the case of adjacent soil maps) and/or perform some type of disaggregation technique in order to retrieve the map unit component information. Some approaches for doing so have been described in Bui and Moran (2003).

More recently, research into the spatial disaggregation of soil map units has emerged as a means of downscaling choropleth soil mapping. Many studies have been published in the last several years (e.g. Häring et al. 2012; Kerry et al. 2012; Odgers et al. 2014; Subburayalu et al. 2014; Thompson et al. 2011). For example, the DSMART algorithm (Odgers et al. 2014) spatially disaggregates a choropleth map by iteratively resampling it to build a series of classification trees that create realisations of the potential soil class distribution. The realisations of the soil class distribution are merged to estimate the probabilities of occurrence for all the soil classes in the choropleth map area. The DSMART algorithm has been used to spatially disaggregate the legacy soil mapping coverage of large river catchments (Odgers et al. 2014), the French region of Brittany (Vincent et al. 2016), the Australian state of Western Australia (Holmes et al. 2015) and the contiguous United States (Chaney et al. 2016). The spatially disaggregated soil mapping can then be used to map soil properties. For example, Odgers et al. (2015b) introduced the PROPR algorithm which takes probability outputs from DSMART together with modal soil profile data of given soil classes, to estimate soil attributes and associated uncertainties.

How do we do digital soil mapping in an area that doesn't have any soil data? In such areas ('recipient' areas), it may be possible to extrapolate using a model that was constructed in an area with plentiful soil data (a 'donor' area). The rationale, called homosoil (Mallavan et al. 2010; Minasny and McBratney 2010), is that if the recipient area is sufficiently homologous to the donor area with respect to the strength and expression of soil-forming factors, then perhaps extrapolation can be made without an unacceptable degree of uncertainty. Identification of potential donor areas is based on the computation of the similarity to the recipient area in terms of the relevant soil-forming factors. Malone et al. (2016) provided a real-world application which compared different extrapolation functions.

12.2.4 Operationalisation of Digital Soil Mapping

Although numerical methods have been available for several decades now, the operationalisation of digital soil mapping has really only been talked about for about the last 10 years (e.g. Grunwald et al. 2011; Lagacherie and McBratney 2007, Sanchez et al. 2009). Nevertheless it is becoming an important tool in mainstream soil survey work, and several important projects have been completed in the last few years, including the Soil and Landscape Grid of Australia (Grundy et al. 2015) and the enterprise suitability mapping programme in Tasmania (Kidd et al. 2014, 2015a). Not all operational work has focussed on the development of digital soil maps ab initio. For example, government-sponsored work has led to the updating of legacy maps using digital techniques (e.g. Kempen et al. 2009). Agencies in other jurisdictions are working to integrate digital soil mapping methods with existing survey procedures (e.g. Hewitt et al. 2010; Moore et al. 2010).

12.3 GlobalSoilMap

12.3.1 Origins: From Pedometric Practice to Global Product

Soil mapping at the global extent is not a recent or new endeavour. The FAO-UNESCO soil map of the world (FAO-UNESCO 1988) was the first published (1981) world soil map and used a single map legend that accounted for the global diversity of soils. During the 2nd Global Workshop on Digital Soil Mapping in Rio de Janeiro in 2006, a proposal for a new global grid of the most important soil functional properties was made. Later that year, the GlobalSoilMap.net consortium was formed with the aim of making a new, high-resolution digital soil map of the world using state-of-the-art and emerging technologies. The consortium was inspired in part because of 'policy-maker's frustrations' about not being able to get quantitative answers to questions such as how much carbon is sequestered or emitted by soils in a particular region? or what is its impact on biomass production and human health? or how do such estimates change over time? (Sanchez et al. 2009).

12.3.2 Technical Specifications of GlobalSoilMap

The technical specifications of GlobalSoilMap are laid out in the specifications document (Arrouays et al. 2014b). That publication articulates not how digital soil maps should be created but the standard to which they should conform to in order to permit collation for the assemblage of a global product. This specification promotes

innovation in digital soil mapping practice and recognises that practitioners may prefer certain methods and that different methods may perform better in different environments.

Key aspects of the specifications include the spatial entity such as the data support and resolution at which mapping should be created, the soil properties to be predicted, the date associated with their prediction as well as an explicit communication of the prediction uncertainty and accuracy. Other aspects include documentation standards and reproducibility and the data release policy. The following paragraphs will discuss these technicalities in more detail which is sourced from Arrouays et al. (2014b) and Arrouays et al. (2014a).

12.3.2.1 Raster Specifications

GlobalSoilMap products will be delivered as raster grids with cell dimensions of 3 arcseconds × 3 arcseconds. The resolution was chosen in order to match the 3-arcsecond Shuttle Radar Topography Mission (SRTM) DEM since elevation is a fundamental environmental covariate and the DEM extent accounts for most of the habitable surface of the Earth.

The horizontal datum of GlobalSoilMap products will be WGS84 (geographic), and the vertical datum will be EGM 96.

12.3.2.2 Tiered Products

The GlobalSoilMap specifications identify four tiers of product. Tier 1 products could be thought of as 'entry-level' products not with respect to quality but with respect to the complexity of the methods involved in producing them. The higher tiers are hierarchical in the sense that they build on products produced to a lower-tier specification, particularly with respect to validation and uncertainty analysis.

It is expected that mapping products meeting Tier 1 or Tier 2 specifications can be derived entirely from legacy soil information. It is possible that collection of additional soil data will be required in order to successfully complete a Tier 3 or Tier 4 product.

12.3.2.3 Soil Attributes

Soil attribute predictions will be made at six depth intervals in the soil profile: 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm and 100–200 cm. The values of 12 attributes will be predicted at each prediction location; they are described in Table 12.7.

Table 12.7 GlobalSoilMap soil attributes (Arrouays et al. 2014a)

Soil attribute	Units	Description
Soil depth		
Depth to rock	cm	Depth to lithic or paralithic contact (Soil Survey Staff (1993))
Plant-exploitable depth	cm	Effective depth as defined in Soil Survey Staff (1993)
Primary soil properties		
Organic carbon	g kg^{-1}	Mass fraction of carbon by weight in <2 mm fraction as determined by dry combustion at 900 °C
pH		pH of 1:5 soil-water suspension
Clay content	g kg^{-1}	<2 μm fraction of <2 mm soil material determined using pipette method
Silt content	g kg^{-1}	20–50 μm fraction of <2 mm soil material determined using pipette method
Sand content	g kg^{-1}	50 μm–2 mm fraction of <2 mm soil material determined using pipette method
Coarse fragments	m^3 m^{-3}	Mass fraction of soil material >2 mm
Effective cation exchange capacity	$mmol_c$ kg^{-1}	Cations extracted using $BaCl_2$ plus exchangeable Al and H
Derived soil properties		
Bulk density	Mg m^{-3}	Bulk density of whole soil, including coarse fragments, equivalent to the core method but determined using pedotransfer function
Bulk density	Mg m^{-3}	Bulk density of <2 mm fraction, equivalent to the core method but determined using pedotransfer function
Available water capacity	mm	Determined by pedotransfer function using previous predictions of organic carbon, sand, silt, clay and bulk density

It is left to individual jurisdictions to determine whether additional soil attributes are worth predicting. For example, the Soil and Landscape Grid of Australia, which is Australia's initial contribution to GlobalSoilMap, includes total N and total P in addition to the attributes identified in Table 12.7 (Grundy et al. 2015).

Definitions and methods of soil analysis for most soil properties are in accordance with Annex 1—Methods for Soil Analysis (ISO standards) of the FAO report outlining the World Reference Base for Soil Resources (IUSS Working Group WRB 2007). Particle size limits conform to the USDA system. Soil properties that are not routinely measured will be predicted with region-specific pedotransfer functions.

12.3.2.4 Uncertainty

It is intended that soil attribute predictions will be accompanied by an estimate of their uncertainty at each prediction location. Although the benefit of uncertainty information is clear, and research into the uncertainty of soil maps is not new, its incorporation in the specifications of a soil mapping project of this extent is unprecedented. It has directly inspired further research (e.g. Helmick et al. 2014; Malone et al. 2014).

Tier 1 and Tier 2 specifications require uncertainty to be expressed as the upper and lower limits of a 90% prediction interval about the soil attribute predictions, and a range of studies have shown how this may be accomplished (e.g. Helmick et al. 2014; Malone et al. 2011; Odgers et al. 2015a; Padarian et al. 2017).

12.3.3 Examples of the GlobalSoilMap Product: What Has Been Achieved So Far?

The GlobalSoilMap project has stimulated a large amount of research since its inception. Some of the first efforts are described in Boettinger et al. (2010) and Minasny et al. (2012). The first GlobalSoilMap conference was held in Orléans, France, in 2013 with the aim of showcasing GlobalSoilMap-inspired research. The proceedings of this meeting are published in Arrouays et al. (2014c). In the following sections, we will describe a few GlobalSoilMap products in more detail.

GlobalSoilMap products have been produced for many regions, including Australia (e.g. Soil Research 53(8), 2015), Chile (Padarian et al. 2017), Denmark (Adhikari et al. 2013, 2014), Europe (de Brogniez et al. 2015; Ballabio et al. 2016), France (Lacoste et al. 2016; Mulder et al. 2016), South Korea (Hong et al. 2012) and the United States (Bliss et al. 2014; Hempel et al. 2014; Odgers et al. 2011); Fig. 12.5 shows maps of 0–20 cm soil texture fractions.

Australia is currently one of the countries that is leading the GlobalSoilMap effort in producing nationwide soil attribute maps, using soil legacy data and new soil property estimates derived from vis-NIR soil spectra (Viscarra Rossel et al. 2015). Australia's contribution is produced under the Soil and Landscape Grid of Australia (SLGA) and is publically available online via the following web service architecture: www.csiro.au/soil-and-landscape-grid. Figure 12.6 shows an example of how the new product compares to existing Australia wide soil attribute maps. It becomes clear that the new Soil Grid of Australia is more accurate and contains less artefacts and generally represents the variability of soils across the country better.

Spatial disaggregation of soil class maps employing the DSMART algorithm has been applied on the national scale (e.g. POLARIS, United States; Chaney et al. 2016), state level (e.g. Western Australia; Holmes et al. 2015) and regional scale (e.g. Dalrymple Shire central Queensland, Australia; Odgers et al. 2014). Figure 12.7 shows an example on the regional scale. Here, DSMART was used to sample

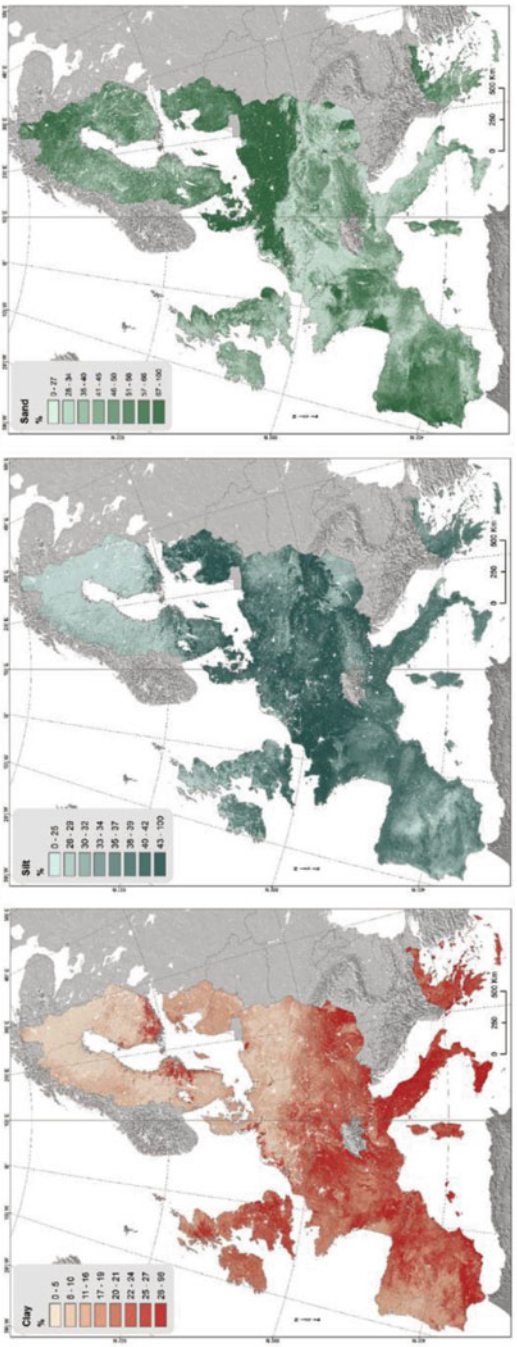


Fig. 12.5 Topsoil (0–20 cm) clay content (%), silt content (%) and sand content (%) of European countries modelled by multivariate additive regression splines and using point data from the land use and cover area frame statistical survey (LUCAS) (After Ballabio et al. (2016), used with permission)

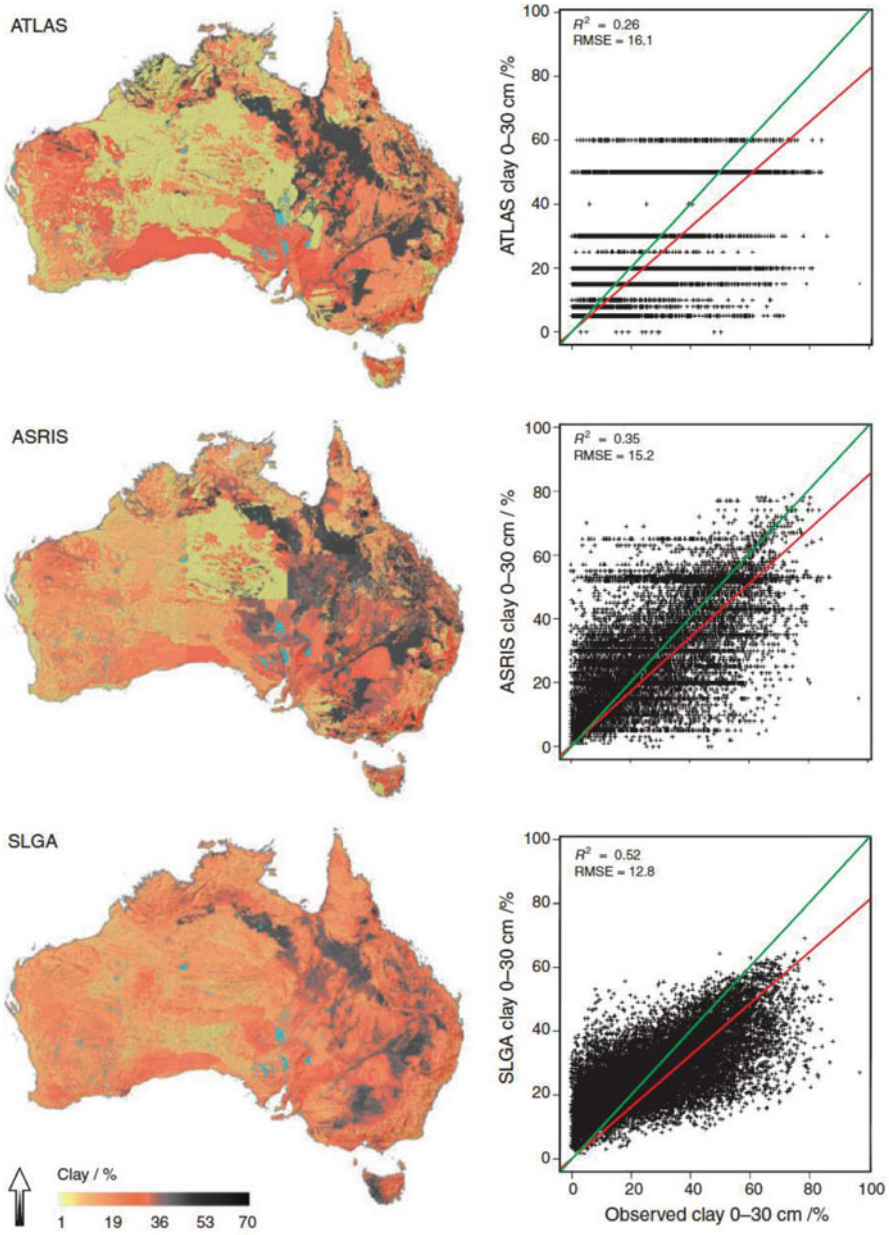


Fig. 12.6 Comparing maps of clay content for the 0–30 cm layer derived by attribution of clay content to polygons from the atlas of Australian soil (AAS); area-weighted means where the level of detail depends on the survey coverage in each region (ASRIS); and depth-weighted average to 30 cm from the fine spatial resolution multi-scale maps of the Soil and Landscape Grid of Australia (SLGA). *Blue* areas on the maps represent lakes and other inland water bodies (Reproduced from Viscarra Rossel et al. (2015), with permission from CSIRO Publishing)

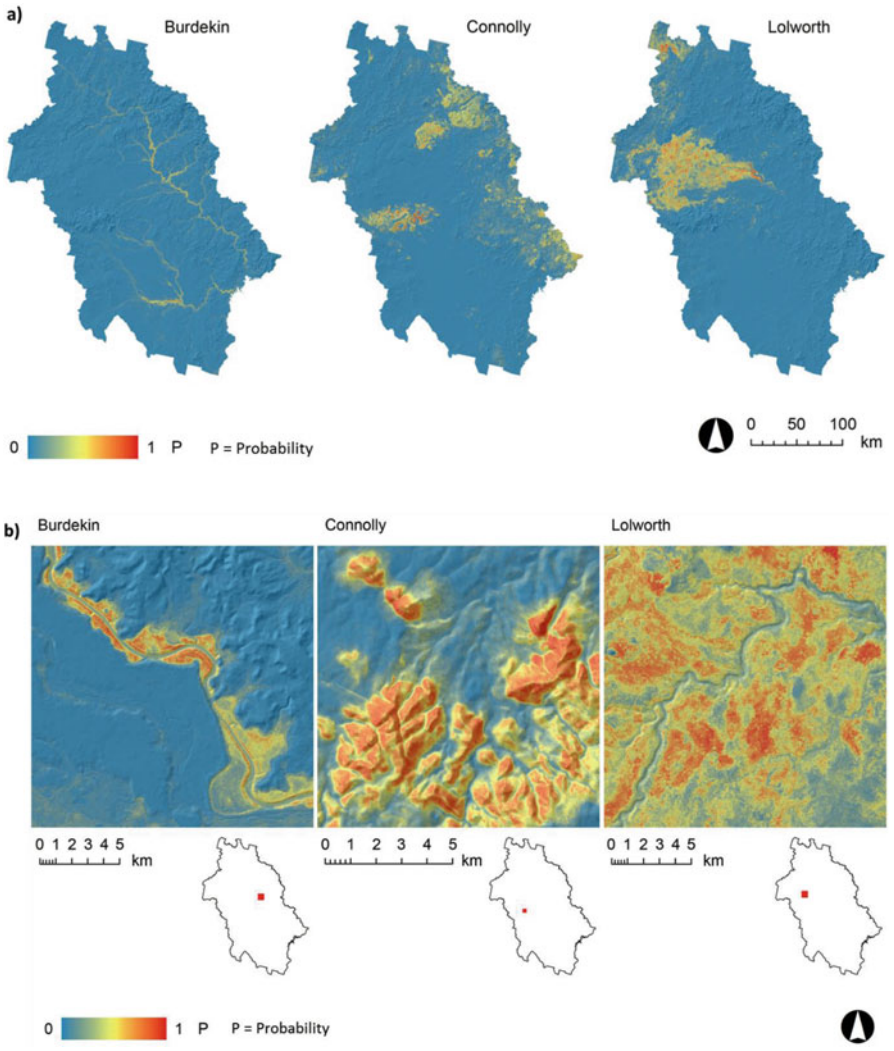


Fig. 12.7 (**a**) Probability surfaces for three of the 72 Dalrymple Shire soil classes, created by running DSMART on the 3058 soil polygons for $n = 100$ iterations at $m = 15$ sampling points per polygon. (**b**) Zoomed-in view (Changed after Odgers et al. 2014)

soil class polygons of a legacy soil map and subsequently derives potential soil class distributions via a classification tree. This was done through a number of realisations to estimate the probability of occurrence of an individual soil class (Fig. 12.7a).

Estimates were conducted on a raster grid to overcome some of the limitations of the discrete polygon boundaries of the original soil legacy map (see Fig. 12.8 which shows the most probable soil class).

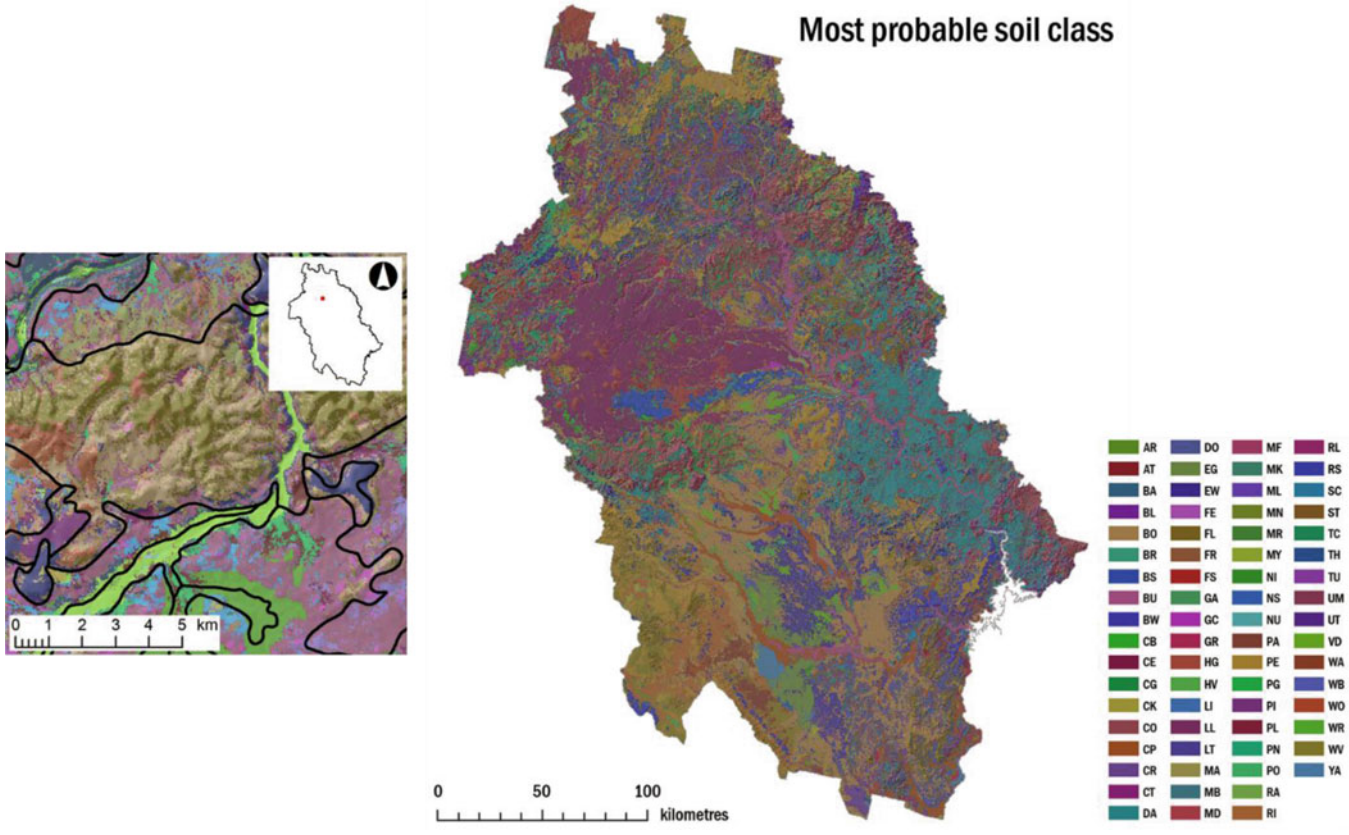


Fig. 12.8 Most probable soil class. It is also shown how the most probable soil class created with DSMART compares to the original soil polygons of the legacy map (After Odgers et al. 2014)

Spatial disaggregation of soil classes employing the PROPR algorithm has also been applied on the state level (e.g. Western Australia; Odgers et al. 2015a) and regional scale (e.g. Dalrymple Shire central Queensland, Australia; Odgers et al. 2015b). This approach is useful when sufficient soil profile point observations are limiting but where soil legacy polygon maps are available.

12.3.4 Uses of GlobalSoilMap Products

As outlined earlier the intention behind the GlobalSoilMap effort is to provide high-resolution soil attribute maps that can ultimately be used in a variety of applications. Sanchez et al. (2009) emphasised that 'maps can provide soil inputs (e.g. texture, organic carbon and soil-depth parameters) to models predicting land-cover changes in response to global climatic and human disturbances'. Here, we will discuss a few examples.

Maps of more readily available soil properties can be used to estimate other more costly or difficult to measure soil attributes. For example, Ballabio et al. (2016) used soil texture maps (%clay, %silt and %sand) created using topsoil data from the European land use and cover area frame statistical survey (LUCAS) (refer to Fig. 12.5) to derive bulk density, USDA soil texture classes and AWC. These soil attributes may then be employed to assess the C sequestration potential of topsoils (bulk density), assess the soil-water holding potential of agricultural soils for informed soil management or estimate soil compaction hazards (texture classes). High-resolution soil attribute maps can also be employed for monitoring and forecasting of biophysical properties. More explicitly, these soil attribute maps may be used as input soil information to drive physical system simulation models (e.g. crop simulation models such as APSIM) or for land suitability assessment.

A nice example of the use of fine-resolution 3D soil attribute maps in land suitability assessment that contributed to the GlobalSoilMap effort in Australia (Kidd et al. 2015b) comes from Tasmania, Australia, and is outlined in Kidd et al. (2015a). Here, digital soil assessment was used to inform on the agricultural land suitability of 20 different crops (perennial horticultural, cereal and vegetable crops) in a new irrigation scheme 'Water for Profit Program' commissioned by the Tasmanian state government. Together with 3D soil attribute maps (80×80 m), climate grids (e.g. growing degree days and frost risk) were applied to a range of defined enterprise suitability rulesets to produce enterprise suitability maps. All maps created are stored on a publically available spatial Internet portal (Land Information Services Tasmania, LISTmap https://www.thelist.tas.gov.au/app/content/data#) and can be used interactively to perform a range of land management-based assessments, such as identifying limiting soil and climate conditions. Outputs also include an enterprise versatility index; and the highest-valued agricultural land for the highest earning potential for individual commodities can also be identified with the product.

References

Adhikari K, Kheir RB, Greve MB, Bøcher PK, Malone BP, Minasny B, McBratney AB, Greve MH (2013) High-resolution 3-D mapping of soil texture in Denmark. Soil Sci Soc Am J 77:860–876

Adhikari K, Hartemink AE, Minasny B, Bou Kheir R, Greve MB, Greve MH (2014) Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS One 9:e105519

Arnold RW (2006) Soil survey and soil classification. In: Grunwald S (ed) Environmental soil landscape modeling: geographic information technologies and Pedometrics. Taylor and Francis, Boca Raton, pp 37–60

Arrouays D, Grundy MG, Hartemink AE, Hempel JW, Heuvelink GBM, Hong SY, Lagacherie P, Lelyk G, McBratney AB, McKenzie NJ, Mendonca-Santos MDL, Minasny B, Montanarella L, Odeh IOA, Sanchez PA, Thompson JA, Zhang G-L (2014b) GlobalSoilMap: toward a fine-resolution global grid of soil properties. Adv Agron 125:93–134

Arrouays D, McBratney A, Minasny B, Hempel J, Heuvelink G, MacMillan R, Hartemink A, Lagacherie P, McKenzie N (2014c) The GlobalSoilMap project specifications. In: Arrouays D, McKenzie NJ, Hempel J, de Forges AR, McBratney AB (eds) GlobalSoilMap: basis of the global spatial soil information system. CRC Press, London, pp 9–12

Arrouays D, McKenzie NJ, Hempel J, de Forges AR, McBratney AB (eds) (2014a) GlobalSoilMap: basis of the global spatial soil information system. CRC Press, London

Ballabio C, Panagos P, Monatanarella L (2016) Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261:110–123

Batjes NH (1996) Total carbon and nitrogen in the soils of the world. Eur J Soil Sci 47(2):151–163

Bell JC, Cunningham RL, Havens MW (1992) Calibration and validation of a soil-landscape model for predicting soil drainage class. Soil Sci Soc Am J 56(6):1860–1866

Bell JC, Cunningham RL, Havens MW (1994) Soil drainage class probability mapping using a soil-landscape model. Soil Sci Soc Am J 58:464–470. doi:10.2136/sssaj1994.03615995005800020031x

Bishop TFA, McBratney AB, Laslett GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91(1–2):27–45

Bliss NB, Waltman SW, West LT, Neale A, Mehaffey M (2014) Distribution of soil organic carbon in the conterminous United States. In: Hartemink EA, McSweeney K (eds) Soil carbon. Springer International Publishing, Cham, pp 85–93

Boettinger JL, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S (eds) (2010) Digital soil mapping. Bridging research, environmental application, and operation. Springer, The Netherlands

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont

Brown DJ (2006) A historical perspective on soil-landscape modelling. In: Grunwald S (ed) Environmental soil landscape modeling: geographic information technologies and Pedometrics. Taylor and Francis, Boca Raton, pp 61–104

Brungard CW, Boettinger JL, Duniway MC, Wills SA, Edwards TC Jr (2015) Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239–240:68–83

Brus DJ, de Gruijter JJ, Walvoort DJJ, de Vries F, Bronswijk JJB, Römkens PFAM, de Vries W (2002) Mapping the probability of exceeding critical thresholds for cadmium concentrations in soils in the Netherlands. J Environ Qual 31:1875–1884. doi:10.2134/jeq2002.1875

Bui EN (2004) Soil survey as a knowledge system. Geoderma 120(1–2):17–26

Bui EN, Moran CJ (2003) A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. Geoderma 111(1–2):21–44

Burgess TM, Webster R (1980a) Optimal interpolation and isarithmic mapping of soil properties: I. The semi-variogram and punctual kriging. J Soil Sci 31:315–331. doi:10.1111/j.1365-2389.1980.tb02084.x

Burgess TM, Webster R (1980b) Optimal interpolation and isarithmic mapping of soil properties: II. Block kriging. J Soil Sci 31:333–341. doi:10.1111/j.1365-2389.1980.tb02085.x

Burrough PA (1989) Fuzzy mathematical methods for soil survey and land evaluation. J Soil Sci 40:477–492. doi:10.1111/j.1365-2389.1989.tb01290.x
Burrough PA, McDonnell RA (1998) Principles of geographic information systems: spatial information systems and geostatistics. Oxford University Press, New York
Burrough PA, Beckett PHT, Jarvis MG (1971) Relation between cost and utility in soil survey. J Soil Sci 22(3):359–394
Burrough PA, van Gaans PFM, Hootsmans R (1997) Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77(2–4):115–135
Chaney NW, Wood EF, McBratney AB, Hempel JW, Nauman TW, Brungard CW, Odgers NP (2016) POLARIS: a 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274:54–67
Colwell JD (1970) A statistical-chemical characterization of four great soil groups in southern New South Wales based on orthogonal polynomials. Aust J Soil Res 8:221–238
Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37:35–46. doi:10.1016/0034-4257(91)90048-B
Cressie NAC (1993) Statistics for spatial data. Wiley, New York
Davies BE, Gamm SA (1970) Trend surface analysis applied to soil reaction values from Kent, England. Geoderma 3:223–231
de Brogniez D, Ballabio C, Stevens A, Jones RJA, Montanarella L, van Wesemael B (2015) A map of the topsoil organic carbon content of Europe generated by a generalized additive model. Eur J Soil Sci 66:121–134
Erh KT (1972) Application of spline functions to soil science. Soil Sci 114(5):333–338
FAO-UNESCO (1988) Soil map of the world: revised legend (with corrections and updates), World Soil Resources Report 60. Rome, Food and Agriculture Organisation
Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89(1–2):1–45
Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis. Geoderma 146(1–2, 102):–113
Grundy MJ, Viscarra Rossel RA, Searle RD, Wilson PL, Chen C, Gregory LJ (2015) Soil and landscape grid of Australia. Soil Res 53:835–844. doi:10.1071/SR15191
Grunwald S (2009) Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152(3–4):195–207
Grunwald S, Lamsal S (2006) The impact of emerging geographic information technology on soil-landscape modelling. In: Grunwald S (ed) Environmental soil landscape modeling: geographic information technologies and Pedometrics. Taylor and Francis, Boca Raton, pp 127–154
Grunwald S, McSweeney K, Rooney DJ, Lowery B (2001) Soil layer models created with profile cone penetrometer data. Geoderma 103(1–2):181–201
Grunwald S, Thompson JA, Boettinger JL (2011) Digital soil mapping and modeling at continental scales: finding solutions for global issues. Soil Sci Soc Am J 75:1201–1213. doi:10.2136/sssaj2011.0025
Häring T, Dietz E, Osenstetter S, Koschitzki T, Schröder B (2012) Spatial disaggregation of complex soil map units: a decision-tree based approach in Bavarian forest soils. Geoderma 185–186:37–47. doi:10.1016/j.geoderma.2012.04.001
Hartemink AE, Hempel J, Lagacherie P, McBratney AB, McKenzie NJ, MacMillan RA, Minasny B, Montanarella L, Mendonca-Santos ML, Sanchez P, Walsh M, Zhang G (2008) GlobalSoilMap.net- a new digital soil map of the world. In: Boettinger JL, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S (eds) Digital soil mapping: bridging research, environmental application, and operation. Springer Science, Dordrecht
Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman and Hall/CRC, London
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. Springer-Verlag, New York

Helmick JL, Nauman TW, Thompson JA (2014) Developing and assessing prediction intervals for soil property maps derived from legacy databases. In: Arrouays D, McKenzie NJ, Hempel JW, Richer de Forges AC, McBratney AB (eds) GlobalSoilMap: basis of the global spatial soil information system. Taylor & Francis, London, pp 359–366

Hempel JW, Libohova Z, Thompson JA, Odgers NP, Smith CAS, Lelyk GW, Geraldo GEE (2014) GlobalSoilMap north American node progress. GlobalSoilMap: basis of the global spatial soil information system – proceedings of the 1st GlobalSoilMap conference, p 41–45

Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG, Gonzalez MR (2014) SoilGrids1km ? Global soil information based on automated mapping. PLoS One 9(8):e105992

Heung B, Ho HC, Zhang J, Knudby A, Bulmer CE, Schmidt MG (2016) An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265:62–77

Heuvelink GBM, Huisman JA (2000) Choosing between abrupt and gradual spatial variation? In: Mowrer HT, Congalton RG (eds) Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing. Ann Arbor Press, Chelsea, pp 111–117

Heuvelink GBM, Webster R (2001) Modelling soil variation: past, present, and future. Geoderma 100(3–4):269–301

Hewitt AE (1993) Predictive modelling in soil survey. Soil Fertilizers 56:305–314

Hewitt AE, McKenzie NJ, Grundy MJ, Slater BK (2008) Qualitative survey. In: McKenzie NJ, Grundy MJ, Webster R, Ringrose-Voase AJ (eds) Guidelines for surveying soil and land resources. CSIRO Publishing, Collingwood, pp 285–306

Hewitt AE, Barringer JRF, Forrester GJ, McNeill SJ (2010) Soilscapes basis for digital soil mapping in New Zealand. In: Boettinger JL, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S (eds) Digital soil mapping: bridging research, environmental application, and operation, progress in soil science. Springer, p 297–307. doi:10.1007/978-90-481-8863-5_24

Hiederer R (2009) Distribution of organic carbon in soil profile data, EUR 23980 EN. Office for Official Publications of European Communities, Luxembourg

Hole FD (1953) Suggested terminology for describing soils as three-dimensional bodies. Proc Soil Sci Soc Am 17:131–135

Hole FD (1961) A classification of pedoturbations and some other processes and factors of soil formation in relation to isotropism and anisotropism. Soil Sci 91(6):375–377

Holmes KW, Griffin EA, Odgers NP (2015) Large-area spatial disaggregation of legacy soil maps: evaluation over Western Australia. Soil Res 53:865–880. doi:10.1071/SR14270

Hong SY, Kim YH, Han KH, Hyun BK, Zhang YS, Song KC, Minasny B, McBratney AB (2012) Digital soil mapping of soil properties for Korean soils. Digital soil assessments and beyond – proceedings of the fifth global workshop on digital soil mapping, p 435–438

Hudson BD (1992) The soil survey as paradigm-based science. Soil Sci Soc Am J 56(3):836–841

Isbell RF (1996) The Australian soil classification. CSIRO Publishing, Melbourne, 152 pp

IUSS Working group WRB (2007): World reference base for soil resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome

Kellogg CE (1949) Soil classification- introduction. Soil Sci 67(2):77–80

Kempen B, Brus DJ, Heuvelink GBM, Stoorvogel JJ (2009) Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma 151(3–4):311–326

Kempen B, Brus DJ, Stoorvogel JJ (2011) Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. Geoderma 162(1–2):107–123

Kempen B, Brus DJ, Heuvelink GBM (2012) Soil type mapping using the generalised linear geostatistical model: a case study in a Dutch cultivated peatland. Geoderma 189:540–553

Kempen B, Brus DJ, de Vries F (2015) Operationalizing digital soil mapping for nationwide updating of the 1:50,000 soil map of the Netherlands. Geoderma 241–242:313–329

Kerry R, Goovaerts P, Rawlins BG, Marchant BP (2012) Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. Geoderma 170:347–358. doi:10.1016/j.geoderma.2011.10.007

Kidd DB, Webb MA, Grose CJ, Moreton RM, Malone BP, McBratney AB, Minasny B (2014) Operational digital soil assessment for enterprise suitability in Tasmania, Australia. In: Arrouays D, McKenzie NJ, Hempel JW, Richer de Forges A, McBratney AB (eds) GlobalSoilMap: basis of the global spatial soil information system. Taylor & Francis, London, pp 113–119

Kidd D, Webb M, Malone B, Minasny B, McBratney A (2015a) Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. Geoderma Reg 6:7–21

Kidd D, Webb M, Malone B, Minasny B, McBratney A (2015b) Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. Soil Res 53:932–955

Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, R Core Team., Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C (2016) Caret: classification and regression training. R Core Team, http://www.R-project.org/

Lacoste M, Mulder VL, Richer-de-Forges AC, Martin MP, Arrouays D (2016) Evaluating large-extent spatial modeling approaches: a case study for soil depth for France. Geoderma Reg 7:137–152

Lagacherie P (2008) Digital soil mapping: a state of the art. In: Hartemink AE, McBratney AB, Mendonca-Santos ML (eds) Digital soil mapping with limited data. Springer Science, Australia, pp 3–14

Lagacherie P, Holmes S (1997) Addressing geographical data errors in a classification tree for soil unit prediction. Int J Geogr Inf Sci 11(2):183–198

Lagacherie P, McBratney AB (2007) Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: Lagacherie P, McBratney AB, Voltz M (eds) Digital soil mapping—an introductory perspective, developments in soil science, vol 31. Elsevier B.V., Amsterdam, pp 3–22

Lagacherie P, Legros JP, Burrough PA (1995) A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. Geoderma 65(3–4):283–301

Lagacherie P, Cazemier DR, vanGaans PFM, Burrough PA (1997) Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area. Geoderma 77(2–4):197–216

Lal R (2004) Soil carbon sequestration to mitigate climate change. Geoderma 123(1–2):1–22

Lark RM, Cullis BR, Welham SJ (2006) On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Eur J Soil Sci 57(6):787–799

Laslett GM, McBratney AB, Pahl PJ, Hutchinson MF (1987) Comparison of several spatial prediction methods for soil pH. J Soil Sci 38(2):325–341

Lin LI-K (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255–268. doi:10.2307/2532051

Mallavan BP, Minasny B, McBratney AB (2010) Homosoil: amethodology for quantitative extrapolation of soil information across the globe. In: Boettinger JL, Howell DW, More AC, Hartemink AE, Kienast-Brown S (eds) Digital soil mapping: bridging research, environmental application, and operation. Springer, London, pp 137–149

Malone BP, McBratney AB, Minasny B, Laslett GM (2009) Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154(1–2):138–152

Malone BP, McBratney AB, Minasny B (2011) Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160(3–4):614–626

Malone B, Minasny B, Odgers N, McBratney A (2014) Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232–234:34–44

Malone B, Jha SK, Minasny B, McBratney A (2016) Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. Geoderma 262:243–253

McBratney AB, de Gruijter JJ (1992) A continuum approach to soil classification by modified fuzzy k-means with extragrades. J Soil Sci 43(1):159–175

McBratney AB, Webster R (1983) Optimal interpolation and isarithmic mapping of soil properties. 5. co-regionalisation and multiple sampling strategy. J Soil Sci 34(1):137–162

McBratney AB, Degruijter JJ, Brus DJ (1992) Spatial prediction and mapping of continuous soil classes. Geoderma 54(1–4):39–64
McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM (2000) An overview of pedometric techniques for use in soil survey. Geoderma 97(3–4):293–327
McBratney AB, Mendonca-Santos ML, Minasny B (2003) On digital soil mapping. Geoderma 117(1–2):3–52
McKenzie NJ, Austin MP (1993) A quantitative Australian approach to medium and small-scale surveys based on soil stratigraphy and environmental correlation. Geoderma 57(4):329–355
McKenzie NJ, Ryan PJ (1999) Spatial prediction of soil properties using environmental correlation. Geoderma 89(1–2):67–94
McSweeney K, Slater BK, Hammer RD, Bell JC, Gessler PE, Petersen GW (1994) Towards a new framework for modeling the soil-landscape continuum. In: Factors of soil formation: a fiftieth anniversary retrospective, SSSA Special Publication. Soil Science Society of America, Madison, p 127–145
Meersmans J, van Wesemael B, De Ridder F, Van Molle M (2009) Modelling the three-dimensional spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium). Geoderma 152(1–2):43–52
Meinshausen N (2006) Quantile regression forests. J Mach Learn Res 7:983–999
Miller DA, White RA (1998) A conterminous United States multilayer soil characteristics dataset for regional climate and hydrological modeling. Earth Interact 2:1–26
Minasny B, McBratney AB (2007) Spatial prediction of soil properties using EBLUP with the Matern covariance function. Geoderma 140(4):324–336
Minasny B, McBratney AB (2010) Methodologies for global soil mapping. In: Boettinger JL, Howell DW, More AC, Hartemink AE, Kienast-Brown S (eds) Digital soil mapping: bridging research, environmental application, and operation. Springer, London, pp 429–436
Minasny B, McBratney AB (2016) Digital soil mapping: a brief history and some lessons. Geoderma 264(Part B):301–311
Minasny B, McBratney AB, Mendonca-Santos ML, Odeh IOA, Guyon B (2006) Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. Aust J Soil Res 44(3):233–244
Minasny B, McBratney AB, Lark RM (2008) Digital soil mapping technologies for countries with sparse data infrastructures. In: Hartemink AE, McBratney AB, Mendonca-Santos MD (eds) Digital soil mapping with limited data. Springer, Australia, pp 15–30
Minasny B, Vrugt JA, McBratney AB (2011) Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation. Geoderma 163(3–4):150–162
Minasny B, Malone B, McBratney A (eds) (2012) Digital soil assessments and beyond. Digital soil assessments and beyond. CRC Press, Leiden
Mishra U, Lal R, Slater B, Calhoun F, Liu D, Van Meirvenne M (2009) Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. Soil Sci Soc Am J 73(2):614–621
Moore AW, Russell JS, Ward WT (1972) Numerical analysis of soils: a comparison of three soil profile models with field classification. J Soil Sci 23:193–209
Moore ID, Gessler PE, Nielsen GA, Peterson GA (1993) Soil attribute prediction using terrain analysis. Soil Sci Soc Am J 57(2):443–452. doi:10.2136/sssaj1993.03615995005700020026x
Moore AC, Howell DW, Haydu-Houdeshell C, Blinn C, Hempel J, Smith D (2010) Building digital soil mapping capacity in the Natural Resources Conservation Service: Mojave Desert operational initiative. In: Boettinger JL, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S (eds) Digital soil mapping: bridging research, environmental application, and operation, progress in soil science. Springer, Dordrecht, pp 357–367. doi:10.1007/978-90-481-8863-5_28
Mulder VL, de Bruin S, Schaepman ME, Mayr TR (2011) The use of remote sensing in soil and terrain mapping—a review. Geoderma 162(1–2):1–19 1235
Mulder VL, Lacoste M, Richer-de-Forges AC, Arrouays D (2016) GlobalSoilMap France: high-resolution spatial modelling the soils of France up to two meter depth. Sci Total Environ 573:1352–1369

Myers DB, Kitchen NR, Sudduth KA, Miles RJ, Sadler EJ, Grunwald S (2011) Peak functions for modeling high resolution soil profile data. Geoderma 166(1):74–83

Nauman TW, Thompson JA (2014) Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213:385–399

Nettleton WD, Brasher BR, Borst G (1991) The tax adjunct problem. Soil Sci Soc Am J 55(2):421–427

Nortcliff S (1978) Soil variability and reconnaissance soil mapping- statistical study in Norfolk. J Soil Sci 29(3):403–418

Odeh IOA, McBratney AB, Chittleborough DJ (1992) Soil pattern recognition with fuzzy c-means: application to classification and soil-landform inter-relationships. Soil Sci Soc Am J 56(2):505–516

Odeh IOA, McBratney AB, Chittleborough DJ (1994) Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63(3–4):197–214

Odeh IOA, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes- heterotopic co-kriging and regression kriging. Geoderma 67(3–4):215–226

Odgers NP, McBratney AB, Minasny B (2011) Bottom-up digital soil mapping. II Soil series classes. Geoderma 163:30–37. doi:10.1016/j.geoderma.2011.03.013

Odgers NP, Libohova Z, Thompson JA (2012) Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189–190:153–163

Odgers NP, Sun W, McBratney AB, Minasny B, Clifford D (2014) Disaggregating and harmonising soil map units through resampled classification trees. Geoderma 214–215:91–100

Odgers NP, Holmes KW, Griffin T, Liddicoat C (2015a) Derivation of soil-attribute estimations from legacy soil maps. Soil Res 53:881–894

Odgers NP, McBratney AB, Minasny B (2015b) Digital soil property mapping and uncertainty estimation using soil class probability rasters. Geoderma 237–238:190–198

Orton TG, Pringle MJ, Bishop TFA (2016) A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262:174–186

Padarian J, Minasny B, McBratney AB (2017) Chile and the Chilean soil grid: a contribution to GlobalSoilMap. Geoderma Reg 9:17–28

Poggio L, Gimona A (2014) National scale 3D modelling of soil organic carbon stocks with uncertainty propagation—an example from Scotland. Geoderma 232–234:284–299

Ponce-Hernandez R, Marriott FHC, Beckett PHT (1986) An improved method for reconstructing a soil-profile from analysis of a small number of samples. J Soil Sci 37(3):455–467

Protz R, Presant EW, Arnold RW (1968) Establishment of the modal profile and measurement of variability within a soil landform unit. Can J Soil Sci 48:7–19. doi:10.4141/cjss68-002

Quinlan JR (1993a) C4.5: programs for machine learning. Morgan Kaufmann, San Mateo

Quinlan JR (1993b) Combining instance-based and model-based learning. In: Utgoff (ed) Machine learning '93. Morgan Kaufmann, San Mateo

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/

Ruhe RV, Walker PH (1968) Hillslope models and soil formation. I. Open systems. Presented at the transactions of the 9th international congress of soil science, Adelaide, South Australia, p 551–560

Russell JS, Moore AW (1968) Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data. Trans 9th Int Congr Soil Sci 4:205–213

Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonça-Santos MDL, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vågen T-G, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang G-L (2009) Digital soil map of the world. Science 325:680–681

Scull P, Franklin J, Chadwick OA, McArthur D (2003) Predictive soil mapping: a review. Prog Phys Geogr 27(2):171–197

Shovic HF, Montagne C (1985) Application of a statistical soil-landscape model to an order III wildland soil survey. Soil Sci Soc Am J 49:961–968. doi:10.2136/sssaj1985.03615995004900040034x
Simonson RW (1989) Historical highlights of soil survey and soil classification with emphasis on the United States, 1899–1970. International Soil Reference and Information Centre Technical Paper 18, Wageningen, The Netherlands
Skidmore AK, Ryan PJ, Dawes W, Short D, O'loughlin E (1991) Use of an expert system to map forest soils from a geographic information system. Int J Geogr Inf Syst 5(4):431–445
Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222
Soil Survey Staff (1993) Soil survey manual, U. S. Department of Agriculture Handbook 18. United States Department of Agriculture Soil Conservation Service
von Steiger B, Webster R, Schulin R, Lehmann R (1996) Mapping heavy metals in polluted soil by disjunctive kriging. Environ Pollut 94:205–215. doi:10.1016/S0269-7491(96)00060-7
Stein A, Corsten LCA (1991) Universal kriging and co-kriging as a regression procedure. Biometrics 47(2):575–587
Story M, Congalton RG (1986) Accuracy assessment: a user's perspective. Photogramm Eng Remote Sens 52:397–399
Subburayalu S, Jenhani I, Slater BK (2014) Disaggregation of component soil series using possibilistic decision trees from an Ohio County soil survey map. Geoderma 213:334–345. doi:10.1016/j.geoderma.2013.08.018
Taalab K, Corstanje R, Zawadzka J, Mayr T, Whelan MJ, Hannam JA, Creamer R (2015) On the application of Bayesian networks in digital soil mapping. Geoderma 259–260:134–148
Taghizadeh-Mehrjardi R, Nabiollahi K, Kerry R (2016) Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma 266:98–110
Thompson JA, Odgers NP, Libohova Z, Waltman SW, D'Avello T (2011) Disaggregation of polygon soil class maps to produce raster soil property maps using digital soil mapping techniques. Presented at the 2011 ASA-CSSA-SSSA International Annual Meetings, San Antonio, Texas
Triantafilis J, Ward WT, Odeh IOA, McBratney AB (2001) Creation and interpolation of continuous soil layer classes in the lower Namoi valley. Soil Sci Soc Am J 65:403–413
Vaysse K, Lagacherie P (2017) Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291:55–64
Veronesi F, Corstanje R, Mayr T (2012) Mapping soil compaction in 3D with depth functions. Soil Tillage Res 124:111–118
Vincent S, Lemercier B, Berthier L, Walter C (2016) Spatial disaggregation of complex soil map units at the regional scale based on soil-landscape relationships. Geoderma
Viscarra Rossel RA, Chen C, Grundy MJ, Searle R, Clifford D, Campbell PH (2015) The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res 53(8):845–864
Wackernagel H (1998) Multivariate geostatistics: an introduction with applications. Springer, Berlin
Webster R, Burgess TM (1980) Optimal interpolation and isarithmic mapping of soil properties: III. Changing drift and universal kriging. J Soil Sci 31:505–524. doi:10.1111/j.1365-2389.1980.tb02100.x
Webster R, Cuanalo de la Cerda HE (1975) Soil transect correlograms of North Oxfordshire and their interpretation. J Soil Sci 26(2):176–194
Webster R, Oliver MA (2006) Modeling spatial variation of soil as random functions. In: Grunwald S (ed) Environmental soil landscape modeling: geographic information technologies and pedometrics. Taylor and Francis, Boca Raton, pp 241–288
Wilding LP, Jones RB, Schafer GM (1965) Variation of soil morphological properties within Miami, Celina, and Crosby mapping units in west-central Ohio. Soil Sci Soc Am Proc 29:711–717. doi:10.2136/sssaj1965.03615995002900060033x
Wright JK (1944) The terminology of certain map symbols. Geogr Rev 34:653–654

Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353
Zhu AX, Band LE (1994) A knowledge-based approach to data integration for soil mapping. Can J Remote Sens 20:408–418
Zhu AX, Band LE, Dutton B, Nimlos TJ (1996) Automated soil inference under fuzzy logic. Ecol Model 90:123–145. doi:10.1016/0304-3800(95)00161-1
Zhu AX, Band L, Vertessy R, Dutton B (1997) Derivation of soil properties using a soil land inference model (SoLIM). Soil Sci Soc Am J 61(2):523–533
Zhu AX, Hudson B, Burt JE, Lubich K, Simonson D (2001) Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Sci Soc Am J 65:1463–1472

Chapter 13
Vis-NIR-SWIR Remote Sensing Products as New Soil Data for Digital Soil Mapping

Philippe Lagacherie and Cécile Gomez

"If you were born without wings, do nothing to prevent them from growing".

Coco Chanel

13.1 Introduction

Since the early ages of soil surveys, air photographs have been intensively used by soil surveyors for depicting the soil variations across landscapes. The variations of soil surfaces, specifically color and ratio of vegetation cover, that were revealed by this early remote sensing product were a great help for interpolating the scarce soil observations and for delineating the soil class boundaries. This was further transposed in digital soil mapping (McBratney et al. 2003), thanks to the large availability of remote sensing images provided by the emerging spatial data infrastructures. Up to now, digital soil mappers have mainly used remote sensing images as spatial data inputs for representing the landscape variables that are related with soil, such as vegetation and parent material (the soil covariates). Boettinger et al. (2008) reviewed the main indicators that could be retrieved for estimating these soil covariates, using multispectral data acquired in the visible near-infrared and short-wave infrared (VIS, 400–700 nm; NIR, 700–1100 nm; SWIR, 1100–2500 nm) spectral domain. After a spatial overlay with the sparse sets of observed and measured sites collected in a given area, the indicators derived from remote sensing have been used as independent variables in regression-like models or as external drift in geostatistic models (McBratney et al. 2003, Chap. 12 of this book).

P. Lagacherie (✉) • C. Gomez
LISAH, Univ Montpellier, INRA, IRD, Montpellier SupAgro, 34060 Montpellier, France
e-mail: philippe.lagacherie@inra.fr; cecile.gomez@ird.fr

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_13

The modern airborne hyperspectral sensors (such as the HyMap sensor and the AISA-Dual sensor) reach spectral resolutions (5–10 nm) that become closer and closer to the laboratory spectrometers (1 nm). Moreover the near future launching of hyperspectral satellites (such as the German EnMAP sensor (Guanter et al. 2015) and the French HYPXIM sensor (Briottet et al. 2013)) would provide additional hyperspectral data sources. Based on the large successful attempts of soil property estimations using Vis-NIR-SWIR laboratory data (Viscarra Rossel et al. 2006), and the more modest successful attempts of mapping soil properties on bare soil surfaces using Vis-NIR-SWIR hyperspectral images (Ben-Dor et al. 2008), the latter can now be envisaged as a potential cost-efficient way for acquiring soil property measurements. Thus, it is expected that Vis-NIR-SWIR hyperspectral data can partially overcome the lack of soil data that has been acknowledged as the main limitation of digital soil mapping performances (Lagacherie et al. 2008). Some recent promising results have been obtained that describe complex spatial patterns of soil properties (Fig. 13.1, Ben-Dor et al. 2006; Schwanghart and Jarmer 2011; Gomez et al. 2012a).

This technological evolution leads to a full revision on how the (digital) soil mapper can use remote sensing products in their activities. This chapter describes the pedometric tools, their applications, and their limitations for producing estimations of soil properties from Vis-NIR-SWIR imagery, denoted further "VNS-I soil products" (Sect. 13.2). It then details how these VNS-I soil products constitute new sources of soil data and can be used to improve the mapping of soil properties across landscapes (Sect. 13.3).

13.2 Soil Property Estimated from Vis-NIR-SWIR Imagery

Some introductive notions about soil spectral signatures are presented in Sect. 13.2.1 (see also Chaps. 5 and 7 of this book). Then the developed approaches for soil property estimation are summarized in Sect. 13.2.2, and their performances are presented in Sect. 13.2.3, both from laboratory to imaging spectrometry.

13.2.1 Introductive Notions

The spectral signature of a material in the Vis-NIR-SWIR (VIS, 400–700 nm; NIR, 700–1100 nm; SWIR, 1100–2500 nm) spectral domain is characterized by (1) its general shape; (2) its reflectance intensity, also called albedo; and (3) its specific absorption bands. The absorption phenomenon results from electronic and vibrational transitions due to the absorption of specific photon energy within ionic and molecular components of the observed material. Spectral characteristics of soils have been largely studied in laboratory VIS-NIR-SWIR spectroscopy for decades

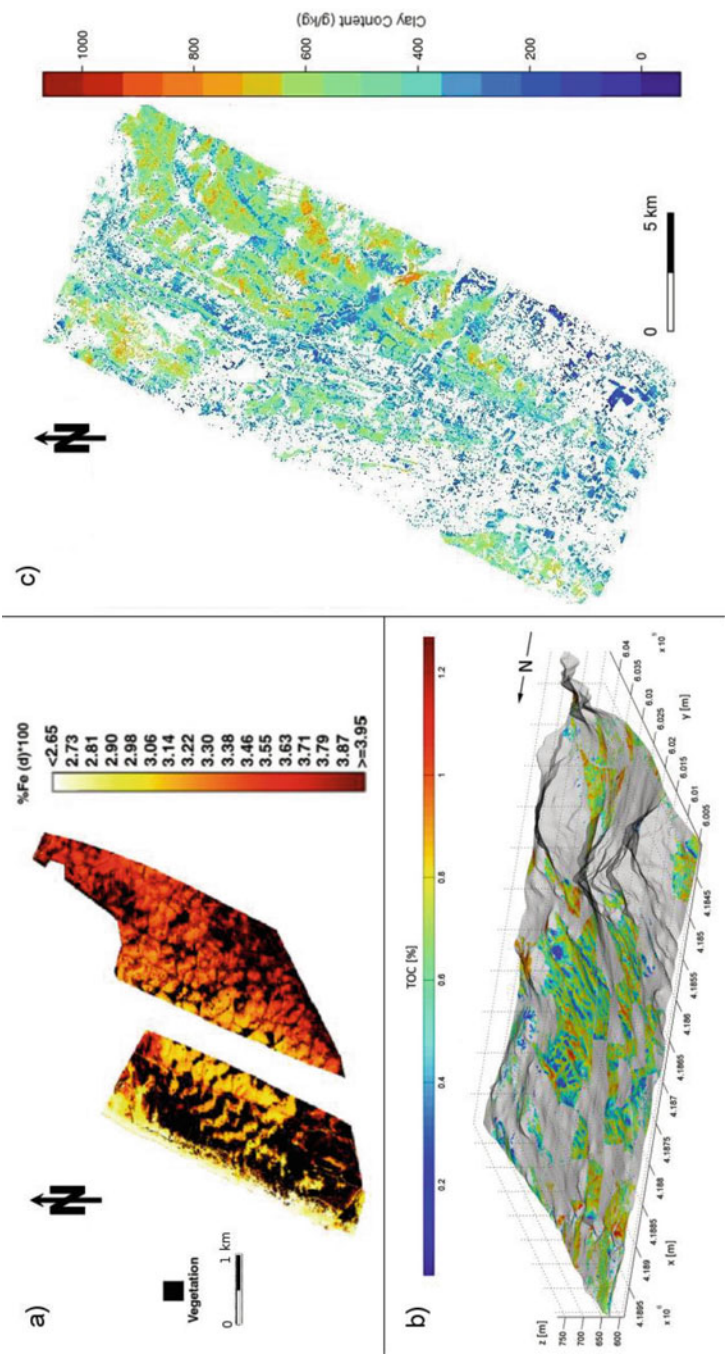


Fig. 13.1 Examples of spatial patterns of soil properties revealed by Vis-NIR-SWIR imagery: (**a**) map of estimated free iron oxides (*100) over the sand dune area of Ashdod, Israel (After Ben-Dor et al. 2006), (**b**) map of estimated total organic carbon (TOC) contents (After Schwanghart and Jarmer 2011), and (**c**) map of estimated clay contents over 300 km^2 of Lebna catchment in Tunisia (After Gomez et al. 2012a). No-soil surfaces (vegetated, water, and urban areas) are not mapped (in *black* in **a**, in *gray* in **b**, and in *white* in **c**)

and described, among others, by, e.g., Ben-Dor et al. (1998). The following section presents a summary of these spectral characteristics.

The general shape of the reflectance spectrum through a change of spectral intensity and absorption bands depth may be influenced by the soil structure linked to its particle size, the sample geometry, the view angle, and the geometry of the light source (incident angle and azimuth angle) (Baumgardner et al. 1985; Ben-Dor and Banin 1995). In particular, the more the soil grain size increases, the more the surface of spectral distribution decreases, and so the more the reflectance spectrum intensity decreases (Baumgardner et al. 1985). Thus, a spectrum of clay soil will tend to have more albedo than a spectrum of sandy soil. Finally, the soil structure usually doesn't affect the absorption band's position (Ben-Dor and Banin 1995).

Three major chemical components affect the soil spectrum, through specific absorption bands: minerals (clay, ferric, and carbonate), organic matter (OM), and water (Stoner and Baumgardner 1981). Calcium carbonate ($CaCO_3$) in soil causes an absorption band around 2340 nm, due to vibrations at the molecular level associated with the C-O bond (Gaffey 1987). Clay minerals such as kaolinite, montmorillonite, and illite in the soil induce an absorption band around 2200 nm due to the combination of vibrations associated with the OH bond and the OH-Al-OH bonds (e.g., Hunt et al. 1971; Chabrillat et al. 2002). Other clay minerals have specific spectral responses in the SWIR spectral domain, relative to vibrations associated with Mg and Fe-OH bonds. Iron in its trivalent state produces three absorption bands between 400 and 1000 nm due to electronic processes. The hematite is characterized by absorption bands centered at 550, 630, and 860 nm, and goethite is characterized by absorption bands centered at 480, 650, and 920 nm (Hunt et al. 1971; Morris et al. 1985). The absorption bands around 500 nm and 900 nm dominate the reflectance in the visible spectrum, while the absorption band around 640 nm is seldom marked.

Organic matter (OM) is an important soil component and has a strong influence on soil reflectance characteristics throughout the entire Vis-NIR-SWIR domain (i.e., Tables 13.2 and 13.3 in Ben-Dor et al. 1997). The vibrations of O-H and C-H bonds cause absorptions, particularly due to lignin and cellulose (i.e., Tables 13.2 and 13.3 in Ben-Dor et al. 1997). The physicochemical complexity of the soil organic matter (fresh or dry) explains the difficulty in precisely identifying specific absorption bands. Nevertheless, OM has a low effect on the spectrum when its content is below 2% (Baumgardner et al. 1970).

Water content in soil induces three phenomena: (i) a general reduction in reflectance along the spectrum due to light absorption by water, (ii) a reduction or increase of absorption bands associated to minerals (e.g., carbonate around 2340 nm and for clay around 2200 nm, respectively), and (iii) the appearance of absorption bands at 1400 and 1900 nm due to the presence of liquid water (H_2O) or associated with the OH bond in soil (Bowers and Hanks 1965).

Some primary soil properties, such as pH and cation-exchange capacity (CEC), have no particular spectral characteristics and therefore cannot be determined by direct spectral analysis.

13.2.2 Soil Property Estimation Approaches

Quantitative estimation of soil properties can be achieved by developing estimation models called spectrotransfer functions (STFs). These STFs link a set of Y-dependent variables (soil property) and a set of X predictor variables (Vis-NIR-SWIR spectrometric measurements). In this section, we present two approaches allowing STF development (Sects. 13.2.2.1 and 13.2.2.2) and the principle of calibration and validation of the STFs (Sect. 13.2.2.3).

13.2.2.1 Spectral Indices

Spectrotransfer functions for estimating some primary soil properties can be developed by analyzing absorption bands and spectrum geometry. The general principle is based on relationships between a spectral index and a soil property. Spectral indices are based on ratio of (at least) two reflectance values focusing on absorption band area, depth, and asymmetry or also slope between two reflectance values (Van Der Meer 2004). Some of these indices are calculated from the continuum removed reflectance. The continuum removed function is usually defined as a convex hull, but can also be defined as a straight line (often named baseline) if a single absorption feature is investigated (Kokaly and Clark 1999). Removing the convex hull allows to highlight absorption bands related to physicochemical properties, by removing the baseline ("continuum") that contains target measurement parameters (humidity differences between targets, illumination between shots, etc.). Table 13.1 presents some spectral indices related to primary soil properties. Initially dedicated to multispectral sensors such as the LANDSAT satellite sensor, some spectral indices, such as the brightness, color, and redness index (Table 13.1), can be used from all spectroscopic sensors (hyperspectral, superspectral, and multispectral, on field, laboratory, or by imaging). Conversely, some spectral indices, such as the SWIR_FI, CR_{clay}, CR_{CaCO3} and SOC, can be computed from hyperspectral sensor outputs only because they require fine spectral resolution.

As spectral indices are based on absorption bands and spectrum geometry, some soil properties such as pH and CEC that have not any specific spectral features may not be studied using spectral indices, which constitutes the major drawback of this STF approach. In the same time, spectral indices are independent from the study area, soil property distribution, and range, which constitute the main advantage of this STF approach.

Finally, spectral indices' development without calibration provides relative predictive values (e.g., Gomez et al. 2016). To obtain absolute values, a calibration process of these spectral indices is needed and is described in Sect. 13.2.2.3.

Table 13.1 Spectral indices related to primary soil properties

Index name	Associated soil property	Formula	Used wavelengths (nm)	Associated spectral feature	References
Brightness index	Iron oxide	$\sqrt{(B^2 + G^2 + R^2)/3}$	Where R, G, and B represent the reflectance at 630–690 nm (red), 520–600 nm (green), and 450–520 nm (blue), respectively	Average reflectance magnitude	Matthieu et al. (1998)
Color index	Iron oxide	$(R - G)/(R + G)$		Hematite and goethite band	Escadafal and Huete (1991) and Madeira et al. (1997)
Redness index	Iron oxide	$R^2/(B^* G^3)$		Hematite band	Madeira et al. (1997) and Pouget et al. (1991)
SWIR fine particles index (SWIR FI)	Textural clay	$\frac{(\lambda_1)^2}{\lambda_2{}^*(\lambda_3)^3}$	Where λ_1, λ_2, and λ_3 represent the reflectance at 2133 nm, 2225 nm, and 2209 nm, respectively	Absorption band at 2200 nm due to the combination of OH and OH-Al vibration modes	Levin et al. (2007)
CR_{clay}	Textural clay	$\min\left(1, \frac{R_{\lambda_3}}{R_{\lambda_2} + \frac{\lambda_1 - \lambda_3}{\lambda_1 - \lambda_2}(R_{\lambda_1} - R_{\lambda_2})}\right)$	Where λ_1, λ_2, and λ_3 represent the reflectance at 2143 nm, 2269 nm, and 2206 nm, respectively	Absorption band at 2200 nm due to the combination of OH and OH-Al vibration modes	Lagacherie et al. (2008)
CR_{CaCO3}	$CaCO_3$		Where λ_1, λ_2, and λ_3 represent the reflectance at 2275 nm, 2375 nm, and 2348 nm, respectively	Absorption band centered around 2340 nm due to vibrations at the molecular level associated with the CO bond	
SOC	Soil organic carbon	$\frac{1}{\mathrm{Sum}(\lambda_1 - \lambda_2)}$	Where λ_1 and λ_2 represent the reflectance at 2138 and 2209 nm, respectively	Biochemical components (lignin)	Bartholomeus et al. (2008)

13.2.2.2 Regression Models

Multivariate regression methods based on building statistical models to estimate the soil property, using a calibration base, are largely used in Vis-NIR-SWIR spectroscopy, from laboratory to field (Chap. 5, Sect. 5.5.2.1 of this book) and remote sensing. Two multivariate regression methods are commonly used:

– Multiple linear regression (MLR) is used for both multispectral (e.g., by Nanni and Demattê 2006) and hyperspectral (e.g., by Ben-Dor et al. 2002) data from remote sensing. This method is useful when the Vis-NIR-SWIR reflectance spectra are acquired in a small number of wavelengths (multispectral data case) and when there is little collinearity between reflectance values at each wavelength. The MLR regression model used to estimate a $\widehat{y}$ soil property is written as

$$\widehat{y} = \Sigma_{i=1}^{N} b_i x_i + a$$

where N is the number of wavelength, b_i is the regression coefficient associated with the wavelength i, x_i is the reflectance value at the wavelength i, and a is the residue.

– Partial least squares regression (PLSR) is widely used in Vis-NIR-SWIR hyperspectral remote sensing (e.g., by Stevens et al. 2008). This method allows the management of (1) colinearity between the reflectance values at different wavelengths and (2) a number of predictors (wavelengths) that are larger than the number of samples composing the database [X, Y]. PLS regression is used to minimize the covariance between X (descriptive variables: wavelengths) and the dependent variable Y (soil property). The PLSR projects the variables X and Y in an area of reduced size (set of orthogonal vectors, called latent variables, maximizing the covariance between X and Y). A detailed description of the PLSR is given in Wold et al. (2001).

13.2.2.3 Cal/Val Models

The implementation of STFs often requires the establishment of a database [X, Y] consisting of Vis-NIR-SWIR reflectance spectra acquired by the sensor on bare soil surfaces (descriptive variable X) associated with the soil property measured on a laboratory soil sample collected on the same bare soil surfaces (dependent variable Y). There are no precise figures on the minimum number of samples needed to build a database for soil property estimation. Nevertheless, it is accepted that a data number [X, Y] inferior to 50 does not allow developing robust estimation model, whatever the size of the study area.

The use of a calibration [X_{cal},Y_{cal}] and validation [X_{val},Y_{val}] database (sub-databases of [X, Y]) allows the calibration of STFs presented above and their validations, respectively. The data allocation in the calibration [X_{cal},Y_{cal}] and

validation [X_{val},Y_{val}] database is an important and complex step. If the two sets ([X_{cal},Y_{cal}] and [X_{val},Y_{val}]) are too similar, the model may be too optimistic (underestimation of the prediction error). Inversely, if the two sets ([X_{cal},Y_{cal}] and [X_{val},Y_{val}]) are too different, the model may seem excessively inaccurate (overestimation of the prediction error).

To ensure model robustness, it is necessary to build a calibration database representative of the study area in terms of either soil properties or of spectral signatures. It is also necessary that when a validation database is available, it is homogeneous regarding the calibration database, again either in terms of soil properties or of spectral signatures. In theory, calibration [X_{cal},Y_{cal}] and validation [X_{val},Y_{val}] databases would have to fulfill the following criteria: (1) similar distribution of Y_{cal} and Y_{val} variables and (2) maximum spectral variability within X_{cal}, while the spectra remain in a realistic domain (avoiding outliers). These two criteria are difficult to meet at the same time. Different calibration [X_{cal},Y_{cal}] and validation [X_{val},Y_{val}] database construction methods exist, relying on either of the above criteria. Similar distribution of Y_{cal} and Y_{val} variables is frequently used. The Y variables need to be sorted in ascending order. Then, both samples having the lowest Y soil properties are assigned to the calibration database. Then, the next sample is placed in the validation set, and the procedure is continued by alternately placing both following samples in the calibration set and the next sample in the validation set. The assignment of samples in each of the bases follows this process until the last sample, thereby constructing two uniform sets containing two thirds and one third of the samples, respectively. To ensure a significant spectral variability, methods such as Kennard-Stone (Kennard and Stone 1969) or DUPLEX (Snee 1977) can also be used.

13.2.3 Prediction Performances and Limitations, Using VIS-NIR-SWIR Imagery

13.2.3.1 Rules to Predict Soil Properties

The soil property estimation approaches presented above have been largely applied in laboratory spectrometry and can also be applied from reflectance spectra provided by VIS-NIR-SWIR imagery. It was highlighted that to be predicted, the soil property has to respect the following rules (Ben-Dor et al. 2002):

- The soil property has a specific spectral signature due to a chemical or physical structure or is correlated to a soil property having a specific spectral signature due to an associated chemical or physical structure.
- And the soil property has a wide range of values. Thus, the predictability of a primary soil property depends, among other things, on the soil diversity of the study site.

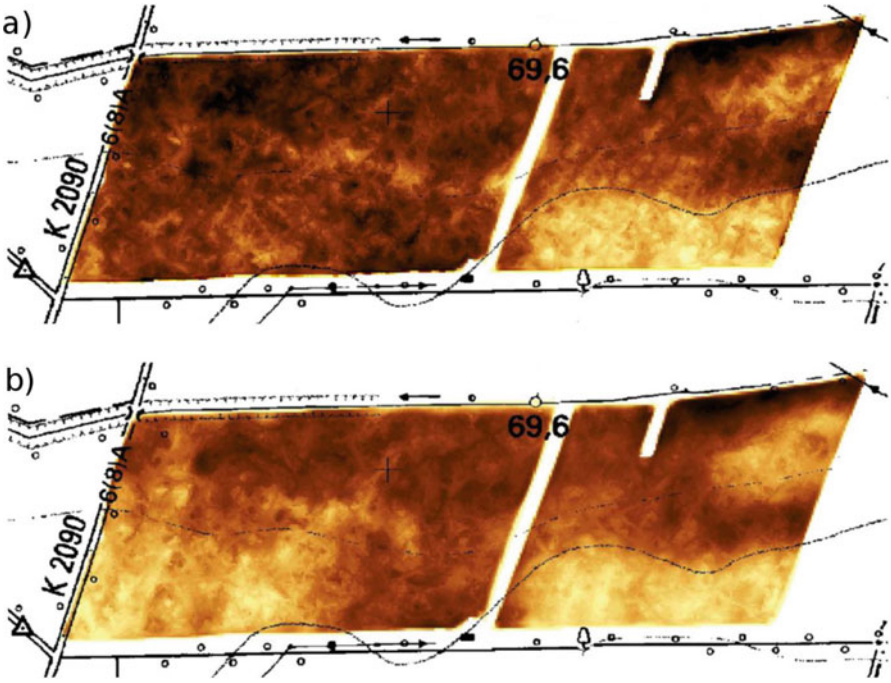


Fig. 13.2 Spatial distribution of predicted (**a**) organic carbon content and (**b**) total nitrogen content across a reference field, using the MLR method (Extracted from Selige et al. 2006)

The case of an accurate prediction (using STF and remote sensing data) of soil properties having specific spectral signatures can be illustrated by the successful mapping of predicted organic carbon in Selige et al. (2006) (Fig. 13.2a). An accurate prediction (using STF and remote sensing data) of soil properties without specific spectral signature but correlated to a soil property having a specific spectral signature due to an associated chemical or physical structure can be illustrated by the successful mapping of predicted total nitrogen (N_t) in Selige et al. (2006) (Fig. 13.2b). Indeed, total nitrogen (N_t) beneficiates from a high correlation (R^2 not communicated) between organic carbon and N_t over the study area (Selige et al. 2006). Finally, an inaccurate prediction (using STF and remote sensing data) of soil properties having specific spectral signatures can be illustrated by the result obtained by Gomez et al. (2012a) for SOC prediction over a Mediterranean context, due to poor values and variation of SOC (mean and standard deviation of 0.9% and 0.4%, respectively).

13.2.3.2 Performances from Lab to Imagery

In laboratory condition, air-dried and sieved soil samples in small cups are measured with laboratory spectrometers (e.g., ASD FieldSpec Pro FR, ASD AgriSpec, FOSS NIRSystem 6500 sensors) under controlled illumination conditions (e.g., using a contact probe or external halogen lamps) and using a Spectralon as a white reference. Drying and sieving the soil samples induce no surface roughness and no humidity perturbation during the spectral measurements (Table 13.2). Thus, all the soil samples belonging to a dataset receive the same pretreatments and have the same surface structure in the cups. Moreover, the same light incident angle and intensity affect spectral measurements of soil samples belonging to a dataset (Table 13.2). And the soil samples are composed of pure soil, so no vegetation residues affect spectral measurements (Table 13.2). Finally, laboratory spectrometers have a high spectral resolution ($\sim$1–3 nm at Vis and $\sim$3–10 at NIR-SWIR) and high signal-to-noise ratio (700:1 < < 2100:1).

In remote sensing conditions, raw soil surfaces are measured with imaging spectrometers (e.g., HyMap, AISA-Dual, CASI sensors) and at solar noon $\pm$2 h (to have maximum solar intensity associated to the minimum incident angle). Imaging raw soil surfaces induces the presence of surface elements (green vegetation, dry residues, coarse elements, etc.), soil roughness variability (which may be due to different tillage methods in cultivated areas), and surface moisture level variability among the targets of the same acquisition (Table 13.2). Acquisition duration, even limited at solar noon $\pm$2 h, induces illumination and incident angle variability within a remote sensing flight. The flight altitude induces atmosphere components in the recorded signals and thus induces spectral perturbations due to different gases (O_3, CO, CO_2, H_2O, CH_4, N_2O), which have to be corrected for (Table 13.2). Finally, imaging sensors have spectral resolutions from 4 nm (e.g., AISA-Dual sensor) to 30 nm (e.g., AHS-160 sensor) and signal-to-noise ratios from 200:1 (e.g., HySpex sensor at Vis-NIR domain) to 850:1 (e.g., AISA-Dual sensor at SWIR domain).

A decrease of performances from laboratory to remote sensing data has been demonstrated by several authors (see examples in Table 13.3). These performance decreases have been assigned to three main factors affecting the soil spectral measurement and so the STF accuracy (e.g., Stevens et al. 2008; Ben-Dor et al. 2009; Lagacherie et al. 2008): (i) the atmosphere, (iii) the disruptive factors affecting the soil areas, and (iii) a low signal-to-noise ratio of the imaging sensor.

Whatever the predicted soil property, an obvious limitation of these predictions is their low depth representativeness, as the depth of Vis-NIR-SWIR measurement is around the order of the wavelength, so around 1–2 μm. However, in cultivated areas where the superficial soil horizons are mixed and homogenized by tillage, the soil surface measured by Vis-NIR-SWIR imagery can be considered as representative of the whole cultivated horizon.

Table 13.2 Spectral measurement conditions for laboratory and imaging spectrometry

	Laboratory Spectrometry			**Imaging Spectrometry**
Measurement object	Soil Horizon samples			Pixels of soil surfaces
	Air-dried	Sieved	Pure soil	Raw
Consequences	No humidity perturbation	No surface roughness	No vegetation residues	Limited to topsoil properties Sensitive to within pixel surface variations (soil, land-use,…) Heterogeneity of vegetation residues, roughness and humidity
Illumination conditions	Controlled			Those of solar noon +-2h.
Consequences	No variation of incident light during spectral measurements			Variation of incident light during spectral imaging acquisition
Spectrometer specifications	• spectral resolution from ~1nm to 10 nm • high Signal to Noise ratio			• spectral resolution from 6 to 10 nm • modest Signal to Noise ratio
Consequences	Very fine description of spectral features			Fine description of spectral features
Atmosphere	None			Yes, and depending on the altitude of the aircraft, location and climate.
Consequences	No atmosphere perturbation			Atmosphere corrections needed (using, e.g., empirical line method , FLAASH model)

13.3 Using Vis-NIR-SWIR Imagery Soil Products in Digital Soil Mapping

Vis-NIR-SWIR imagery (VNS-I) soil products are a new source of quantitative characterization of soil that could contribute to overcome the lack of soil input that is recognized as the major limitation of Digital Soil mapping. In the following, different possible uses of such data in DSM are presented according to whether VNS-I

Table 13.3 Table of performances obtained from laboratory and imaging sensors, extracted from the literature

Properties	References	Method	Performance obtained from laboratory sensor *(sensor name)*	Performance obtained from imaging sensor *(sensor name)*
Iron	Bartholomeus et al. (2007)	Spectral index (redness index)	$R^2 = 0.48$ *(ASD FieldSpec)*	$R^2 = 0.22$ *(ROSIS)*
SOC	Gomez et al. (2008a)	PLSR	$R^2_{cv} = 0.66$ RMSECV = 6.1 g/kg *(ASD AgriSpec)*	$R^2_{cv} = 0.55$ RMSECV = 7.3 g/kg *(Hyperion)*
SOC	Stevens et al. (2008)	PLSR	R^2_{cv} not communicated RMSECV = 1.7 g/kg *(ASD FieldSpec)*	R^2_{cv} not communicated RMSECV = 1.2 g/kg *(AHS-160)*
Clay	Gomez et al. (2008b)	PLSR	$R^2_{cv} = 0.85$ RMSECV = 31.2 g/kg *(ASD FieldSpec)*	$R^2_{cv} = 0.64$ RMSECV = 49.6 g/kg *(HyMap)*
Clay	Gomez et al. (2008b)	Spectral index (CR)	$R^2_{cv} = 0.73$ RMSECV = 44 g/kg *(ASD FieldSpec)*	$R^2_{cv} = 0.58$ RMSECV = 82 g/kg *(HyMap)*
CaCO3	Gomez et al. (2008b)	PLSR	$R^2_{cv} = 0.94$ RMSECV = 38.3 g/kg *(ASD FieldSpec)*	$R^2_{cv} = 0.77$ RMSECV = 76.6 g/kg *(HyMap)*
CaCO3	Gomez et al. (2008b)	Spectral index (CR_{CaCO3})	$R^2_{cv} = 0.92$ RMSECV = 52 g/kg *(ASD FieldSpec)*	$R^2_{cv} = 0.47$ RMSECV = 132 g/kg *(HyMap)*

products are used in DSM models as substitutes of soil property measurements, as complements of soil property measurements, or as soil covariates.

13.3.1 Using VIS-NIR-SWIR Imagery Soil Products as Substitutes of Site Measurements

In this situation, the estimations of soil properties at each pixel of a hyperspectral image (see Sect. 13.2.3) is assumed to be precise enough to be considered as a measurement of a soil property and to be used as such for feeding DSM models. This potentially provides spatial sampling with densities largely greater than usually handled by DSM models, which has the advantage of better capturing soil variations, especially short-scale ones. However, these estimations convey uncertainties that may impact the model results.

Table 13.4 Statistic and variogram parameters estimated from different soil inputs in the Peyne Valley (Southern France, Hérault) after Gomez et al. (2012b)

Input	Mean (g/kg)	Standard deviation (g/kg)	Nugget (g^2/kg^2)	Sill (g^2/kg^2)	Range (m)
95 measured	233	74	2702	2315	1901
95 estimated	233	65	1422	2232	1928
37,320 estimated	248	74	0	2895	184
				2393	2297

In the following, we analyzed this trade-off by considering two basic operations of DSM: estimating means and variances of a soil property within a study area and building experimental variograms of a soil property. These two operations were performed for the Peyne watershed case study (Gomez et al. 2012b). Three different soil inputs were considered (i) a set of 95 sites with real measurements (wet laboratory analysis), (ii) the same set of 95 sites with estimates from VNS-I estimates, and (iii) the 37,320 sites with HI estimates at the bare soil pixels of the image. The results obtained for the two above-cited basic operations are presented in Table 13.4 for one example of soil property, namely, clay content, which has been estimated with fairly good precision ($R2 = 0.67$) from Vis-NIR-SWIR hyperspectral imagery. More results can be found for other soil properties in Gomez et al. (2012b).

In the example provided, the substitution of real measurements of clay content by their estimations from Vis-NIR-SWIR hyperspectral data over 95 sites has a moderate impact on calculated statistics and variogram parameters (first two lines of Table 13.4). The same mean values are registered, whereas standard deviation moderately decreases because of a loss of precision caused by the spectrotransfer function. This also induces differences in nugget values of the variograms, whereas these variograms are very similar with regard to range and sill.

Conversely, increasing the number of considered sites has a significant impact on calculated statistics and variogram parameters (last two lines of Table 13.4). A better coverage of the spatial sampling makes the estimated mean and standard deviation increase. Besides, the greater density of sites allows to reveal a more complex spatial structure, combining a short-range variation of clay content due to erosion-redeposition along slopes (range of 184 m) and a larger one (range of 2297) due to lithological variations across the study area. This result demonstrates a posteriori that the important nugget values obtained with 95 sites (first two lines of Table 13.4) are probably caused more by the inability to capture the real clay variations from a small number of sites than by any measurement uncertainties.

This example tends to demonstrate that substituting expensive wet laboratory analysis by VNS-I estimates could be a valuable solution in regions where dense spatial sampling is required for capturing short-range variations. This could be a step forward for the use of laboratory soil spectroscopy as soil data input of DSM models that has been observed in 16.7% of the DSM studies published in Geoderma and SSSAJ in 2007–2008 (Grunwald 2009).

13.3.2 Using VIS-NIR-SWIR Imagery Soil Products as Complements of Site Measurements

In some situations, the soil input for digital soil mapping includes both a VNS-I product that provides estimates of a given soil property at many locations (as the "VNS-I estimates" evoked in the previous section) and a sparser set of sites with the measurements of the same soil property. The latter data is often the calibration dataset of the spectrotransfer function used to produce the former. In these situations, the VNS-I estimates are used as auxiliary data that help with interpolating the measurements available for the sparse sites.

The suitable model for dealing with this input data is bivariate cokriging (Wackernagel 1995). In bivariate cokriging, the variable of interest (here a soil property) and its auxiliary variable (here the VNS-I product) are modeled by the random functions $Z(x)$ and $Y(x)$, respectively, where x denotes the location index. Let's suppose that $Z(x)$ is sampled at n_i sites x_i and $Y(x)$ is sampled at n_j sites x_j ($j \gg i$), the values at these locations being denoted further Z_i and Y_j. Let us then consider the covariance function of Z for a pair of points $C_{ZZ}(xi - xi')$(noted further $C_{i,i'}^{(ZZ)}$) and the cross covariance between Z and Y C_{ZY}(xi–xj) (denoted further $C_{i,j}^{(ZY)}$). Covariances and cross covariances are directly derived from fitted variograms and co-variograms. The cokriging predictor is the following:

$$Z^*(x_0) = \sum_{i=1}^{n_i} \lambda_i Z_i + \sum_{j=1}^{n_j} \lambda'_j Y_j \tag{13.1}$$

where the λ_i and the λ_j solve the following cokriging system with $n_i + n_j + 1$ equations to ensure unbiasedness and minimization of the MSE:

$$\begin{gathered} \sum_{i'=1}^{n_i} \lambda_{i'} C_{i'}^{(ZZ)} + \sum_{j=1}^{n_j} \lambda'_j C_{i,j}^{(ZY)} - \mu_1(x_i) = C_{i,0}^{(ZZ)} \\ \sum_{j'=1}^{n_j} \lambda'_{i'} C_{j,j'}^{(YY)} + \sum_{j=1}^{n_j} \lambda_i C_{i,j}^{(ZY)} - \mu_2(x_i) = C_{j,0}^{(ZY)} \\ \sum_{i=1}^{n_i} \lambda_i + \sum_{j=1}^{n_j} \lambda'_j = 1 \end{gathered} \tag{13.2}$$

Compared to the ordinary bivariate cokriging system, the constraints on λ_i and λ_j are summed up considering Z and Y have the same theoretical mean.

This approach was applied both in southern France (Lagacherie et al. 2012) and in northern Tunisia (Ciampalini et al. 2012). As an example, the latter application is summarized hereafter (Fig. 13.3). Cokriging was applied across a 339 km^2 area located in the Cap Bon region (northern Tunisia) with 262 sites with measured topsoil properties (clay, sand, iron contents, and cation-exchange capacity). The auxiliary variables were the estimated values of the same topsoil properties that were derived from an Vis-NIR-SWIR AISA-Dual hyperspectral image by means of a partial least square regression (PLSR) function calibrated from a subset of 129 sites (Gomez et al. 2012a). The hyperspectral image covered 42% of the

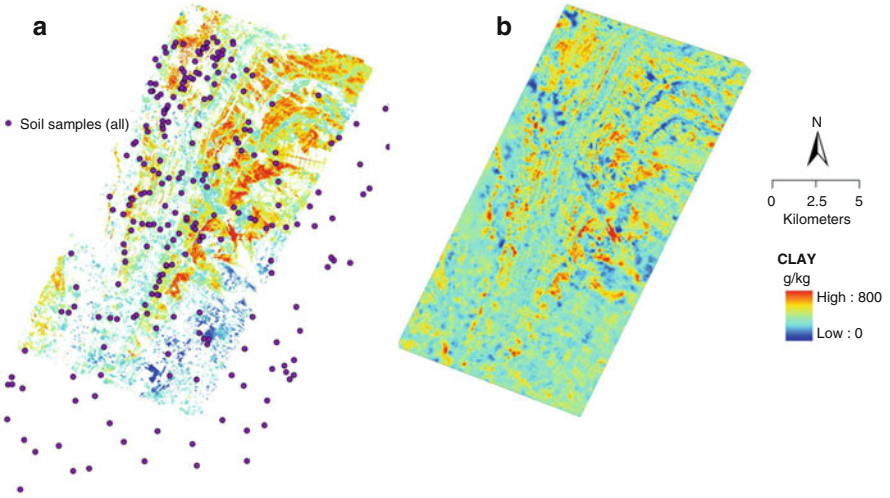


Fig. 13.3 Combining spatial sampling of measured sites with hyperspectral imagery estimates for mapping clay content in the Cap Bon region (After Ciampalini et al. 2012). (**a**) input data (VNS-I product), (**b**) final estimated clay map obtained by cokriging

study area, which allowed to get 156,693 estimated values of topsoil properties at 30×30 m pixels after averaging initial estimated values at 5×5 meter pixels for the sake of computing time. The cokriged map (Fig. 13.3, right) allowed to capture a soil pattern in the locations not covered by the hyperspectral image that was in good continuity with the one revealed by the hyperspectral image (Fig. 13.3, left). However, quantitative assessments of performances (by cross validation) showed that it remained a substantial uncertainty (R^2 between 0.31 and 0.34) in such locations, which can be jointly explained by (i) the initial uncertainty of the PLSR function (R^2 between 0.7 and 0.8), (ii) the short-scale soil variations of this lithology-driven Mediterranean study area, and (iii) the non-stationarities of soil properties' variations across the study area that were not taken into account by the co-regionalization model used by the cokriging.

13.3.3 Using VIS-NIR-SWIR Imagery Soil Products as Soil Covariates

The two previous subsections dealt only with soil properties that were directly predictable from Vis-NIR-SWIR images. A step further is to extent the use of such images to the larger set of soil properties that are assumed to be correlated with the latter. As a straightforward example, it can be considered that a VNS-I estimate of a given topsoil property can improve the predictions of the same soil properties but

considered at a greater depth. In that case, The VNS-I estimate is used as one of the covariates ("s" following the "scorpan" model of McBratney et al. (2003)) of a "surface-subsurface" soil model. This model will infer the value of a soil property S at the ith interval of depth (S_i) through the expression:

$$\begin{aligned} &S_i = S_1 + \Delta_i \ \ [1] \\ &\text{with } \Delta_i = S_i - S_1 = f\left(\{S_1, L\}\right) + \varepsilon_i \end{aligned} \tag{13.3}$$

where S_1 is the estimate of the topsoil surface property S obtained from a VNS-I product and $\{L\}$ is a set of easily available landscape covariates. The surface-subsurface soil model f is a statistical function that estimates the differences in the soil property values between the surface and the ith interval of depth (Δ_i) with a random error ε_i. The model requires a prior calibration from a dataset of existing measured soil profiles retrieved from a soil database.

Such a model was applied for a 339 km^2 study area located in the Cap Bon region, Northern Tunisia (Lagacherie et al. 2013). A set of 152 legacy-measured soil profiles collected in this region was used for calibrating the above-presented model for three soil properties (clay content, sand content, and cation-exchange capacity). The calibration data were the values of these three soil properties at four fixed interval depths (0–15, 15–30, 30–60, and 60–100 cm) that were derived by means of an equal-area spline function (Bishop et al. 1999) from the measurements made at each horizons of the profiles. A set of geomorphometric indicators (elevation, slope, total curvature, profile curvature, multi-resolution valley bottom flatness index, multi-resolution ridge top flatness index, flow accumulation, and wetness index) were calculated from a 30 m ASTER digital elevation model for serving as input of the model (the $\{L\}$ set), in addition of the VNS-I estimates.

The validation results on the Cap Bon study area are presented in Table 13.5. When real measurements were used as inputs, the surface-subsurface models provided fairly satisfactory predictions at 15–30 cm and 30–60 cm but were less accurate at 60–100 cm (third column of Table 13.5). As expected, a moderate decrease of performances was observed when real measurements are replaced by Vis-NIR-SWIR imagery estimates of topsoil properties (fourth column of Table 13.5). Finally, it was observed that combining legacy data and VNS-I products allowed to capture between one third and two thirds of the total variance of the subsurface soil properties with a significantly decreasing performance with depth. This result, although modest, clearly outperformed prior attempts for mapping this property in this region.

This result suggests that legacy data and hyperspectral imagery products could be combined for improving precision of DSM outputs, even for subsurface soil properties.

Table 13.5 Performances of the surface-subsurface model in the Cap Bon region (After Lagacherie et al. 2013)

		Validation with different input (R^2)	
Soil property	Depth interval (cm)	Measured topsoil property	Vis-NIR-SWIR-estimated top soil property
Clay	15–30	0.81	0.51
	30–60	0.62	0.38
	60–100	0.38	0.33
Sand	15–30	0.83	0.67
	30–60	0.61	0.49
	60–100	0.56	0.46
CEC	15–30	0.71	0.69
	30–60	0.55	0.44
	60–100	0.53	0.35

13.4 Toward Further Improvements (These Are Still Research Questions)

Extending the Use of Vis-NIR-SWIR Imagery to Partly Vegetated Conditions
Spectrotransfer functions are build linking a set of soil surface properties (response Y variables) to a set of imaging reflectance spectra over bare soil pixels (predictor X variables) and, then, are applied to bare soil pixels to map the soil surface properties (Sect. 13.2.3). The applicability of these spectrotransfer functions is therefore inaccurate over surfaces which are partially covered by vegetation. Bartholomeus et al. (2011) estimated that for 5% of vegetation in a mixed pixel, soil organic carbon prediction is inaccurate. In addition, Ouerghemmi et al. (2011) estimated that from 10% of vegetation in a mixed pixel, clay content prediction is inaccurate. Few studies have addressed the estimation of soil properties over mixed pixels, including semi-vegetated pixels, by overcoming the problem of mixed spectra. Bartholomeus et al. (2011) and Ouerghemmi et al. (2011, 2016) proposed a "double-extraction" approach of soil property estimation from mixed spectra. This "double-extraction" approach consists of (1) a first extraction of an estimated soil spectrum, $\widehat{s}_{soil}$, from mixed Vis-NIR-SWIR spectra X and (2) a second extraction of soil property content from the estimated soil spectrum $\widehat{s}_{soil}$. Ouerghemmi et al. 2016 obtained acceptable predictions using the "double-extraction" approach for a significant part of their study area (63%) that was moderately covered by vegetation (NDVI <0.55). Finally, the performances of these studies were encouraging, and these approaches should be improved to be applied to pixels highly covered by vegetation and should be tested on higher diversity case studies.

Improving Soil Property Prediction from Soil Moisture

Using soil spectra measured with hyperspectral imaging spectrometers remains challenging due to uncontrolled variations in surface soil conditions (see Sect. 13.2.3.2), such as the soil moisture. The decrease of Vis-NIR-SWIR model performance, when dealing with field-moist soil samples, has been documented in the literature (Bricklemyer and Brown 2010; Minasny et al. 2009).

Several authors show the performance of the external parameter orthogonalization (EPO) algorithm (Roger et al. 2003), to correct for the effects of soil moisture in Vis-NIR-SWIR spectra and so improve the prediction of soil properties. Minasny et al. (2011) showed that EPO could remove the effect of water content from Vis-NIR-SWIR spectra allowing for successful predictions of soil organic carbon using PLSR models calibrated with air-dried spectra (Chap. 5, Sect. 5.4.2.1. of this book). The EPO-PLS method was also used for predicting clay content and soil organic carbon of intact and field-moist soil cores (Ge et al. 2014). And Ackerson et al. (2015) show that EPO-PLS is an effective tool for Vis-NIR-SWIR spectroscopy of tropical soils with mineralogy consisting of a mixture of kaolinite and iron-aluminum oxides.

In another way, Nocita et al. (2013) proposed an approach to improve SOC predictions from moist samples with unknown moisture content. The approach consist of (1) a calculation of the normalized soil moisture index (NSMI) (Haubrock et al. 2008) to estimate the soil moisture content for each moist soil sample which is used to spectrally classify the samples according to their moisture content and (2) a SOC content prediction using PLSR models developed on groups of samples with similar NSMI values.

Reducing Ground Measurements Using Legacy Data

Vis-NIR-SWIR hyperspectral imaging has been demonstrated to be a potential tool for topsoil property mapping (such as free iron, clay, and organic matter) over bare soils of large areas. Nevertheless, one of the limiting factors of hyperspectral data use for soil property mapping is the need for an "ad hoc" database for the study site to calibrate predictive models of the soil properties. These "ad hoc" databases are composed of soil spectra extracted from bare soil pixels of the Vis-NIR-SWIR imaging data and the corresponding soil property values measured over soil samples collected over the bare soil pixels for which soil spectra are extracted.

An alternative calibration approach should be to replace the expensive "ad hoc" field sampling datasets by legacy soil datasets which are much less expensive to collect. In this way, Gomez et al. (2016) proposed to use a spectral index to predict clay contents, which were standardized with soil legacy data for soil properties mapping. The main performance limitations were due to the bias of clay measurements, which affected the legacy soil databases. As explained by Ciampalini et al. (2013), the bias of the legacy databases results from two main causes: (1) differences in laboratory soil analysis protocols which cause converges toward a general problem of digital soil mapping and should receive more attention in future studies (Ciampalini et al. 2013, Baume et al. 2011) and (2) sampling locations, as the sites of legacy data are not always located inside the study area. Finally, the

combination use of legacy and "ad hoc" soil sampling for the calibration of Vis-NIR-SWIR hyperspectral imaging models would open up new ways to deal with acceptable costs of sampling and acceptable prediction performances.

Using DSM Models Adapted to Peculiarities of Vis-NIR-SWIR Products
Up to now, the use of Vis-NIR-SWIR products in DSM has involved classical DSM models that were tailored for using as soil input a set of sites with exact values of soil property, usually sparsely sampled in the study area. Advances are expected by taking into account some peculiarities of this new soil input. First, the high densities of observed sites allow to envisage models that would better account for the complexity of the soil-landscape relationships that are often observed in the field. For example, anisotropic spatial interpolation driven by topography (Schwanghart and Jarmer 2011) can be suitable in pedological contexts characterized by toposequences. Another example is to account for non-stationarities that may affect the soil-landscape relationships across a study area by applying, e.g., local regression kriging (Sun et al. 2012). Another peculiarity of Vis-NIR-SWIR hyperspectral imagery products is that they provide estimated values of soil properties that convey non-negligible uncertainties. There exists, in the geostatistical toolbox, models that are able to handle such uncertain inputs, e.g., soft kriging approach (Journel 1986). However, they have not been involved yet in DSM applications. A prerequisite for using such models would be to progress in estimating the local uncertainty associated with Vis-NIR-SWIR hyperspectral imagery products (Gomez et al. 2015).

13.5 Conclusions

- Soils have specific spectral signatures related with chemical and physical components, themselves related with some soil properties.
- These soil properties can be estimated from reflectance spectra by spectrotransfer functions calibrated from a limited set of soil property measurements (soil spectroscopy combined with soil properties).
- Satisfactory results are obtained for soil properties (i) related with chemical species impacting soil reflectance or (ii) correlated with the latter and (iii) variable enough over the study area.
- Vis-NIR-SWIR remote sensing is limited to topsoil characterization and can only be applied on dry and bare soil surfaces.
- By increasing dramatically the number of measured sites, Vis-NIR-SWIR hyperspectral imaging can be a valuable input for digital soil mapping in the near future.

References

Ackerson JP, Demattê JAM, Morgan CLS (2015) Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. Geoderma 259–260:196–204

Bartholomeus H, Epema G, Schaepman M (2007) Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. Int J Appl Earth Obs Geoinf 9:194–203

Bartholomeus H, Schaepman ME, Kooistra L, Stevens A, Hoogmoed WB, Spaargaren OSP (2008) Spectral reflectance based indices for soil organic carbon quantification. Geoderma 145(1–2):28–36

Bartholomeus H, Kooistra L, Stevens A, Van Leeuwen M, Van Wesemael B, Ben-Dor E, Tychon B (2011) Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. Int J Appl Earth Obs Geoinf 13:81–88

Baume O, Skøien JO, Heuvelink GBM, Pebesma EJ (2011) A geostatistical approach to data harmonization – application to radioactivity exposure data. Int J Appl Earth Obs Geoinf 13(3):409–419

Baumgardner MF, Kristof SJ, Johannsen CJ, Zachary AL (1970) Effects of organic matter on the multispectral properties of soils. Proc Indian Acad Sci 79:413–422

Baumgardner MF, Silva LF, Biehl LL, Stoner ER (1985) Reflectance properties of soils. Adv Agron 38:l–44

Ben-Dor E, Banin A (1995) Near infrared analysis (NIRA) as a simultaneous method to evaluate spectral featureless constituents in soils. Soil Sci 159:259–269

Ben-Dor E, Inbar Y, Chen Y (1997) The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500) during a controlled decomposition process. Remote Sens Environ 61:1–15

Ben-Dor E, Irons JA, Epema A (1998) Soil spectroscopy. In: Rencz A (ed) Manual of remote sensing, 3rd edn. Wiley, New-York, pp 111–189

Ben-Dor E, Patkin K, Banin A, Karnieli A (2002) Mapping of several soil properties using DAIS-7915 hyperspectral scanner data – a case study over clayey soils in Israel. Int J Remote Sens 23:1043–1062

Ben-Dor E, Levin N, Singer A, Karnieli A, Braun O, Kidron GJ (2006) Quantitative mapping of the soil rubification process on sand dunes using an airborne hyperspectral sensor. Geoderma 131:1–21

Ben-Dor E, Taylor RG, Hill J, Demattê JAM, Whiting ML, Chabrillat S, Sommer S (2008) Imaging spectrometry for soil applications. Adv Agron Acad Press 97:321–392

Ben-Dor E, Chabrillat S, Demattê JAM, Taylor GR, Hill J, Whiting ML, Sommer S (2009) Using imaging spectroscopy to study soil properties. Remote Sens Environ 113(1):S38–S55. 2009

Bishop TFA, McBratney AB, Laslett GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45

Boettinger J, Ramsey RD, Bodily JM, (2008) Landsat spectral data for digital soil mapping. In: Hartemink A, McBratney AB Mendonca Santos L (eds), Digital soil mapping with limited data. pp. 193–199

Bowers SA, Hanks AJ (1965) Reflection of radiant energy from soil. Soil Sci 100:130–138

Bricklemyer RS, Brown DJ (2010) On-the-go VisNIR: potential and limitations for mapping soil clay and organic carbon. Comput Electron Agric 70(1):209–216

Briottet X, Marion R, Carrere V, Jacquemoud S, Bourguignon A, Chami M, Chanussot J, Chevrel S, Deliot P, Dumont M, Foucher PY, Minghelli-Roman A, Sheeren D, Weber C, Prastault P, Hosford S, Lefevre MJ (2013) HYPXIM: HYPXIM: a second generation high spatial resolution hyperspectral satellite for dual applications, Fifth Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)., June 2013. Gainesville, Florida

Chabrillat S, Goetz AFH, Krosley S, Olsen HW (2002) Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. Remote Sens Environ 82:431–445

Ciampalini R, Lagacherie P, Monestiez P, Walker E, Gomez C (2012) Co-kriging of soil properties with Vis-NIR hyperspectral covariates in the Cap Bon region (Tunisia). In: Minasny, Malone, McBratney (eds), Digital soil assessments and beyond. CRC Press 2012, pp. 393–398

Ciampalini R, Lagacherie P, Gomez C, Grunberger O, Hamrouni H, Mekki I, Richard A (2013) Detecting and correcting biases of measured soil profiles data. A case study in the Cap Bon Region (Tunisia). Geoderma 192:68–76

Escadafal R, Huete A (1991) Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices végétation obtenus par télédétection. C R Acad Sci Paris Série II 312:1385–1391

Gaffey SJ (1987) Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 pm): calcite, aragonite and dolomite. J Geophys Res Solid Earth 71:151–162

Ge Y, Morgan CLS, Ackerson JP (2014) VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. Geoderma 213:61–69

Gomez C, Viscarra Rossel RA, McBratney AB (2008a) Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: an Australian case study. Geoderma 146(3–4):403–411

Gomez C, Lagacherie P, Coulouma G (2008b) Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma 148(2):141–148

Gomez C, Lagacherie P Bacha S (2012a) Using an Vis-NIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia). In: Minasny, Malone & McBratney (eds), Digital soil assessments and beyond. CRC Press 2012, pp. 387–392

Gomez C, Coulouma G, Lagacherie P (2012b) Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. Geoderma 189–190:176–185

Gomez C, Drost APA, Roger J (2015) Remote sensing of environment analysis of the uncertainties affecting predictions of clay contents from VNIR/SWIR hyperspectral data. Remote Sens Environ 156:58–70

Gomez C, Gholizadeh A, Borůvka L, Lagacherie P (2016) Using legacy data for predicting soil surface clay content from VNIR/SWIR hyperspectral airborne images. Geoderma 276:84–92

Grunwald S (2009) Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152:195–207

Guanter L, Kaufmann H, Segl K, Förster S, Rogass C, Chabrillat S, Küster T, Hollstein A, Rossner G, Chlebek C, Straif C, Fischer S, Schrader S, Storch T, Heiden U, Müller A, Bachmann M, Mühle H, Müller R, Habermeyer M, Ohndorf A, Hill J, Buddenbaum H, Hostert P, van der Linden S, Leitao PJ, Rabe A, Doerffer R, Krasemann H, Xi H, Mauser W, Hank T, Locherer M, Rast M, Staenz K, Sang B (2015) The EnMAP Spaceborne imaging spectroscopy mission for earth observation. Remote Sens 7(7):8830–8857

Haubrock S, Chabrillat S, Lemmnitz C, Kaufmann H (2008) Surface soil moisture quantification models from reflectance data under field conditions. Int J Remote Sens 29:3–29

Hunt GR, Salisbury JW, Lenhoff CJ (1971) Visible and near-infrared spectra of minerals and rocks: III. Oxides hydroxides. Mod Geol 2:195–205

Journel AG (1986) Constrained interpolation and qualitative information – the soft kriging approach. Math Geol 18:269

Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11(1):137–148

Kokaly RF, Clark RN (1999) Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sens Environ 67:267–287

Lagacherie P, Baret F, Feret JB, Madeira NJ, Robbez-Masson JM (2008) Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sens Environ 112(3):825–835

Lagacherie P, Bailly JS, Monestiez P, Gomez C (2012) Using scattered soil sensing field surveys to map soil properties over a region. An example with airborne hyperspectral imagery. Eur J Soil Sci 63:110–119

Lagacherie P, Sneep A-R, Gomez C, Bacha S, Coulouma G, Hamrouni MH, Mekki I (2013) Combining Vis–NIR hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia). Geoderma 209–210:168–176

Levin N, Tsoar H, Maia LP, Sales VC, Herrmann H (2007) Dune whitening and inter-dune freshwater ponds in NE Brazil. Catena 70(1):1–15

Madeira J, Bedidi A, Pouget J, Cervelle B, Flay N (1997) Spectrometric indices (visible) of hematite and goethite contents in lateritic soils. Application to a TM image for soil mapping of Brasilia area. Int J Remote Sens 18(13):2835–2852

McBratney AB, Mendonca Santos ML, Minasny B (2003) On digital soil mapping. Geoderma 117:3–52

Minasny B, McBratney AB, Pichon L, Sun W, Short MG (2009) Evaluating near infrared spectroscopy for field prediction of soil properties. Soil Res 47(7):664–673

Minasny B, McBratney AB, Bellon-Maurel V, Roger JM, Gobrecht A, Ferrand L, Joalland S (2011) Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167–168:118–124

Morris RV, Lauer HV, Lawson CA, Gibson EK Jr, Nace GA, Stewart C (1985) Spectral and other physicochemical properties of submicron powders of Hematite (α-Fe 2 O 3), Maghematite (γ-Fe 2 O 3), Magnetite (Fe 3 O 4), Goethite (α-FeOOH), and Lepidocrocite (γ-FeOOH). J Geophys Res 90:3126–3144

Nanni MR, Demattê JAM (2006) Spectral reflectance methodology in comparison to traditional soil analysis. Soil Sci Soc Am J 70(2):393–407

Nocita M, Stevens A, Noon C, Van Wesemael B (2013) Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. Geoderma 199:37–42

Ouerghemmi W, Gomez C, Naceur S, Lagacherie P (2011) Applying blind source separation on hyperspectral data for clay content estimation over partially vegetated surfaces. Geoderma 163(3–4):227–237

Ouerghemmi W, Gomez C, Nacer S, Lagacherie P (2016) Semi-blind source separation for estimation of clay content over semi-vegetated areas, from VNIR/SWIR hyperspectral airborne data. Remote Sens Environ 181:251–263

Pouget M, Madeira J, Le Floc'h E, Kamal S (1991) Caractéristiques spectrales des surfaces sableuses de la région cotière nord-ouest de l'Egypte: application aux données satellitaires SPOT. In Proc. 2e'me Journées Télédétection. In: Caractérisation et Suivi des Milieux Terrestres en Régions Arides et Tropicales. ORSTOM, Bondy, pp. 27–38

Roger JM, Chauchard F, Bellon-Maurel V (2003) EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemom Intell Lab Syst 66(2):191–204

Schwanghart W, Jarmer T (2011) Linking spatial patterns of soil organic carbon to topography – a case study from south-eastern Spain. Geomorphology 126:252–263

Selige T, Bohner J, Schmidhalter U (2006) High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136(1–2):235–244

Snee RD (1977) Validation of regression models: methods and examples. Technometrics 19(4):415–428

Stevens A, Wesemael B, Bartholomeus B, Rosillon D, Tychon B, Ben-Dor E (2008) Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma 144(1–2):395–404

Stoner ER, Baumgardner MF (1981) Characteristic variations in reflectance of surface soils. Soil Sci Soc Am J 45:1161–1165

Sun W, Minasny B, McBratney AB (2012) Analysis and prediction of soil properties using local regression-kriging. Geoderma 171–172:16–23

Van Der Meer F (2004) Analysis of spectral absorption features in hyperspectral imagery. Int J Appl Earth Obs Geoinf 5(1):55–68

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59–75

Wackernagel H (1995) Multivariate geostatistics. Springer Verlag edition

Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130

Chapter 14
Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling

Gerard B. M. Heuvelink

"We demand rigidly defined areas of doubt and uncertainty"!

Douglas Adams
The Hitchhiker's Guide to the Galaxy

In previous chapters, the use of geostatistical modelling for soil mapping was addressed. We learnt that one of the advantages of kriging is that it not only produces a map of predictions but that it also quantifies the uncertainty about the predictions, through the kriging standard deviation. In this chapter we will look into this in more detail. We will also examine another way to assess the accuracy of soil prediction maps, namely, through independent validation. This approach has the advantage that it is model-free and hence makes no assumptions about the structure of the spatial variation and relationships between the target soil property and covariates. Finally, we will examine how uncertainties in soil maps propagate through environmental models and spatial analyses. Throughout this chapter we will use the Allier data set and case study, Limagne rift valley, central France, to illustrate concepts and methods. We will only consider soil properties that are measured on a continuous-numerical scale. Many of the concepts presented can also be extended to categorical soil variables, but this is more complicated and beyond the scope of this chapter.

14.1 What Is Uncertainty?

Suppose that the bulk density of the topsoil at some location in some study area equals 1.33 g/cm^3. Suppose further that this value is unknown to us because we did not measure the bulk density at the location. All that we have is a soil property map

G.B.M. Heuvelink (✉)
Soil Geography and Landscape group, Wageningen University and ISRIC – World Soil Information, PO Box 47, 6700 AA Wageningen, The Netherlands
e-mail: gerard.heuvelink@wur.nl

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_14

that contains predictions of the bulk density for all locations in the study area, also at the location where the true bulk density is 1.33 g/cm^3. Let the bulk density at that location according to the map be 1.45 g/cm^3. This shows that the soil map is in error. The error equals 1.33–1.45 = −0.12 g/cm^3. Here, *error* is defined as the difference between the true and predicted value of the soil property.

In practice, we usually do not know the error, because we would need the true value to calculate it and we do not have the resources to perfectly measure the soil everywhere. In other words, we are *uncertain* about the error (and the true value). Here, *uncertainty* refers to a state of mind of a person or people that expresses a lack of confidence about reality (Heuvelink 2014). Note that uncertainty is a property of people. It is not the soil bulk density that is uncertain; it is we that are uncertain about the soil bulk density. We are uncertain because we have limited and imperfect information (i.e. only a soil map) and are aware that the information we have may be in error.

Although we are uncertain, this does not mean that we are completely ignorant. For instance, we may know that the chances are equal that the error in the bulk density prediction is positive or negative (because we used an unbiased mapping method), we may know that it is unlikely that the absolute value of the error is greater than 0.50 g/cm^3, etc. Thus, it is not unreasonable to assume that we can come up with a large number of possible error values and attach a probability to each of these. Since the true value is the sum of the (known) prediction and the error, we can also list all possible values of the soil property and attach a probability to each. If we can do this, then we have characterised (our uncertainty about) the soil property by a probability distribution.

Now that we have characterised the soil property by a probability distribution, it has effectively become a random variable. After all, a random variable is nothing else than a variable that can take on many values, where each value has a certain probability of occurrence. Since we deal with spatially distributed variables, we must extend this concept to that of a random field. A random field is a collection of indexed random variables, where in our case the index is geographic location. We can characterise the variable at each location by a (univariate, marginal) probability distribution, but we must also characterise the (spatial) correlation between the variables at multiple locations. Geostatistics provides the methods and tools to do this (i.e. variogram estimation and kriging), and this has been explained in detail in Chaps. 9 and 10 (but see also Goovaerts 2001). However, while in previous chapters the focus was on the predictions made by kriging, in this chapter we will concentrate on the uncertainty associated with these predictions. In the next section, we will explain how geostatistics can be used to model uncertainty in mapped soil properties by means of a cokriging example.

14.2 Geostatistical Modelling of Uncertainty

14.2.1 Mapping Soil Properties for the Allier Study Area

As part of a research study in quantitative land evaluation, a crop simulation model was used to calculate potential crop yields for floodplain soils of the Allier River in the Limagne rift valley, central France. The moisture content at wilting point (Θ_{wp}, cm^3/cm^3) is an important input attribute for the crop simulation model. Because Θ_{wp} varies considerably over the area in a way that is not linked directly with soil type, it was necessary to map its variation separately to see how moisture limitations affect the calculated crop yield.

Unfortunately, because Θ_{wp} must be measured on samples in the laboratory, it is expensive and time-consuming to determine it for a sufficiently large number of data points for creating the prediction map by kriging. An alternative and cheaper way is to calculate Θ_{wp} from other soil properties which are cheaper to measure or using pedotransfer functions (see Chap. 7). Because moisture content at wilting point is often strongly correlated with moisture content at field capacity (Θ_{fc}, cm^3/cm^3) and soil porosity (Φ, cm^3/cm^3), both of which can be measured more easily and cheaply, it was decided to map these first and next derive a map of Θ_{wp} from these using multiple linear regression. We will come back to this in Sect. 14.4 and concentrate first on the kriging of Θ_{fc} and Φ.

Sixty-two measurements of Θ_{fc} and Φ were made in the field at the sites indicated in Fig. 14.1. From these data experimental variograms and an experimental cross-

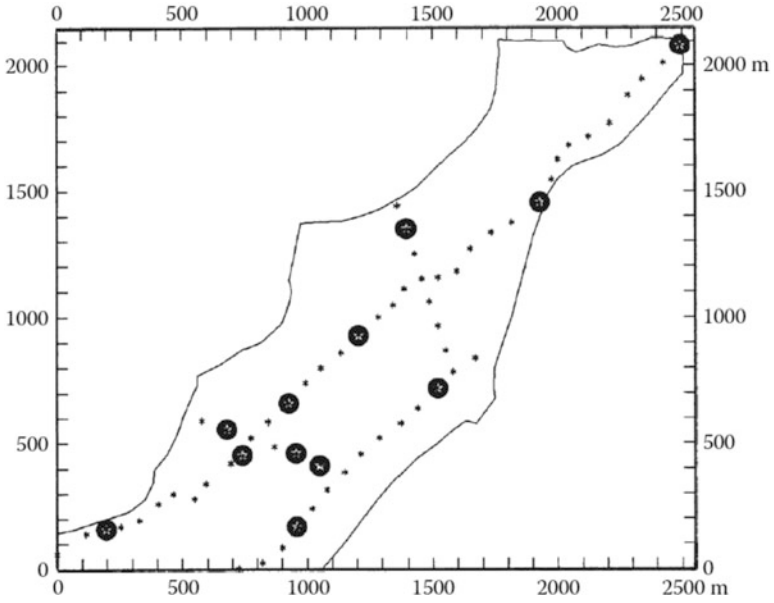


Fig. 14.1 The Allier study area showing sampling points of moisture content at field capacity and porosity. Circled sites are those where in addition moisture content at wilting point was measured

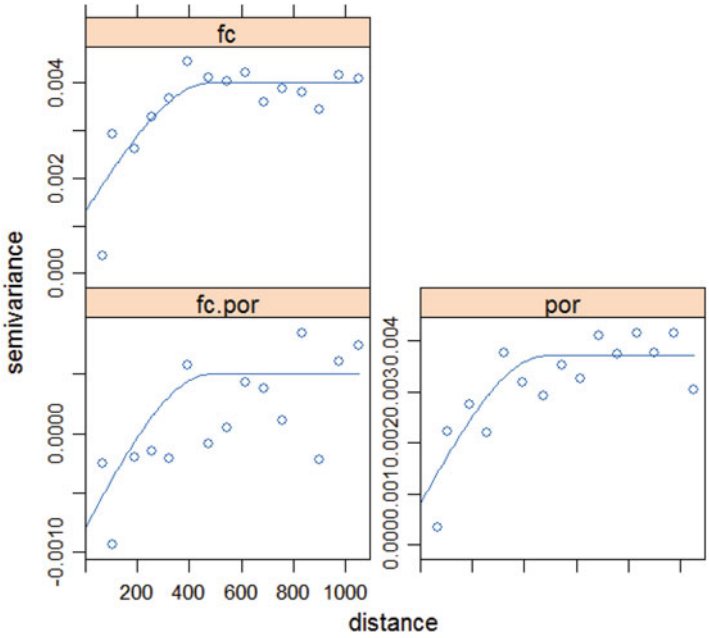


Fig. 14.2 Experimental variograms (*open circles*) and fitted variogram models (*solid lines*) for Θ_{fc} (*top left*), Φ (*bottom right*) and the experimental and modelled cross-variogram (*bottom left*)

variogram were computed (see Sect. 10.5). These were then fitted using the linear model of coregionalisation (Fig. 14.2). Next both soil properties were mapped to a regular 50×50 m grid using ordinary cokriging. The cokriging yielded raster maps of means and standard deviations for both Θ_{fc} and Φ, as well as a map of the correlation of the cokriging prediction errors. Figure 14.3 displays these maps.

14.2.2 Interpreting the Kriging Standard Deviation Maps

The kriging standard deviation maps shown in Fig. 14.3 are summary measures of the uncertainty about Θ_{fc} and Φ in the study area. These uncertainties are the result of interpolation errors: while we know the true values of Θ_{fc} and Φ at the 62 observation locations (assuming that measurement errors are negligibly small), we are uncertain about their true value at non-observation locations. As explained in Sect. 14.1, we are uncertain because the true value is unknown to us, and so we cannot identify a single true reality. At best we can list all possible values of the soil property and attach a probability to each of them. This is exactly what we do in kriging, because under the assumptions made (i.e. normality, stationarity, isotropy), we derived a (conditional) probability distribution of Θ_{fc} and Φ for each grid cell. In this case, the uncertainty at each grid cell is characterised by a zero-mean

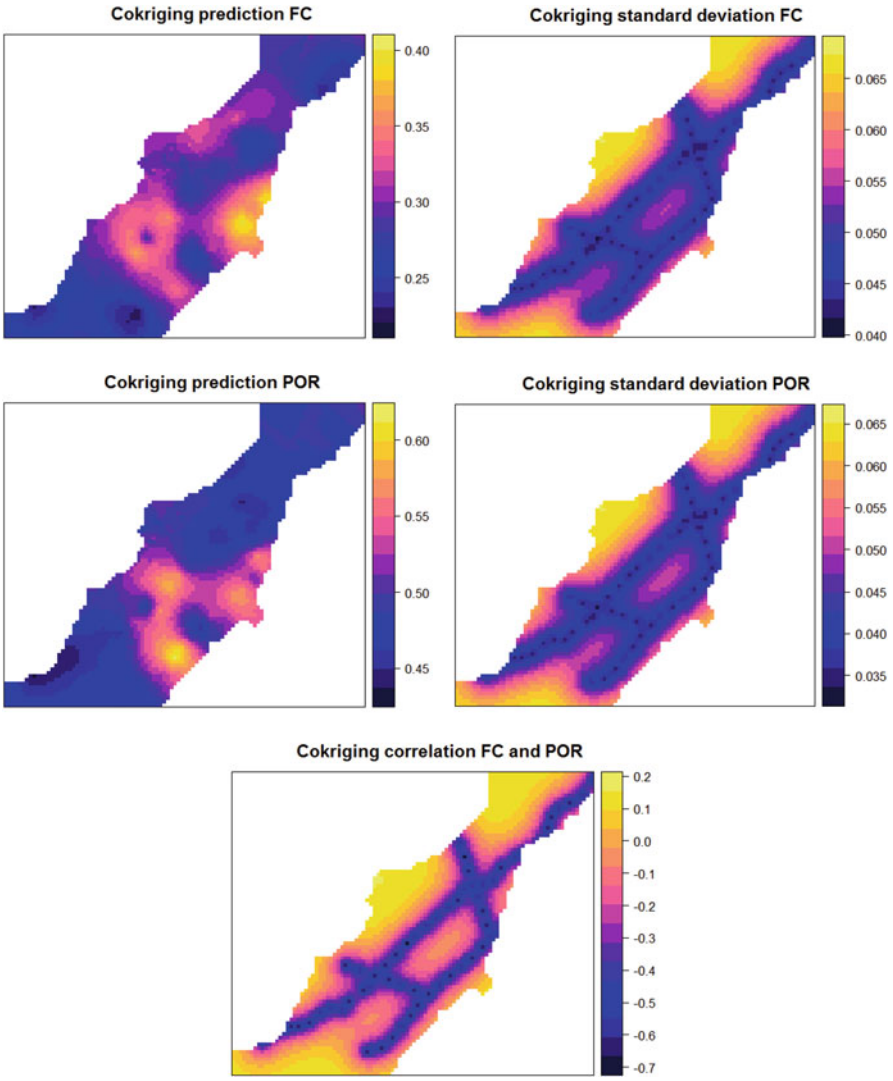


Fig. 14.3 Cokriging results for the Allier study area (50 × 50 m grid): conditional mean and standard deviation of Θ_{fc} (cm^3/cm^3), conditional mean and standard deviation of Φ (cm^3/cm^3) and correlation of cokriging prediction errors of Θ_{fc} and Φ

normal distribution with a standard deviation as given in Fig. 14.3. The magnitude of uncertainty is captured by the width of the probability distribution, although the shape of the distribution is important as well (see Fig. 14.4). Because of the assumption of normality, the uncertainties of the kriged Θ_{fc} and Φ all have a shape such as shown in the left-hand panels of Fig. 14.4. The width of the distribution varies in space, as is clear from Fig. 14.3.

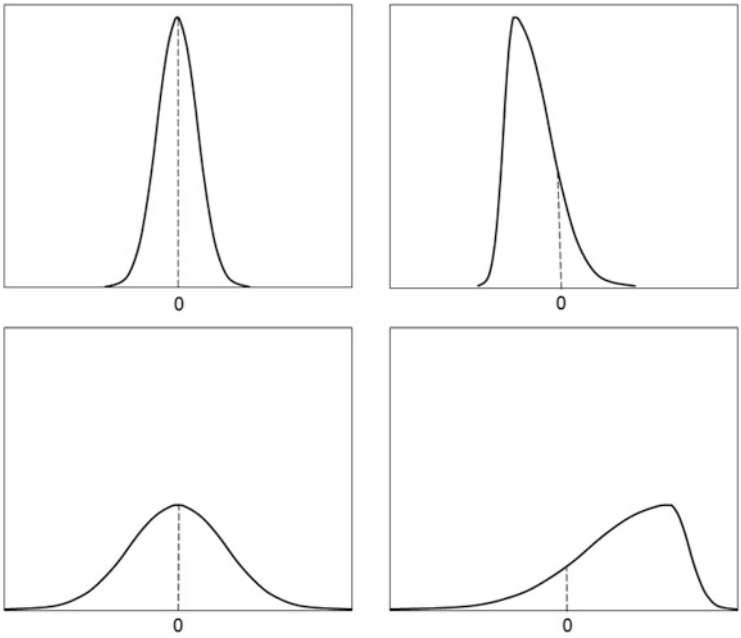


Fig. 14.4 Examples of probability distributions to characterise uncertainty. Probability distributions can be narrow (small uncertainty, *top*) or wide (large uncertainty, *bottom*). They can also be symmetric around zero (*left*) or asymmetric and biased (*right*)

The probability distributions shown in Fig. 14.4 refer to the value of a single variable. In our case study, we considered two variables, Θ_{fc} and Φ, and so each of these will have their own marginal probability distribution at any one location in the study area. But the uncertainties associated with these two variables are also correlated, because of the cross-correlation between Θ_{fc} and Φ, as characterised by the cross-variogram. It is difficult to predict how large the correlation between the prediction errors at any given location is, because the correlation between the cokriging prediction errors is not the same as that between the variables themselves. In other words, in general we have $\text{corr}\left(\widehat{\Theta}_{\text{fc}} - \Theta_{\text{fc}}, \widehat{\Phi} - \Phi\right) \neq \text{corr}\left(\Theta_{\text{fc}}, \Phi\right)$. Fortunately, cokriging provides the correlations between the cokriging errors at each location, as shown in the bottom map in Fig. 14.3. Note that there are clear spatial variations in the correlation between the cokriging errors and that these may be positive as well as negative, depending on location. A graphical illustration of the joint (bivariate) probability distribution of two uncertain variables is shown in Fig. 14.5. If the correlation between the uncertain variables were zero, then the major and minor axes of the ellipses would be along the axes of variables S_1 and S_2, i.e. they would not be rotated. The example in the right panel of Fig. 14.5 shows a case in which there is a non-zero correlation. In this example the correlation is positive because the major axis has a positive angle. The contour lines would be circular if

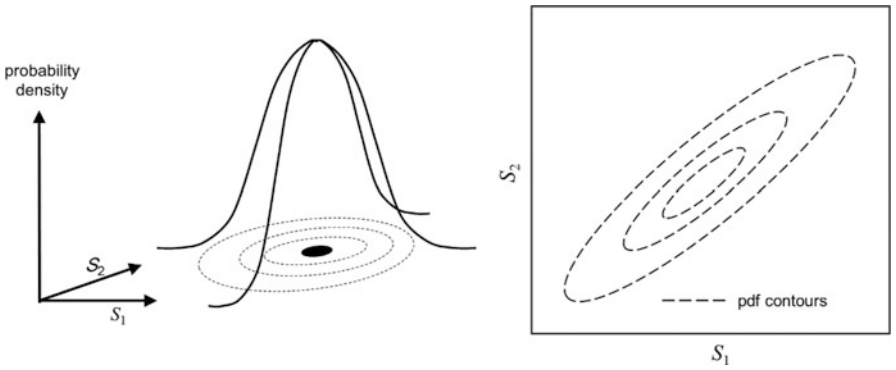


Fig. 14.5 Examples of a bivariate probability distribution of two uncertain variables S_1 and S_2. The rotated ellipses of equal probability density shown in the *right panel* indicate positive correlation

the two variables would be uncorrelated and have equal standard deviation. Figure 14.5 refers to two uncertain variables S_1 and S_2, which could be $S_1 = \widehat{\Theta}_{\mathrm{fc}} - \Theta_{\mathrm{fc}}$ and $S_2 = \widehat{\Phi} - \Phi$, but note that it might as well refer to $S_1 = \widehat{\Theta}_{\mathrm{fc}}(x) - \Theta_{\mathrm{fc}}(x)$ and $S_2 = \widehat{\Theta}_{\mathrm{fc}}(x') - \Theta_{\mathrm{fc}}(x')$ for two arbitrary locations x and x' in the study area.

The kriging standard deviations shown in Fig. 14.3 vary spatially and tend to be small in the neighbourhood of observation locations and are large further away from these, particularly at the boundary of the study area. This is as one would expect intuitively, because the magnitude of the interpolation error depends on the closeness of observations and their local density and because (spatial) extrapolation is more error prone than interpolation. It can also be inferred from the kriging variance equation (see Sect. 10.3):

$$\sigma_K{}^2(x_0) = E\left[\left(\widehat{S}(x_0) - S(x_0)\right)^2\right] = \sum_{i=1}^{n} \lambda_i \cdot \gamma(|x_i - x_0|) + \varphi \qquad (14.1)$$

where n is the number of observations used in kriging, the λ_i are kriging weights, the x_i are observation locations and x_0 is the prediction location, γ is the variogram model and φ is a Lagrange parameter. In most practical cases, the latter is relatively small so that we can concentrate on the summation part. Inspection shows that this part will be small when the distances between the x_i and x_0 are small, hence when observation locations are close to the prediction location. Of course the exact result also depends on the shape of the variogram model. For instance, in case of a pure nugget variogram, the kriging variance (and hence its square root, the kriging standard deviation) will be constant: if there is no spatial correlation, then interpolation cannot benefit from nearby observations, and the interpolation error (variance) will be equal everywhere. Note that the kriging variance can never be smaller than the nugget variance, except when we interpolate to an observation location. For Φ, which has a nugget variance of 0.0008 (see Fig. 14.2), this means

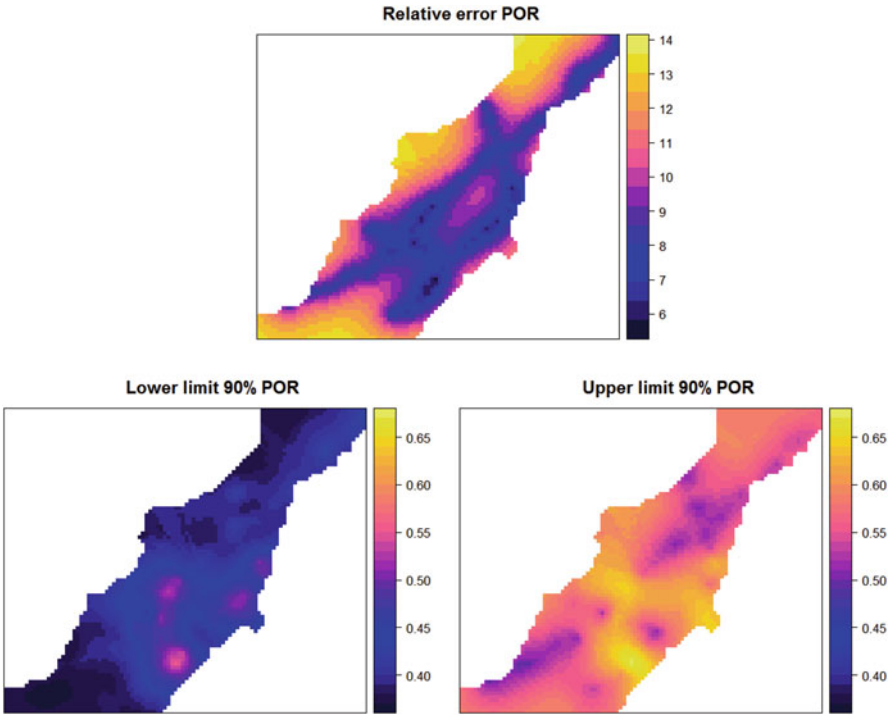


Fig. 14.6 Relative error (*top*) and lower and upper limits of the central 90% prediction interval of the topsoil porosity for the Allier study area

that the uncertainty about the predicted Φ at any non-observation location in the Allier study area will be at least 0.028 cm^3/cm^3 (as confirmed by Fig. 14.3). Note also that in the Allier example, the kriging variance will be calculated in a slightly different way than using Eq. 14.1, because in that case predictions were made using cokriging instead of kriging (Wackernagel 2003).

For users with little background in (geo)statistics, it may not be that easy to interpret a standard deviation map. More appealing for uncertainty communication are maps of the relative error (computed as the ratio of the kriging standard deviation and kriging prediction maps, multiplied with 100%) and maps of the lower and upper limits of a central 90% prediction interval, which are derived by subtracting and adding 1.64 times the kriging standard deviation map from the kriging prediction map, respectively. Note that here we assumed that the kriging error is normally distributed. These maps are shown for the soil porosity in the Allier study area in Fig. 14.6. The relative error is nowhere greater than 15%, indicating that the uncertainty about soil porosity is small compared to its predicted value. Nonetheless, the differences between the lower and upper limits of the 90% prediction interval maps are large, indicating that the kriging interpolation error is far from negligible.

14.2.3 Spatial Stochastic Simulation

Kriging makes predictions, such that the expected squared prediction error is minimised. This is attractive because it means that the predicted value is on average closest to the true (unknown) value. As explained in Chap. 9, s*patial stochastic simulation* has an entirely different objective. Here, the goal is to generate 'possible realities' from the probability distribution of the uncertain variable. This is done by sampling from the probability distribution using a *pseudo-random number generator*. The result of a spatial stochastic simulation exercise is not unique, because there are an infinite number of possible realities, from which just one or several are taken. To illustrate the difference between optimal prediction and stochastic simulation, take the example of the outcome of a throw of a fair die. Optimal prediction would produce the value of 3.5, because on average this is the number closest to any of the outcomes 1–6. However, stochastic simulation would randomly take one of the numbers 1–6, where each of the six values would have equal chance of being selected. Figure 14.7 shows four realisations ('possible realities') from the kriging probability distribution of topsoil porosity in the Allier study area. These were created using *conditional* simulation, meaning that the

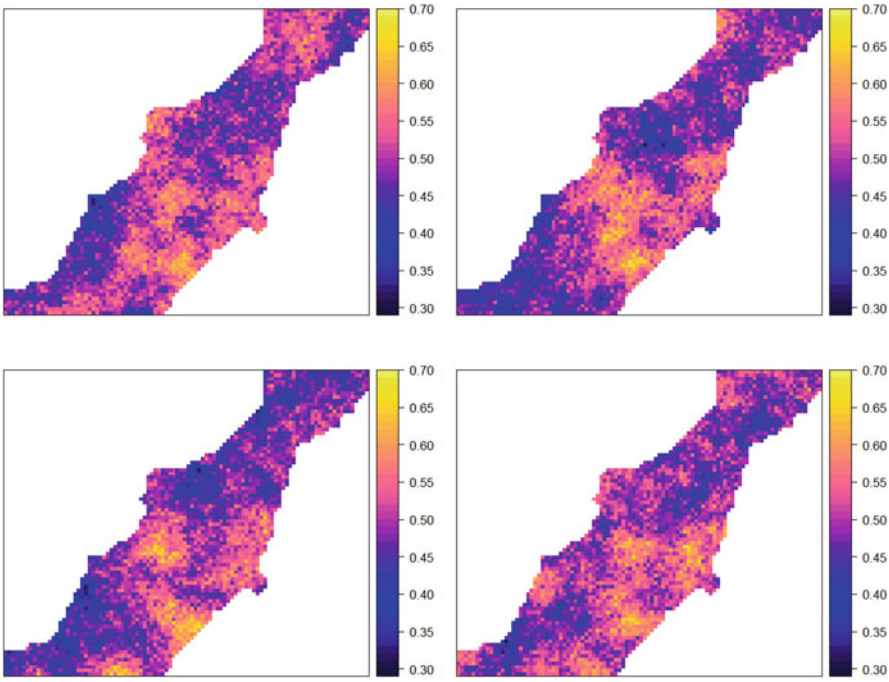


Fig. 14.7 Four possible realities of the topsoil porosity for the Allier study area generated with conditional spatial stochastic simulation

62 observations of soil porosity were used as conditioning data. The differences between the simulated maps convey the uncertainty about the true porosity, and when shown in animation mode, they are an attractive means to communicate the uncertainty of an interpolated map to non-experts. We will also make use of simulated maps in Sect. 14.4, when we explain the Monte Carlo method for analysing uncertainty propagation through environmental models.

If we would generate many more than four realisations of topsoil porosity, then their average would equal the kriging prediction map shown in Fig. 14.3, while their standard deviation would equal the kriging standard deviation map, also shown in Fig. 14.3. Thus, although perhaps not easily noticeable from Fig. 14.7, the differences between the realisations are greater far from observation locations than close to observation locations.

14.2.4 Change of Support

Often users do not want to predict soil properties at points but instead are interested in the average value over a larger piece of land. For instance, perhaps for a farmer it is not that relevant to know Θ_{fc} and Φ at point locations within the Allier study area, but instead the real interest is in parcel averages. Such averages over spatial units or 'blocks' can be predicted using *block kriging*, as explained in Sect. 10.3. The blocks need not be rectangular or square but may take irregular shapes as well. They may even be as large as the entire study area. When the blocks are relatively small, then block kriging produces similar predictions as point kriging, but the associated kriging standard deviation is usually much smaller, especially when the nugget variance is large. Figure 14.8 shows the cokriging standard deviation maps of Θ_{fc} and Φ for the case where the blocks are equal to the grid cells. Note that the standard deviations are substantially smaller compared to those of point kriging shown in Fig. 14.3. The explanation is that within-block spatial variation averages

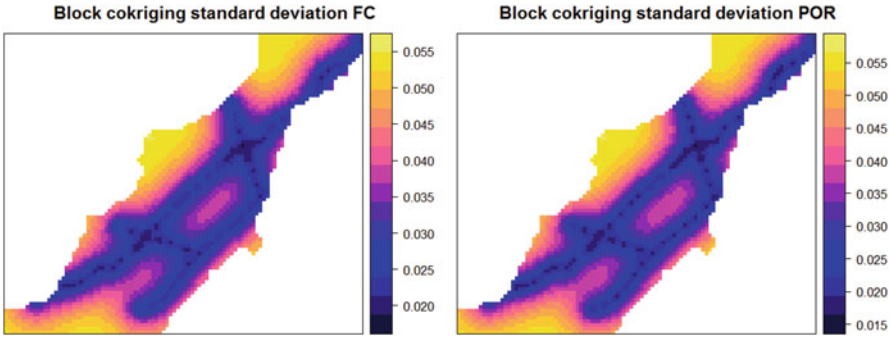


Fig. 14.8 Block cokriging standard deviation maps of Θ_{fc} and Φ using blocks of 50×50 m

out when the block mean is predicted. In other words, in case of large short-distance spatial variation (i.e. high nugget), predictions of block averages are much more accurate than point predictions. This tells us that it is crucially important to choose the right support when addressing uncertainty in interpolated soil maps.

Block kriging is used when *spatial aggregation* is the objective, in other words it is used in a case where the observations have a smaller support than the predictions. The opposite, i.e. making predictions at a smaller support than the observations, is known as *area-to-point kriging*. It will be no surprise that in this case uncertainty increases instead of decreases. There are not many examples of area-to-point kriging in soil science, because usually the starting point is observations at point support, but an exception is vertical spatial interpolation. In this case observations are often averages over soil horizons or layers, while predictions may be required for smaller or different depth intervals (Orton et al. 2016).

14.2.5 Extension to Kriging with External Drift

So far we discussed the uncertainty resulting from ordinary (co)kriging. But in recent years ordinary kriging is used less frequently and often replaced with kriging with external drift (KED), also termed universal kriging and regression kriging (Odeh et al. 1995; Hengl et al. 2004). Chapter 9 provides the details. This is because we rarely have only soil point observations as a source of information, but in addition we may have a large suit of covariate maps that provide valuable information about the soil property of interest. The additional information may then be used to improve the mapping and reduce uncertainty. The mathematical expression for the KED variance, which quantifies the uncertainty in the resulting map, is more complicated than that of the ordinary kriging variance given in Eq. 14.1. It is the sum of the trend estimation variance and the kriging variance. Even though the trend estimation variance is added, in practice the KED variance will usually be smaller than the ordinary kriging variance. This is because the KED variance is based on the variogram of the residual (defined as the difference between the soil property and the trend), which typically is much smaller than the variogram of the soil property itself, and hence the kriging component of the KED variance will decrease. See Sect. 10.3 for a more detailed discussion of KED and comparison with ordinary kriging.

Recall from Sect. 14.2.2 that the kriging standard deviation tends to be small near observation locations and large further away from them. As noted, this can be explained from a closer look at Eq. 14.1, which shows that the kriging variance will be larger if the distance between observation and prediction locations is large. Note also from Eq. 14.1 that the kriging variance does not depend on the observations themselves but only on the variogram and configuration of the observation locations. This allows optimisation of spatial sampling designs that minimise the spatially averaged kriging variance, as explained in Sect. 11.6. Sampling design optimisation under the ordinary kriging model typically leads to a fairly uniform distribution of

the sampling locations, with a slightly higher concentration of sampling locations near the study area boundary. In case of KED another aspect is also included in the sampling design optimisation. In that case the trend estimation error variance needs also to be minimised, which calls for a joint optimisation in geographic and 'feature' space. In other words, we must make sure that the observations also cover the covariate space well (Minasny and McBratney 2006; Brus and Heuvelink 2007).

14.2.6 Uncertainty Quantification of a Given Soil Map

The geostatistical uncertainty quantification approach works well in situations where one starts from scratch and where it is feasible to build a geostatistical model of reality. However, what to do in situations in which a soil property map has already been derived, without using geostatistical models? These could be maps made using deterministic algorithms, such as inverse distance or nearest neighbour interpolation. Alternatively, soil property maps may have been derived from an existing soil class map, by assuming that soil properties within each soil class are constant and assign these map-unit mean values using expert judgement or data from 'representative' profiles. It may also be that a soil property map is provided without additional information about how the map was made and without quantification of the uncertainty.

In such situations the map uncertainty may still be modelled geostatistically if there are sufficient independent observations of the soil property. This boils down to building a geostatistical model of the differences between the soil property map and the independent observations and kriging these errors (Heuvelink 2014). Here, it is essential that the observations are truly independent, i.e. have not been used for map making, because otherwise it might result in a severe underestimation of the map uncertainty. In practice, truly independent data are rarely available unless a new sampling campaign is initiated after the map was made. If uncertainty quantification is important, it is worthwhile to spend extra budget on collecting new data and quantifying the map uncertainty as described above. Note that this will not only quantify the map uncertainty but will also improve the map accuracy, because the existing map could be adjusted by adding the interpolated error to it. In a way, this approach comes close to the KED approach described in Sect. 14.2.5, but now using a single external covariate that is an existing map of the target soil property.

When soil property maps are derived without an underlying geostatistical model and there are no independent observations to build a geostatistical model of the map error, then the only resort is to base the uncertainty model of the map on expert judgement (Truong and Heuvelink 2013). This introduces subjectivity because different experts tend to have different opinions. Also, it will prove to be practically impossible to extract from experts a full probabilistic uncertainty model that also includes spatial and cross-correlations. Expert elicitation procedures are cumbersome and often limited to estimation of quantiles of the (marginal) distribution (O'Hagan et al. 2006).

14.3 Uncertainty Assessment Through Statistical Validation

Uncertainty quantification as described in Sect. 14.2 takes a model-based approach, by defining a geostatistical model of the soil property of interest and deriving an interpolated map and the associated uncertainty from that or by constructing a geostatistical model of the error in an existing map. The approach yields a complete probabilistic characterisation of the map uncertainty, but such characterisation is only valid under the assumptions made. Perhaps the stationarity assumption of ordinary kriging can be relaxed by using a more elaborate geostatistical model, such as that underlying kriging with external drift, but such a model typically needs more data, and in the end no modelling approach is free of assumptions. Therefore it is worthwhile to discuss an alternative, model-free approach to assess the accuracy of soil property maps. This is achieved through (statistical) *validation*.

Validation is defined here as an activity in which the soil map predictions are compared with independent observations. Unlike in Sect. 14.2.6, the outcomes are not used to build a geostatistical model of the map error, but instead summary measures of the observed errors are computed and reported. Common summary measures are the mean error and the root mean squared error. As before, it is essential that the validation observations are independent and have not been used in map making. The safest way to ensure this is to collect validation data after the map was made.

In practice, often we are not that much interested in how well the map predicts the soil property at the limited set of validation locations, but instead we want to know how well the map performs for the entire study area. Summary measures of the entire area cannot be computed but only estimated, because we cannot afford to collect validation observations everywhere. It is then strongly advised to select the validation locations using probability sampling (Brus et al. 2011). The important advantages are that in such case unbiased estimation of summary measures can be ensured and that confidence intervals around the estimated summary measures can be calculated, which is also a prerequisite for significance testing (e.g. to compare whether map A is more accurate than map B). The simplest probability sampling design is *simple random sampling*, but efficiency can be improved by using more elaborate designs. In practice, *stratified simple random sampling* is often used. Model-free estimation of map accuracy has the important advantage that it makes no assumptions, but the disadvantages are that a probability sample is required and that the method can only produce summary measures of the map accuracy.

Validation is based on a comparison of map predictions with independent observations. Typically the observed differences are attributed to map errors. However, it is important to recognise that part of the differences may also be caused by errors in the observations. It is not difficult to incorporate this if the observation error is known in statistical terms (i.e. bias and variance). If observation error is negligibly small compared to map error, as may be the case when a poor map is validated with observations analysed in a high-quality lab, then the influence of observation error on validation statistics may be ignored.

Although it is advised to collect an independent validation data set using probability sampling, this does not mean that summary measures of map accuracy cannot be calculated in case of a non-probability sample, such as a convenience or purposive sample. But in such case, it is important to be aware that there is a risk that the measures may not represent the overall map accuracy very well, such as when the validation data are from specific parts of the study area that have a different map accuracy as other parts.

14.3.1 Cross-Validation

Summary accuracy measures may also be derived using *cross-validation* (see Sect. 11.5.2). In the case of leave-one-out cross-validation, all observations are put aside one by one and the remaining data are used to calibrate the soil mapping model and predict at the location that was put aside. Validation measures are then computed by comparing the predictions with the put-aside observations for all observation locations.

Table 14.1 shows the accuracy measures for Θ_{fc} and Φ as obtained using leave-one-out cross-validation. The *mean error* is close to zero for both properties, indicating that cokriging is unbiased. The *root mean squared error* is 0.050 cm^3/cm^3 for Θ_{fc} and 0.044 cm^3/cm^3 for Φ. These values are not much smaller than the spatial variation of these properties, which can be gleaned from comparison with the square root of the variogram sills shown in Fig. 14.2 (because the sill of the variogram is approximately equal to the variance of the variable of interest). Poor prediction performance is also evidenced by the low values for the *amount of variance explained*, which is defined as one minus the ratio of the mean squared error and the variance. Apparently the sampling density is insufficient to capture a large part of the spatial variation. In fact this is already foretold by the variograms in Fig. 14.2, which have fairly high nugget variances and small ranges. The poor prediction accuracy is also confirmed by the scatter plots of cross-validation predictions against observations (Fig. 14.9). Note also that the cross-validation accuracy measures might still be somewhat too optimistic about the overall map accuracy, since all observations are on transects, and hence any cross-validation location always has at least a few nearby observations. The last column of Table 14.1 shows the

Table 14.1 Accuracy measures of cokriging predictions of Θ_{fc} and Φ as obtained with leave-one-out cross-validation

	Mean error (cm^3/cm^3)	Root mean squared error (cm^3/cm^3)	Amount of variance explained (−)	Standardised root mean squared error (−)
Water content at field capacity (Θ_{fc})	−0.002	0.050	0.30	1.05
Porosity (Φ)	−0.001	0.044	0.37	1.05

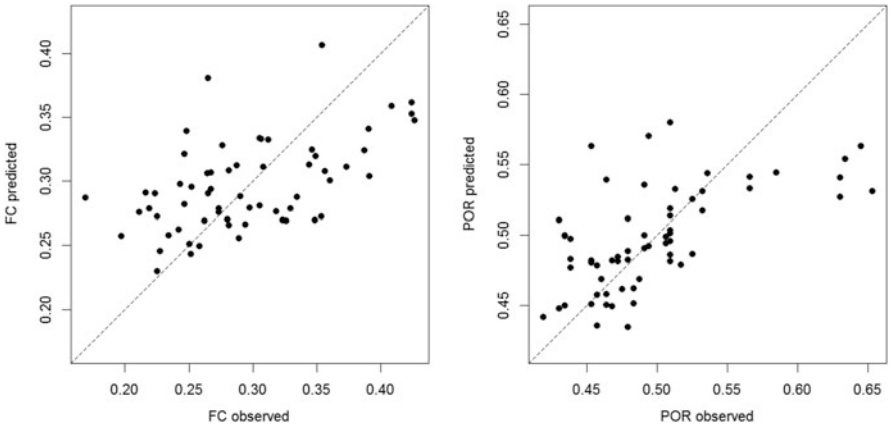


Fig. 14.9 Scatter plots of predicted against observed Θ_{fc} (*left*) and Φ (*right*) as obtained using leave-one-out cross-validation

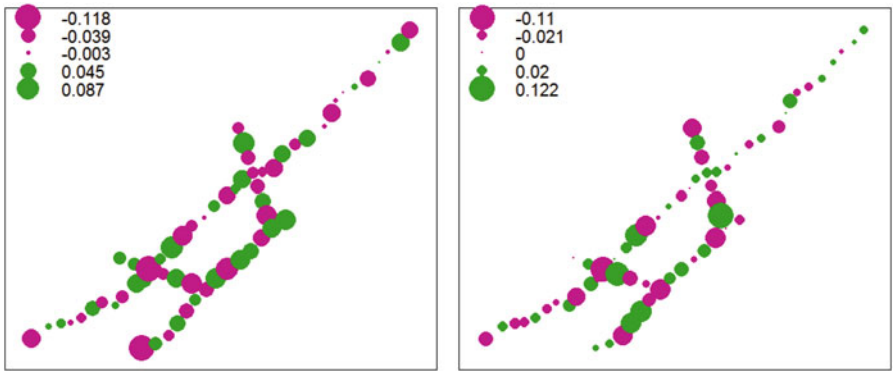


Fig. 14.10 Spatial plot of leave-one-out cross-validation errors for Θ_{fc} (*left*) and Φ (*right*)

standardised root mean squared error (SRMSE), which is obtained by taking the square root of the average squared z_{score}, where z_{score} is defined as the difference between the observed and predicted soil property, divided by the cokriging standard deviation. If the cokriging standard deviation is a proper measure of the map prediction error, then SRMSE should be close to one. The obtained values are fairly close to one and do not indicate a significant over- or underestimation of the uncertainty. Figure 14.10 shows bubble plots of the spatial distribution of the cross-validation errors.

14.4 Uncertainty Propagation

Previous sections of this chapter explained that uncertainty about soil properties can be conveniently represented using probability distributions. Specific attention was paid to quantification of spatial interpolation errors using geostatistics. The methods were illustrated using a case study on mapping water content at field capacity (Θ_{fc}) and soil porosity (Φ) in the Allier study area, France. This chapter takes the analysis a step further by analysing how uncertainties in soil properties propagate through an environmental model that uses these soil properties as input (Heuvelink 1998). More specifically, we will analyse how uncertainties in Θ_{fc} and Φ propagate through a multiple linear regression model that predicts the soil water content at wilting point (Θ_{wp}) from Θ_{fc} and Φ. Before we do this, we first present the statistical uncertainty propagation methodology.

The uncertainty propagation analysis can be formulated mathematically as follows. Let U be the output of an environmental model g on m input variables S_i:

$$U = g(S_1, S_2, \ldots, S_m) \tag{14.2}$$

The model g may be of various types, ranging from a simple pedotransfer function to a complex soil erosion or crop yield model. The objective of the uncertainty propagation analysis is to determine the uncertainty in the output U, given the operation g and the inputs S_i and their associated uncertainties. Let us denote the means and variances of the S_i by μ_i and σ_i^2, respectively. Since the inputs are random variables or random fields, the output will be a random variable or random field as well. Important parameters of U are its mean ξ and variance τ^2. From an uncertainty propagation perspective, the main interest is in the uncertainty of U, as contained in its variance τ^2.

It must first be observed that the uncertainty propagation problem is relatively easy when g is a linear function of its inputs S_i. In that case the mean and variance of U can be directly and analytically derived. In case of non-linear models, analytically driven methods exist only in a few cases, and one must nearly always rely on approximation methods for a complete evaluation. Two of these methods will now be discussed.

14.4.1 Taylor Series Method

The idea of the Taylor series method is to approximate g by a truncated Taylor series centred at the means μ_i. In case of the first-order Taylor method, g is linearised by taking the tangent of g in μ_i. Fig. 14.11 illustrates this for the one-dimensional case ($m = 1$). The linearisation greatly simplifies the uncertainty analysis, but only at the expense of introducing an approximation error.

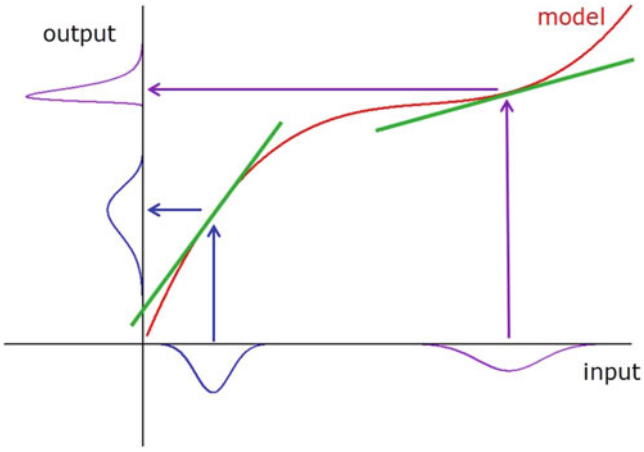


Fig. 14.11 Graphical illustration of the first-order Taylor method for the case where the model has a single input. The model (*red line*) is approximated by a linear function (*green line*) that has a small approximation error near the centre of the input probability distribution. Two cases are depicted: model is sensitive to changes in input (*blue*), and model is insensitive to changes in input (*purple*)

Using the first-order Taylor series method, the variance τ^2 of the output U is given by (Heuvelink 1998):

$$\tau^2 \cong \sum_{i=1}^{m} \sum_{j=1}^{m} \rho_{ij} \sigma_i \sigma_j g_i' g_j' \tag{14.3}$$

where ρ_{ij} is the correlation coefficient between the uncertainties in S_i and S_j and g_i' is the first derivative of g with respect to S_i, which is evaluated in the means $\mu_i, i = 1, \ldots, m$. Equation 14.3 shows that the variance of U is the sum of various terms, which contain the correlations and standard deviations of the S_i and the first derivatives of g. These derivatives reflect the sensitivity of U to changes in the inputs (see Fig. 14.11 for a graphical illustration). From Eq. 14.3 it also appears that the correlations of the input uncertainties can have a marked effect on the variance of U.

To decrease the approximation error invoked by the first-order Taylor method, one option is to extend the Taylor series of g to include a second-order term as well. This is particularly useful when g is a quadratic function, in which case the second-order Taylor method is free of approximations and the first-order Taylor method is not. The application to the Allier case study discussed later in this section gives an example. However, it should be noted that in other cases including a second-order term might worsen the results. For instance, this might happen if the variance of the input is large and the approximation by a quadratic function, which is more accurate locally, is less accurate than the linear approximation at a greater distance from the approximation point.

14.4.2 Monte Carlo Method

The Monte Carlo method uses an entirely different approach to analyse the propagation of uncertainty. The idea of the method is to compute the result of the model repeatedly, with input values s_i that are randomly sampled from their joint distribution. The model results form a random sample from the distribution of U, so that parameters of the distribution, such as the mean ξ and variance τ^2, can be estimated from the sample.

The method thus consists of the following steps:

1. Repeat N times:
   (a) Generate a set of realisations $s_i, i = 1, \ldots, m$.
   (b) For this set of realisations s_i, compute and store the output $u = g(s_1, \ldots, s_m)$.
2. Compute and store sample statistics from the N outputs u.

A random sample from the m inputs S_i can be obtained using an appropriate pseudorandom number generator (Lewis and Orav 1989; Ross 1990). Note that a conditioning step will have to be included when the S_i are correlated. In case of spatial inputs, these may be sampled using spatial stochastic simulation as explained in Sect. 14.2.3.

The accuracy of the Monte Carlo method is inversely related to the square root of the number of runs N. This means that to double the accuracy, four times as many runs are needed. The accuracy thus slowly progresses as N increases.

14.4.3 Evaluation and Comparison of Uncertainty Propagation Techniques

The main problems of the Taylor method are that it only works with models that are continuously differentiable with respect to their uncertain inputs, that it only provides estimates of the mean and variance of the model output and that the results are approximate only. It will not always be easy to determine whether the approximations involved using this method are acceptable. The Monte Carlo method also involves approximation errors, but these can be made arbitrarily small by increasing the number of Monte Carlo runs.

The Monte Carlo method brings along other problems, though. High accuracies are reached only when the number of runs is sufficiently large, which may cause the method to become extremely time-consuming. This will remain a problem even when variance reduction techniques such as Latin hypercube sampling are employed. Another disadvantage of the Monte Carlo method is that the results do not come in an analytical form.

As a general rule, it seems that the Taylor method may be used to obtain crude preliminary answers for simple models. These should provide sufficient detail to be able to obtain an indication of the quality of the model output. When exact values or

quantiles and/or percentiles are needed, the Monte Carlo method may be used. The Monte Carlo method will probably also be preferred when uncertainty propagation with complex models is studied, because the method is easily implemented and generally applicable. It is no more than an extra loop around an existing model. This, and the fact that computer power is ever increasing, means that nowadays the majority of uncertainty propagation studies uses the Monte Carlo method. Some examples from soil science are Brown and Heuvelink (2005), Bishop et al. (2006), Hastings et al. (2010), Kros et al. (2012), Van Den Berg et al. (2012), Brodsky et al. (2013), Poggio and Gimona (2014), Malone et al. (2015) and Xiong et al. (2015).

14.4.4 Sources of Uncertainty Contributions: The Balance of Errors

When the uncertainty propagation analysis reveals that the output of g contains too large an error, then measures will have to be taken to improve accuracy. When there is a single input to g, then there is no doubt where the improvement must be sought, but what if there are multiple inputs? Also, how much should the uncertainty of a particular input be reduced in order to reduce the output uncertainty by a given factor? It is useful to explore these questions briefly.

To obtain answers to the questions above, consider Eq. 14.3 again, which gives the variance of the output U using the first-order Taylor method. When the inputs are uncorrelated, this reduces to:

$$\tau^2 \cong \sum_{i=1}^{m} \sigma_i^2 g'^2_i \tag{14.4}$$

This shows that the variance of U is a sum of parts, each to be attributed to one of the inputs S_i. This *partitioning property* allows one to analyse how much each input contributes to the output variance. Thus from Eq. 14.4, it can directly be seen how much τ^2 will reduce from a reduction of σ_i^2. Clearly the output will mainly improve from a reduction in the variance of the input that has the largest contribution to τ^2. Note that this need not necessarily be the input with the largest error variance, because the sensitivity of the model g to changes in the input is also important. Figure 14.11 shows an example where in the purple case the input uncertainty is greater than in the blue case, but where the output uncertainty still is the greatest in the blue case. This is because in the blue case, the model is more sensitive to changes in the input. Note also that Eq. 14.4 is derived under rather strong assumptions. When these assumptions are not realistic, it may be advisable to derive the uncertainty source contributions using a modified Monte Carlo approach.

14.4.5 Application of Uncertainty Propagation to the Allier Case Study

Recall from Sect. 14.2.1 that our aim is to map the soil moisture content at wilting point Θ_{wp} from maps of the moisture content at field capacity Θ_{fc} and soil porosity Φ. We obtained maps of both input soil properties and their associated uncertainties using cokriging in Sect. 14.2. We now need to discuss how these can be used to derive a map of Θ_{wp} and its associated uncertainty. Recall that we use a multiple linear regression model to predict Θ_{wp} from Θ_{fc} and Φ. The model is very simple, and hence we can use the Taylor series method to analyse the uncertainty propagation.

Figure 14.1 shows 12 circled sites where all three properties Θ_{wp}, Θ_{fc} and Φ were determined in the laboratory. These measurements were used to set up a pedotransfer function, relating Θ_{wp}, Θ_{fc} and Φ, which took the form of a multiple linear regression:

$$\Theta_{wp} = \beta_0 + \beta_1 \Theta_{fc} + \beta_2 \Phi + \varepsilon \tag{14.5}$$

The coefficients β_0, β_1 and β_2 were estimated using standard ordinary least squares regression. The estimated values for the regression coefficients and their respective standard deviations were $\beta_0 = -0.263 \pm 0.031$, $\beta_1 = 0.408 \pm 0.096$ and $\beta_2 = 0.491 \pm 0.078$. The standard deviation of the stochastic residual ε was estimated as 0.0114. The correlation coefficients of the estimation errors of the regression coefficients were $\rho_{01} = -0.221$, $\rho_{02} = -0.587$ and $\rho_{12} = -0.655$. The regression model explains 94.8% of the variance of the observed Θ_{wp}, indicating that the model is satisfactory. Note that presence of spatial correlations between the observations at the 12 locations was ignored in the regression analysis.

The maps of Θ_{fc} and Φ as derived in Sect. 14.2.1 were substituted in the regression (Eq. 14.5) yielding a map of Θ_{wp}. The associated uncertainty was computed using the Taylor series method. Note that Eq. 14.5 is a quadratic function of six uncertain inputs. To avoid approximation errors, it was therefore decided to use the second-order Taylor method, which is a logical extension of the first-order Taylor method. Because the model coefficients and the field measurements were determined independently, the correlation between the β_i and cokriging prediction errors was taken to be zero. Also, the stochastic residual ε is uncorrelated with all other uncertain inputs.

The results of the uncertainty propagation are given in Fig. 14.12. The accuracy of the map of Θ_{wp} is reasonable: the standard deviation of Θ_{wp} rarely exceeds 50% of the predicted value. The uncertainty is much larger in those parts of the study area where there are no observations. This suggests that uncertainty in the maps of Θ_{fc} and Φ are the main source of uncertainty because these uncertainty maps had similar spatial patterns. Indeed Fig. 14.13 shows that the contribution of the regression model uncertainty is small. Improvement of the Θ_{wp} map can thus best be done by improving the maps of Θ_{fc} and Φ, for instance, by taking more measurements over

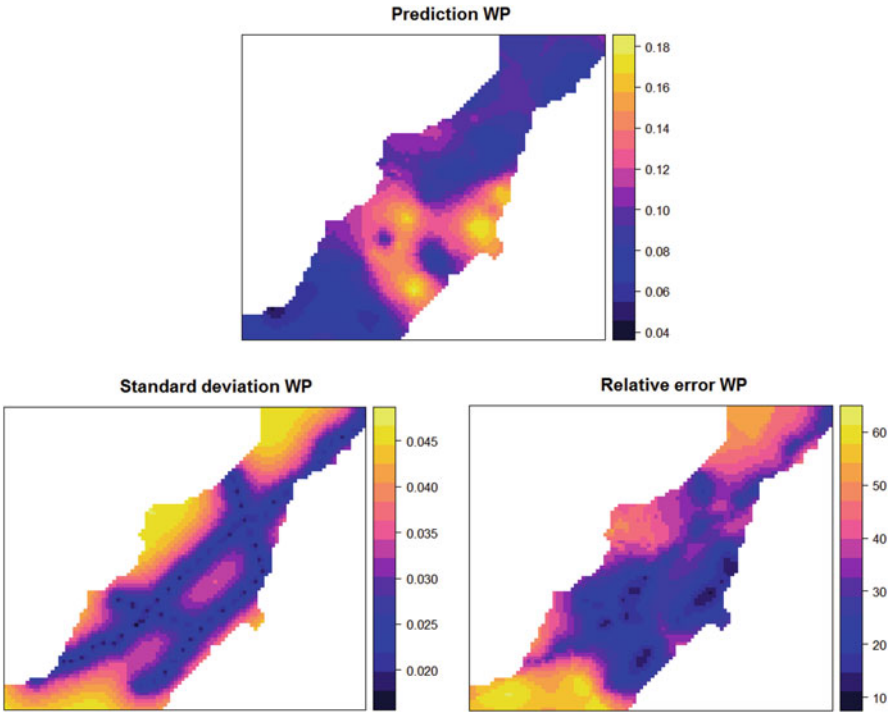


Fig. 14.12 Prediction map (*top*) of soil moisture at wilting point computed from cokriging maps of soil moisture at field capacity and porosity using a pedotransfer function, associated prediction error standard deviation map (*bottom left*) derived using the second-order Taylor method and relative error (*bottom right*)

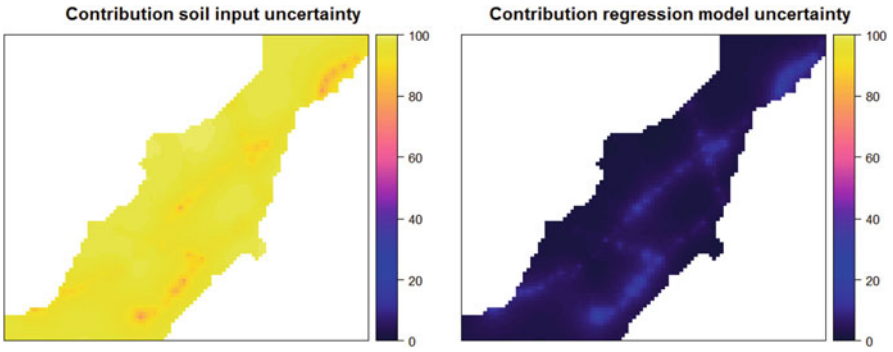


Fig. 14.13 Contributions in percentages to the overall error variance of the soil moisture at wilting point predictions as caused by cokriging errors in soil input maps (*left*) and by uncertainty in the multiple linear regression model (*right*)

the study area. The variograms and cross-variogram of Θ_{fc} and Φ could be used to assist in optimising sampling. This technique would allow judging in advance how much improvement is to be expected from the extra sampling effort.

14.5 Conclusions

No soil map is perfect. It is important to quantify the errors and uncertainties associated with soil maps because this determines whether a map is usable for an intended purpose. Any end user of soil maps should therefore require that the maps are accompanied by accuracy measures. Such measures can be computed from comparison of map predictions with independent validation data, but for spatially explicit uncertainty measures, a geostatistical approach that quantifies the map accuracy through the kriging standard deviation is recommended. Geostatistics also provides the tools to generate 'possible realities' by sampling from the conditional spatial probability distribution of the uncertain soil property. These possible realities may be used to communicate uncertainty and are also useful in Monte Carlo uncertainty propagation analyses.

Uncertainty propagation analysis is used to analyse how uncertainty in input (soil) maps propagates through spatial analyses and environmental models. This not only quantifies the uncertainty in the model output but can also tell us which are the main sources of uncertainty, which is essential information for taking informed decisions about how to improve the quality of maps and model results.

This chapter was limited to uncertainty quantification and uncertainty propagation of continuous-numerical soil properties and variables, but generalisation to categorical variables can be made, although it is more complicated.

This chapter concentrated on errors and uncertainties that arise from spatial interpolation and from fitting and applying linear regression models. There are many more sources of uncertainty, such as field and lab measurement error, positional error, classification error and model parameter and structural errors. These can be handled in similar ways, but the main challenge often is to characterise the error sources with realistic probability distributions. Once this is done, the uncertainty propagation analysis itself is not difficult, although it might be computationally demanding.

References

Bishop TFA, Minasny B, McBratney AB (2006) Uncertainty analysis for soil-terrain models. Int J Geogr Inf Sci 20:117–134

Brodsky L, Vasat R, Klement A, Zadorova T, Jaksik O (2013) Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping. Geoderma 199:54–63

Brown JD, Heuvelink GBM (2005) Assessing uncertainty propagation through physically based models of soil water flow and solute transport. In: Anderson MG et al (eds) Encyclopaedia of hydrological sciences. Wiley, Chichester, pp 1181–1195

Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138:86–95

Brus DJ, Kempen B, Heuvelink GBM (2011) Sampling for validation of digital soil maps. Eur J Soil Sci 62:394–407

Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. Geoderma 103:3–26

Hastings AF, Wattenbach M, Eugster W, Li C, Buchmann N, Smith P (2010) Uncertainty propagation in soil greenhouse gas emission models: an experiment using the DNDC model and at the Oensingen cropland site. Agric Ecosyst Environ 136:97–110

Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil properties based on regression-kriging. Geoderma 120:75–93

Heuvelink GBM (1998) Error propagation in environmental modelling with GIS. Taylor & Francis, London. 127 pp

Heuvelink GBM (2014) Uncertainty quantification of GlobalSoilMap products. In: Arrouays D, McKenzie N, Hempel J, Richer de Forges A, McBratney A (eds) GlobalSoilMap. Basis of the global spatial soil information system. CRC Press, Boca Raton, pp 335–340

Kros J, Heuvelink GBM, Reinds GJ, Lesschen JP, Ioannidi V, De Vries W (2012) Uncertainties in model predictions of nitrogen fluxes from agro-ecosystems in Europe. Biogeosciences 9: 4573–4588

Lewis PAW, Orav EJ (1989) Simulation methodology for statisticians, operations analysts, and engineers, vol. 1. Wadsworth and Brooks/Cole, Pacific Grove

Malone BP, Kidd DB, Minasny B, McBratney AB (2015) Taking account of uncertainties in digital land suitability assessment. Peer J 3:e1366

Minasny B, McBratney AB (2006) A conditioned latin hypercube method for sampling in the presence of ancillary information. Comput Geosci 32:1378–1388

O'Hagan A, Buck C, Daneshkhah A, Eiser L, Garthwaite P, Jenkinson D, Oakley J, Rakow T (2006) Uncertain judgements: eliciting experts' probabilities. Wiley, Chichester

Odeh IOA, McBratney AB, Chittleborough DJ (1995) Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67:215–226

Orton TG, Pringle MJ, Bishop TFA (2016) A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262:174–186

Poggio L, Gimona A (2014) National scale 3D modelling of soil organic carbon stocks with uncertainty propagation – an example from Scotland. Geoderma 232:284–299

Ross SM (1990) A course in simulation. Macmillan, New York

Truong PN, Heuvelink GBM (2013) Uncertainty quantification of soil property maps with statistical expert elicitation. Geoderma 202–203:142–152

Van den Berg F, Tiktak A, Heuvelink GBM, Burgers SLGE, Brus DJ, de Vries F, Stolte J, Kroes JG (2012) Propagation of uncertainties in soil and pesticide properties to pesticide leaching. J Environ Qual 41:253–261

Wackernagel H (2003) Multivariate geostatistics. An introduction with applications. Springer, Berlin

Xiong X, Grunwald S, Brenton MD, Kim J, Harris WG, Bliznyuk N (2015) Assessing uncertainty in soil organic carbon modeling across a highly heterogeneous landscape. Geoderma 251:105–116

Chapter 15
Complex Soil Variation over Multiple Scales

R. Murray Lark and Alice E. Milne

"So, naturalists observe, a flea
Has smaller fleas that on him prey;
And these have smaller still to bite 'em,
And so proceed ad infinitum".

Jonathan Swift

15.1 Introduction

15.1.1 Spatial Frequency and the Fourier Transform

Like Swift's fleas, the soil is organized at multiple spatial scales from the clay particle, interacting with its neighbours and the soil solution according to the laws of electrochemistry, to the continent, at which the properties of the soil are organized according to general climate trends and the contingencies of geological history. Pedometrics can help the soil scientist to understand these processes in so far as it is possible to analyse soil properties into scale-dependent components which can be modelled and visualized. This is done, for example, by the spatially nested sampling introduced by Youden and Mehlich (1937). Geostatistical methods achieve it to some extent, with the use of nested models of regionalization and coregionalization, and methods of analysis to visualize components of different spatial scales (factorial kriging, reference) and scale-dependent correlation between variables (e.g. Goovaerts and Webster 1994). However, geostatistical analysis is primarily undertaken to support spatial prediction; the variogram, while reflecting

R.M. Lark (✉)
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
e-mail: mlark@bgs.ac.uk

A.E. Milne
Sustainable Agricultural Sciences, Rothamsted Research, Harpenden, UK
e-mail: alice.milne@rothamsted.ac.uk

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_15

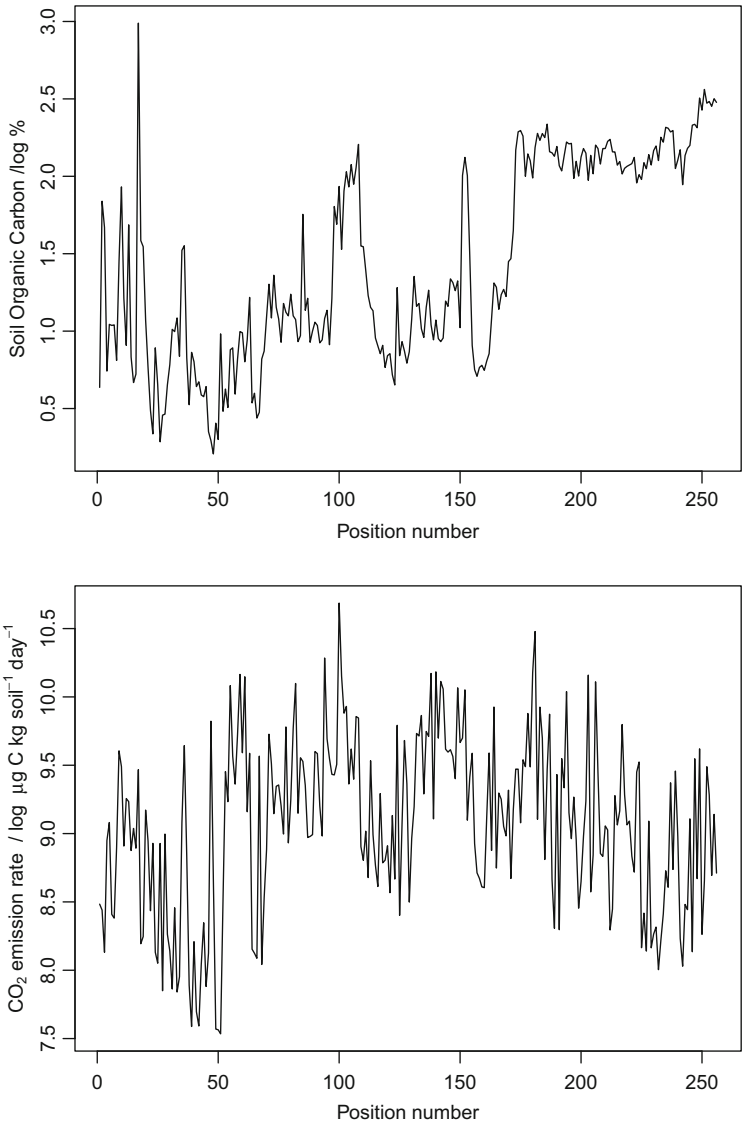


Fig. 15.1 Measurements of (*top*) soil organic carbon content and (*bottom*) rate of carbon dioxide emission from soil cores collected at intervals of 30 m on a transect in eastern England

the influence of processes at multiple spatial scales, is not particularly suited to the interpretation of such processes. For this we must look elsewhere.

Consider a soil property measured from sample sites at equal intervals on a linear transect. Figure 15.1 shows two variables from such a data set, the rate of carbon dioxide emission from incubated soil cores (transformed to natural logarithms)

collected at approximately 30 m intervals on a transect of just over 7 km long across part of Bedfordshire in eastern England. Haskard et al. (2010) and Milne et al. (2011) describe the data and their collection. The variations of this soil property depend on a range of factors from differences in the underlying parent material, land use, microtopography, etc. The great insight of the French mathematician and physicist Joseph Fourier (1768–1830) was to show that such a sequence of measurements, subject to certain assumptions, can be decomposed into a sum of components as follows:

$$z(x) \;=\; a_0 + \sum_{m=1}^{\lfloor T/2 \rfloor} (a_m \cos 2\pi f_m x + b_m \sin 2\pi f_m x)\,, \tag{15.1}$$

where $z(x)$ denotes the value of the soil property at location x on the transect, T is the length of the transect (units are the sampling interval), $\lfloor \cdot \rfloor$ is the integer part of the enclosed term and $f_m = \frac{m}{T}$, $m = 1, 2, \ldots \lfloor T/2 \rfloor$. In this setting, f_m is a *spatial frequency*, the analog in spatial terms of the frequency more familiarly encountered in the analysis of time series. The largest spatial frequency in this sequence is $\frac{1}{2}$ or $\frac{1}{2x_0}$ where x_0 denotes the basic sampling interval. This maximum frequency, called the Nyquist frequency, represents the most detailed (fine-scale) component of soil variation that can be resolved by the analysis of transect data. The constant, a_0, is the mean of the variable (i.e. the component of zero frequency or infinite wavelength).

The component $(a_m \cos 2\pi f_m x + b_m \sin 2\pi f_m x)$ is a sinusoidal function of wavelength T/m and amplitude $\sqrt{a_m^2 + b_m^2}$, so the Fourier representation is a decomposition of the original signal into sinusoidal components of different frequency. The sinusoid comprises sine and cosine functions of the specified frequency, which are called *basis functions* of the transformation. The amplitude of the component of frequency f_m is $\sqrt{a_m^2 + b_m^2}$. Let us call this c_m. It can be shown that the variance of the component of frequency f_m is

$$\frac{c_m^2}{2},\ m < T/2; \quad c_m^2,\ m = T/2$$

A plot of variance (power) against frequency is called the *spectrum* of the time series. Such a spectrum is shown in Fig. 15.2 for the (transformed) emission rates. In this case, the frequency is standardized so that the Nyquist frequency is 0.5. The horizontal dashed line shows the threshold at for evidence for a significant peak in the spectrum. In addition to the large power at the lowest frequency, there is a peak at a relatively low frequency (corresponding to a wavelength of about 1300 m). This may all seem very promising (at least for the analysis of soil data measured on transects). We have a method to represent such data as a sum of components of different spatial frequency and to represent the partition of the variance of the property, a familiar statistic, between these frequencies. Such an analysis has been used to examine soil properties (Webster 1977; Milne et al. 2010). In both these studies, the underlying process of interest was the periodic variation arising from

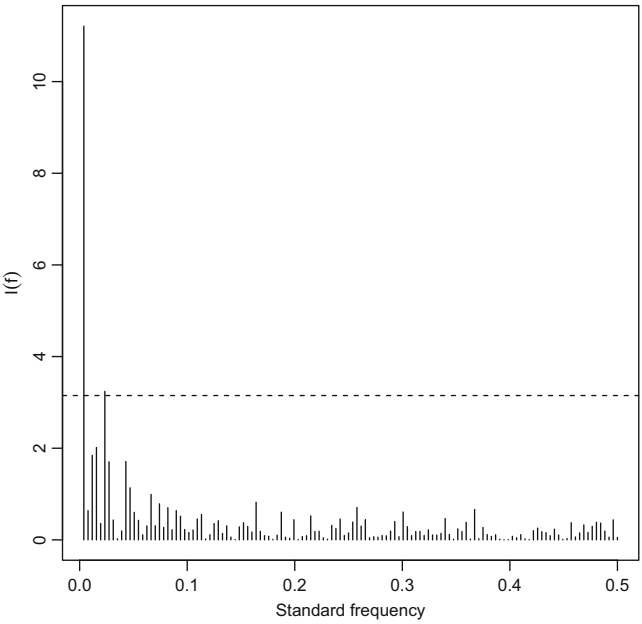


Fig. 15.2 Power spectrum for the data on (transformed) emission rate

the gilgai landscapes of Australia. The Fourier analysis is particularly useful in such a context because the periodic component of the signal is particularly apparent as a peak in the spectrum. However, in other settings where strictly periodic variation is absent or relatively insignificant, the power spectrum does not offer particular insight. In fact, the power spectrum is directly related, via the Fourier transform, to the covariance function of a variable and so, under weak stationarity, to its variogram, so, in general, the spectrum will not provide us with any information that the variogram does not.

At this point, the reader could be forgiven some impatience with us. We have led you through some tedious notation but taken you no further than earlier chapters on geostatistics. The reason for this is that the Fourier transform provides a gateway to some more powerful methods for analysis in terms of spatial frequencies which are valuable in themselves but also help to exhibit some of the limitations imposed by stationarity assumptions in geostatistics.

Consider Eq. (15.1) again. Two coefficients represent the contribution of sinusoidal variation at some spatial frequency to the variation of a soil property. These coefficients determine the amplitude of that variation (and so the variance) and the phase (i.e. where the peaks and troughs of the sinusoid occur relative to the origin of the transect). With those two coefficients fixed, the variability of the soil property at the frequency of interest is fixed everywhere on the transect. This is directly equivalent to the stationarity assumptions made in geostatistics, including

the weakest assumption of intrinsic stationarity under which the expected square difference between two observations made some fixed lag apart depends only on that lag and not on absolute position.

Consider a real transect across a soil landscape. This might start near a drainage channel, traversing a floodplain with contemporary alluvium including braided deposits, before traversing a series of terrace gravels deposited by the outwash of previous glaciations and then eroded by subsequent drainage. Above the oldest terrace, we move to a Jurassic clay, partly covered by solifluction deposits from overlying sandstone strata which themselves are capped by a coarse limestone with some eroded residuals of an overlying clay and later aeolian deposits. This is the first author's default soil landscape in which he worked as a student. The reader may have others. All are likely to exhibit similar complexity, and the assumption that a soil property (e.g. clay content) can be meaningfully represented by components at spatial frequencies which oscillate with uniform amplitude and phase from one part of this notional transect to another is unconvincing. The algebra of Fourier's transform will still work; the question is whether the representation of variation in terms of different spatial frequencies can be given a realistic *pedological* interpretation.

Alternative methods for analysis, primarily the wavelet transformation, are presented in this chapter. The wavelet transforms are analyses in terms of spatial frequency, but their basis functions are not sines and cosines with uniform amplitude and phase over the whole transect but rather are local functions which oscillate over a short window or *support* and go to zero elsewhere. With an appropriate set of wavelet basis functions, we can achieve an analysis of a data set which shows scale dependency of its variation but also allows us to examine how this variation may change from place to place in a way which, while unsurprising in a pedological context, is not consistent with assumptions of stationarity. The following sections of this chapter introduce and illustrate the wavelet transform. We then complete the chapter with an examination of an alternative analysis in terms of spatial frequencies, the empirical mode decomposition, and compare it with wavelet analysis.

15.2 Wavelets and the Discrete Wavelet Transform

15.2.1 Wavelet Functions and Their Properties

As noted in the introduction, wavelet transformation uses basis functions which differ from the sines and cosines of Fourier analysis. Wavelet functions allow us to examine variation at distinct spatial frequencies but do not require, for their interpretation, the assumption that the variation at this frequency is uniform in space. This is achieved by using basis functions that are dilations (changes in frequency) and translations (changes in position) of a *mother wavelet function*. There are various such functions. The mother wavelet is not an arbitrary function but must satisfy three criteria.

1. Its mean is zero, i.e.

$$\int_{-\infty}^{\infty} \psi(x)\mathrm{d}x \;=\; 0 \tag{15.2}$$

   from which we can infer that a wavelet function oscillates.
2. Its squared norm is 1, i.e.

$$\int_{-\infty}^{\infty} |\psi(x)|^2\mathrm{d}x \;=\; 1\,. \tag{15.3}$$

3. It has a compact support, by which we mean that the mother wavelet damps rapidly to zero and so operates very locally.

   The support of a wavelet depends on M, its number of *vanishing moments* where

$$\int_{-\infty}^{\infty} x^c \psi(x)\mathrm{d}x = 0 \tag{15.4}$$

for $c = 0, 1, 2, \ldots, M-1$. From Eq. (15.2), it follows that any wavelet function must have at least one vanishing moment. There are several families of wavelet functions which meet the requirements detailed above and which differ with respect to the number of vanishing moments. Within a family of wavelet functions, increasing M makes the wavelet smoother and increases the support.

Recall that the Fourier transform involves finding the sets of coefficients that can be used to represent a series of observations by summing sines and cosines of different frequency. In effect these basis functions are mathematical building blocks from which the data are rebuilt, and the coefficients tell us something about the variation of the data at different spatial frequencies (subject to stationarity assumptions). The same is true of wavelet functions. It is clear that a single wavelet function can only function as a basis over some limited range of spatial frequencies (depending on how it oscillates) and over the support of the wavelet (where it takes non-zero values). To build up a complete wavelet basis for the analysis of some variable, we use *dilations* of the mother wavelet function (i.e. rescalings to cover different spatial frequencies) and *translations* of these (i.e. shifts of the dilated wavelet to provide a support in the basis at all locations of interest).

The set of dilated and translated versions of a mother wavelet $\psi(x)$ is denoted by $\psi_{\lambda,u}(x)$, where

$$\psi_{\lambda,u}(x) \;=\; \frac{1}{\sqrt{\lambda}}\psi\left(\frac{x-u}{\lambda}\right), \quad \lambda > 0, \quad u \in \mathcal{R}\,, \tag{15.5}$$

where $\mathcal{R}$ is the set of real numbers. Here λ defines the dilation of the wavelet. If the mother wavelet takes non-zero values over a support of unit length, then changing its argument to $\left(\frac{x-u}{\lambda}\right)$ changes the support to length λ. This process also changes the

range of spatial frequencies to which the wavelet responds. Similarly, changing the value u changes the location at which the support of the wavelet is found.

The wavelet transform of $g(x)$, $Wf(x|\lambda, u)$, is obtained by integrating $g(x)$ with the wavelet; thus

$$\begin{aligned} Wg(x|\lambda, u) &= \int_{-\infty}^{\infty} g(x)\psi_{\lambda,u}(x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x)\frac{1}{\sqrt{\lambda}}\psi\left(\frac{x-u}{\lambda}\right). \end{aligned} \tag{15.6}$$

If $g(x)$ is some continuous function, then it can be analysed with continuous variation of the dilation parameter $\lambda > 0$ and translation parameter, u. For the analysis of real data, a discrete version of the transformation is required. Some wavelet analyses, called continuous wavelet transforms, use as close to a continuous translation and dilation of the wavelet as possible. These methods have proved useful for many purposes and have been applied in soil science (e.g. Si 2003; Si and Farrell 2004). However, most methods for statistical inference from wavelets have used a particular approach to discretization commonly referred to as the discrete wavelet transform (DWT), and we consider this approach, and methods derived from it, in the remainder of this chapter.

15.2.2 The Discrete Wavelet Transform and Multiresolution Analysis

In the discrete wavelet transform (DWT), we consider a discrete set of dilations of the mother wavelet function $\psi(x)$ and an associated set of translations. The general DWT basis function is

$$\psi_{m,n}(x) = \frac{1}{\sqrt{2^m}}\psi\left(\frac{x-n2^m}{2^m}\right), \tag{15.7}$$

which is the basis function for the mth dilation of the wavelet and its nth translation. Under this scheme, the dilations of the discrete wavelet transform are in the dyadic sequence $2^1, 2^2, 2^3, \ldots$. The wavelet coefficients from all wavelets in the mth dilation are said to represent variation at scale $2^m x_0$ where x_0 is the basic sample interval. We refer to scales with large m as 'coarse' scales, contrasted with 'fine' scales. A scale corresponds to a nominal frequency interval; increasing m reduces the frequency. Translation of the wavelet takes place in steps of 2^m, i.e. larger steps at coarser scales. The nominal frequency interval for the mth dilation of the wavelet is $\left[\frac{1}{2^m x_0}, \frac{1}{2^{m+1}x_0}\right]$. The finest scale therefore has a lower bound of $\frac{1}{2x_0}$, the Nyquist frequency, given the sample spacing.

Note that, as with any real partition of spatial frequencies in an analysis, there is some 'leakage' between these intervals, i.e. DWT bases in the *m*th dilation will respond to some extent to variation at frequencies outside the nominal range. The cutoff of the response is sharpest for wavelets in a family with a larger number of vanishing moments.

The DWT, with dilations on the dyadic sequence, has some desirable properties when applied to suitable mother wavelet functions. In particular, the dilations and translations of such a wavelet are orthogonal to each other, that is to say, their cross products sum to zero. This provides the basis for the partition of the variance of a variable into components associated with the wavelet basis functions, as we shall see.

A DWT wavelet coefficient is the inner product of a discrete series of data values, $f(x)$, with the particular dilation and translation of the wavelet:

$$D_{m,n} = \langle f(x), \psi_{m,n} \rangle. \tag{15.8}$$

The dilations and translations of a wavelet are basis functions for the wavelet analysis, and the DWT coefficients provide the basis for representing a data set in terms of these basis functions. The original data sequence can be analysed into components as follows:

$$f(x) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} D_{m,n} \psi_{m,n}(x) \, . \tag{15.9}$$

The attentive reader might be concerned by the two summations over $-\infty, \infty$ in this expression, when applied to finite data sampled at a finite interval. The summation over infinite translations is reflected in the practical difficulty of applying Eq. (15.7) to data near the beginning and end of the transect where components in the dilated wavelet function overlap the ends of the data. A standard procedure is to pad the ends of the data. The most satisfactory way to do this for data analysis is to pad by reflection at each end of the sequence so that, when padding a transect length n by adding values at locations $n+1,\ n+2, \ldots$ and at $n-1,\ n-2, \ldots$ one pads at $n+j$ with the value at location $n-j+1$ on the transect, and similarly, one pads at $n-j$ with the observed value at location $n+j-1$ (see Percival and Walden 2000). Because the DWT basis functions are of finite length, one can therefore evaluate Eq. (15.9) just for those locations in our finite transect by adding sufficient padding observations.

Dealing with the summation over infinite dilations is a little more complex. We note that, for finite data, the smallest value of the dilation parameter that we can consider is $m = 1$. If we modify Eq. (15.9) to include only the wavelet coefficients for the *j*th dilation $m = j$, we obtain

$$Q_j(x) = \sum_{n=-\infty}^{\infty} D_{j,n} \psi_{j,n}(x) \, . \tag{15.10}$$

The infinite translations no longer concern us in practice because we require only those for which the finite support of the wavelet overlaps our transect. The term Q_j is an additive component of the data with variations at scale $2^j x_0$. This is called a 'detail' component of the signal. If we computed the detail components for $m = 1, 2, \ldots, j$ and subtracted these from $f(x)$, the residual would be what is called a 'smooth' component of $f(x)$, P_k, equivalent to

$$P_k(x) = \sum_{m=k+1}^{\infty} \sum_{n=-\infty}^{\infty} D_{m,n} \psi_{m,n}(x) \, . \tag{15.11}$$

As discussed by Lark and Webster (1999), one can obtain the smooth component directly using a single dilation of a scaling function, which is a unique partner to the particular wavelet function. The set of detail components for some set of dilations of the wavelet and the residual smooth component are known as a *multiresolution analysis* of the finite data set into a finite set of scale-specific components (Mallat 1989).

Figure 15.3 shows the wavelet functions and scaling functions for Daubechies's (1988) extremal phase wavelet filter with increasing numbers of vanishing moments. Note how increasing the number of vanishing moments increases the number of coefficients in the filter and also the smoothness of the filter function.

To summarize, we have shown that a mother wavelet function can be dilated and translated in finite steps to provide a basis function for a real data set and that the analysis can be presented as a multiresolution analysis. In subsequent sections, we shall examine statistical analyses of these data based on related basis functions. However, to illustrate what has been done so far, we now consider a DWT analysis and multiresolution representation of our data on soil carbon content (after transformation to logarithms).

First it is necessary to select the mother wavelet function. Here we consider Daubechies's extremal phase family of wavelet functions. Within a family, the key choice is the number of vanishing moments, M, to specify. As M is increased, the wavelet basis function becomes smoother, which may be advantageous for representing smoothly varying properties. The amount of leakage between adjacent scales, i.e. the extent to which a particular dilation of the wavelet responds to spatial frequencies outside the nominal scale range, is also reduced. The disadvantage of increasing M is that the support of the wavelet is increased, and so the degree of localization of any wavelet coefficient is reduced. A wavelet with M vanishing moments has $2M$ coefficients in the dilated wavelet function $\psi_{m,n}$. In this example, we chose $M = 3$ as a compromise between a more compact basis (with smaller M) and reduced leakage between scales (with larger M). Figure 15.4 shows the multiresolution analysis of the data on emission rates using this wavelet basis.

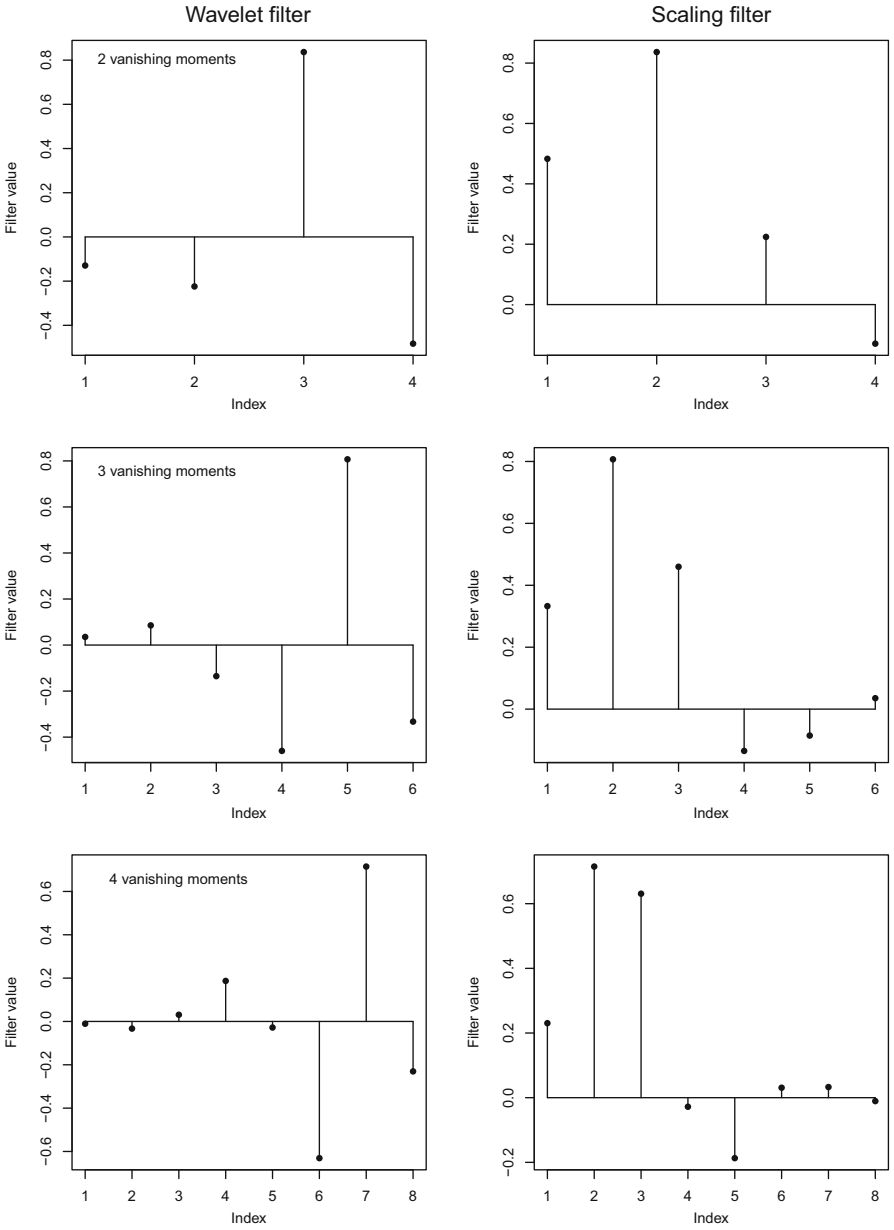


Fig. 15.3 Wavelet and scaling functions for Daubechies's extremal phase wavelet with two, three and four vanishing moments

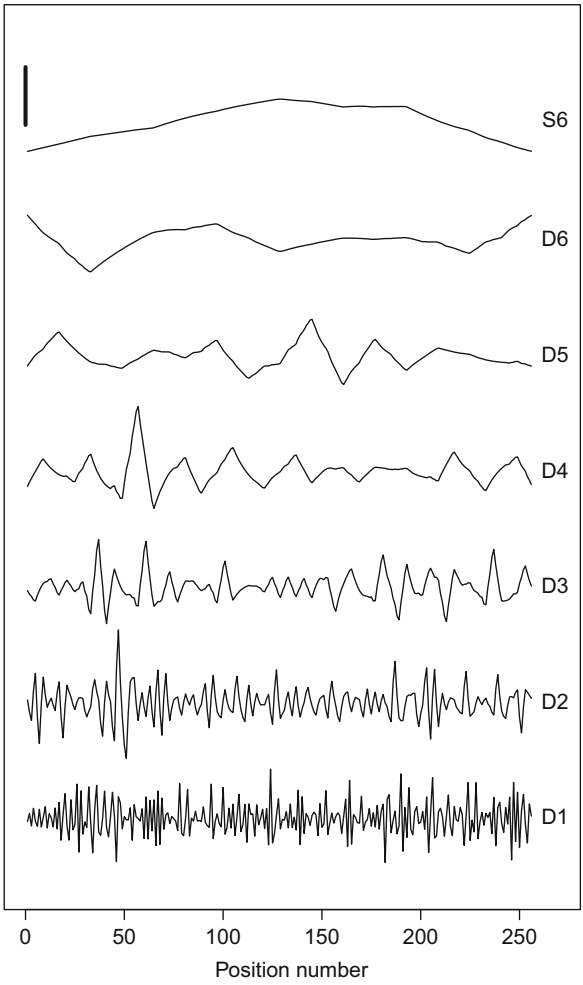


Fig. 15.4 Multiresolution analysis of the data on (transformed) emission rate using six successive dilations of Daubechies's extremal phase wavelet with three vanishing moments. The six detail components are labelled *D1–D6*; the smooth component is labelled *S6*. The components of the analysis are stacked above each other; the *vertical bar* corresponds to 1 unit on the scale

15.3 Wavelets for Inference, the Maximal Overlap Discrete Wavelet Transform

15.3.1 Wavelet Variance

In the introduction, we noted that the Fourier coefficients for a particular frequency can be used to compute the power of a particular data set at that frequency, which is a component of the variance of the data. It is possible to compute a similar component of variance from the DWT coefficients. From the unit square norm of

the mother wavelet, Eq. (15.3), and certain other conditions (Daubechies 1992), the DWT partitions the variance of set of data over all scales. The sample wavelet variance (Percival 1995) for scale $2^j x_0$ can be computed as

$$\widehat{\sigma}_j^2 = \frac{1}{2^j n_j} \sum_{n=1}^{n_j} D_{j,n}^2, \tag{15.12}$$

where n_j is the number of DWT coefficients at the jth scale.

The wavelet coefficients $D_{j,n}$ for the jth scale are obtained by translations of the dilated wavelet in steps of 2^j sample intervals, which is equivalent to subsampling, at interval 2^j, the convolution of the data sequence with the dilated mother wavelet. A consequence of the subsampling is that all wavelets in the basis are orthogonal to all others, so the sample wavelet variance given above has n_j degrees of freedom. However, Percival and Guttorp (1994) showed that the wavelet variance could be more efficiently estimated by retaining all output from the filtering of the signal with the dilated wavelet. This is called the maximal overlap discrete wavelet transform (MODWT). The MODWT is equivalent to the DWT but with wavelet functions for the jth scale translated in unit steps rather than steps of 2^j. The kth MODWT coefficient out of $\breve{n}_j$ for the jth scale is denoted $d_{j,k}$, and the MODWT wavelet variance is computed as

$$\breve{\sigma}_j^2 = \frac{1}{2^j \breve{n}_j} \sum_{k=1}^{\breve{n}_j} d_{j,k}^2. \tag{15.13}$$

These coefficients are not orthogonal, so do not provide $\breve{n}_j$ independent pieces of information about the variance at this scale. Percival and Walden (2000) propose some methods to compute effective degrees of freedom for $\breve{\sigma}_j^2$. The simplest, and most conservative, is to use n_j.

Figure 15.5 shows estimated wavelet variances for the six dilations of Daubechies's extremal phase wavelet (three vanishing moments) applied to the data on emission rates in the MODWT. Also shown are the 95% confidence intervals for these estimates, treating n_j as the effective degrees of freedom.

The wavelet variance, like power, is a global property of a variable. Its interpretation therefore depends on stationarity assumptions, i.e. that the contribution to variance in the interval of a particular DWT scale is homogeneous across the transect. However, the particular value of wavelet transforms is that their basis functions have a local support. One can use the wavelet coefficients for some scale to examine the evidence that variance at this scale is not spatially homogeneous. This was developed by Whitcher et al. (2000) and has been used to examine spatial variation of soils (Lark and Webster 2001; Milne et al. 2011).

Define the normalized cumulative sum of squared MODWT coefficients for scale j at the kth coefficient by

$$S_k = \frac{\sum_{m=1}^{k} d_{j,m}}{\sum_{m=1}^{\breve{n}_j} d_{j,m}}, \; k = 1, 2, \ldots, n_j - 1. \tag{15.14}$$

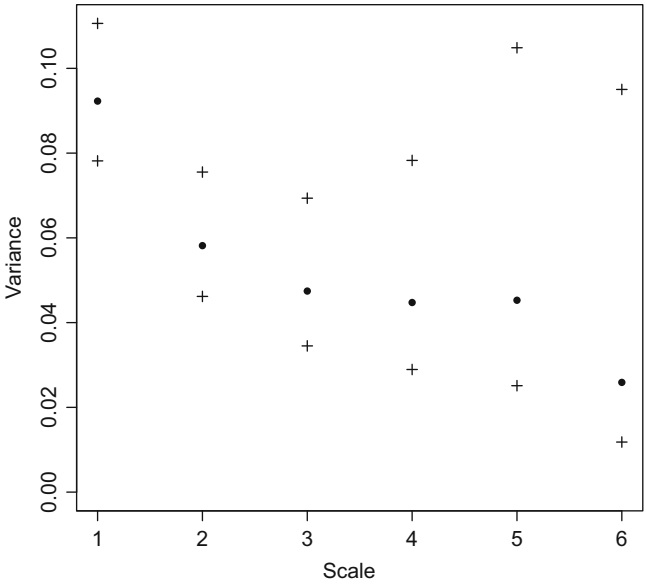


Fig. 15.5 Wavelet variances of (transformed) emission rate using six successive dilations of Daubechies's extremal phase wavelet with three vanishing moments and the MODWT. The *crosses* show the 95% confidence interval for each estimate

Under null hypothesis of homogeneity of the variance at the *j*th scale, we expect S_k to increase linearly with k. Consider an alternative where the variance changes at location k'. Now we expect the graph of S_k against k be approximately a bilinear piecewise function with the breakpoint at k'. The null hypothesis at some scale is tested by finding a candidate change point at which the graph of *Sk* departs from the expected linear function under the null hypothesis. The magnitude of this difference is measured by a statistic B. Define B^+ and B^- as

$$B^+ \equiv \max_{1 \le k \le \breve{n}_j - 1} \left(\frac{k}{\breve{n}_j - 1} - S_k \right) \tag{15.15}$$

and

$$B^- \equiv \max_{1 \le k \le \breve{n}_j - 1} \left(S_k - \frac{k-1}{\breve{n}_j - 1} \right) ; \tag{15.16}$$

then

$$B \equiv \max \left[B^+, B^- \right] . \tag{15.17}$$

The evidence against a null hypothesis of homogeneity of variance is measured by B. Percival and Walden (2000) propose that the null hypothesis for some j and

$\check{n}_j$ is tested by Monte Carlo simulation when $\check{n}_j < 128$ and otherwise by applying a large sample result from Inclán and Tiao (1994) to compute the P-value. Whitcher et al. (2000) propose that one tests the first candidate change point on a transect. If the null hypothesis is rejected, then the change point is accepted, and the same procedure is then used to test the homogeneity of variance of the two segments of the original sequence, divided at the accepted change point.

Figure 15.6 shows an example of change detection in the case of the wavelet variance of the transformed data on emission rates. In this case, we consider the second dilation of Daubechies's extremal phase wavelet with three vanishing moments. This dilation therefore corresponds to a nominal scale interval from 120 to 240 m on the ground. The plot of S_k against k is seen in the top left of the figure, with the 1:1 line superimposed. The biggest deviation, the candidate change point, is at the location where the solid symbol appears on the 1:1 line (position 71 on the transect), and the distance between the two graphs, shown by the vertical line, is the B statistic (0.196 in this case). The graph on the top right shows the distribution of the B statistic for this case obtained for the stationary null hypothesis by Monte Carlo sampling. The empirical p-value is 0.012, which provides a basis to reject the null hypothesis. The wavelet variance appears to decrease at this change point, which, in fact, coincides exactly with the location at which mixed drift over the Lower Greensand, a Cretaceous sandstone, gives way to, predominantly, heavier drift over the Gault clay (see Milne et al. 2011). The figure in the lower part of the plot shows the corresponding detail component in the multiresolution analysis based on the MODWT, with the change point shown by a solid symbol.

15.3.2 Wavelet Covariance and Correlation

Just as the variance of some variable, $u = f(x)$, can be partitioned into components associated with each scale of the DWT, one may compute the contribution to the covariance of variables $u = f(x)$ and $v = f_2(x)$ from variation at some scale j. This is called the wavelet covariance and can be computed from the MODWT coefficients $d^u_{j,n}$ and $d^u_{j,n}$ for $n = 1, 2, \ldots \check{n}_j$ by

$$\tilde{C}_{u,v,j} = \frac{1}{2^j \tilde{n}_j} \sum_{n=1}^{\tilde{n}_j} d^u_{j,n} d^v_{j,n}, \tag{15.18}$$

where $d^u_{j,n}$ and $d^v_{j,n}$ are the MODWT coefficients for scale 2^j and location n for variables u and v, respectively.

Just as the covariance of two variables can be most easily interpreted by scaling it to the interval $[-1, 1]$, also called the correlation, Whitcher et al. (2000) defined the wavelet correlation by rescaling the wavelet covariance by the corresponding wavelet variances at some scale 2^j. The wavelet correlation of u and v at scale 2^j is estimated by

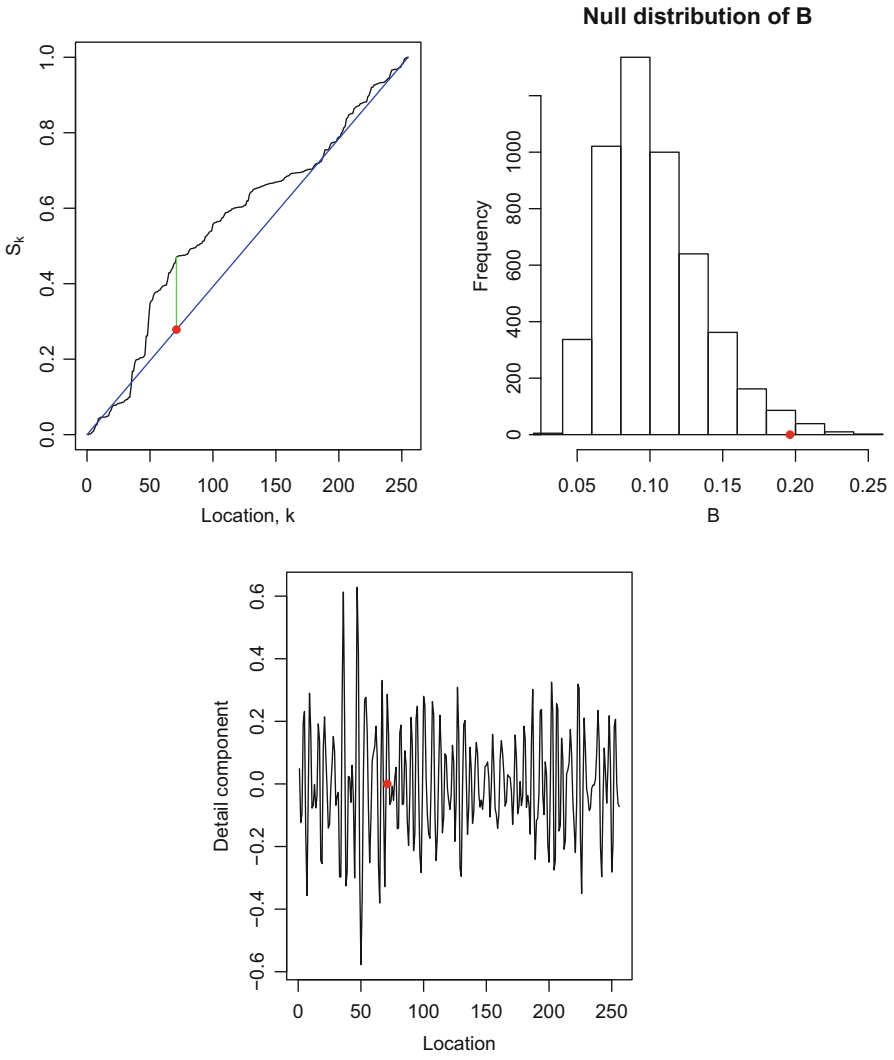


Fig. 15.6 Change detection for the variance of (transformed) emission rate on the second dilations of Daubechies's extremal phase wavelet with three vanishing moments. The three plots are explained in detail in the text

$$\tilde{\rho}_j = \frac{\tilde{C}_{u,v,j}}{\tilde{\sigma}_{u,j}\tilde{\sigma}_{v,j}} \,. \tag{15.19}$$

The wavelet correlation can be useful for identifying scale-specific relationships between variables. Variables may be correlated differently at different scales, and the standard correlation between variables, in which the spatial scales are not distinguished, may obscure such a relationship. For example, Lark et al. (2004)

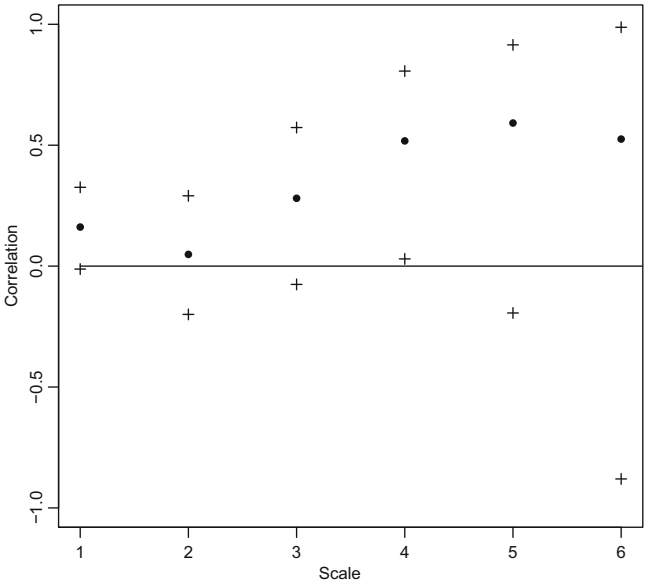


Fig. 15.7 Wavelet correlations of (transformed) emission rate and SOC using six successive dilations of Daubechies's extremal phase wavelet with three vanishing moments and the MODWT. The *crosses* show the 95% confidence interval for each estimate

found strong negative correlations between the rate of nitrous oxide emission from soil cores and soil pH at the coarsest scales on a transect across several fields on a farm. In contrast, soil organic carbon was only correlated with emission rates at the finest scales and soil mineral nitrogen at intermediate scales.

Figure 15.7 shows the wavelet correlation between (transformed) emission rates and SOC content of soil cores on the transect, with the corresponding 95% confidence intervals, on the same MODWT basis as the wavelet variances in Fig. 15.5. Note that the correlations at finer scales are small. The 95% confidence interval excludes zero only for the fourth dilation (nominal scale range of 240–480 m) at which the wavelet correlation is 0.5. The correlation is larger at the fifth dilation, but the confidence interval is wider and includes zero. This indicates that the variation of these two variables at finer scales does not appear to be correlated (at least overall), but that coarser-scale spatial variations of the variables are indeed correlated.

As with the wavelet variance, it is possible to examine the MODWT coefficients for two variables to examine evidence that their covariance is not spatially homogeneous. One computes the cumulative sum of cross products of the MODWT coefficients for some scale of interest:

$$S_k^{u,v} = \frac{\sum_{m=1}^{k} d_{j,m}^u d_{j,m}^v}{\sum_{m=1}^{\breve{n}_j} d_{j,m}^u d_{j,m}^v}, \; k = 1, 2, \ldots, \breve{n}_j - 1. \tag{15.20}$$

The same B statistic can then be computed to find a candidate change point. Lark et al. (2004) and subsequent studies (Milne et al. 2011) use a Monte Carlo method to identify change points at which the difference in wavelet correlation between the two segments of the transect is large enough to reject the null hypothesis of a homogeneous wavelet correlation. In the particular example examined in this chapter, there is no evidence that the correlation between emission and SOC is not homogeneous at any scale of the wavelet transform.

15.4 The Hilbert-Huang Transform

Classical methods of analysing series of data, such as spectral analysis, are based on the assumptions that the series comes from a system which is (i) stationary in both the mean and the variance (i.e. the expectation of these statistics is constant over the series) and (ii) linear (i.e. if the inputs $x_1(t)$ and $x_2(t)$ to a system result in outputs $y_1(t)$ and $y_2(t)$, then if the system is linear, the input $x = x_1(t) + x_2(t)$ will result in the output $y = y_1(t) + y_2(t)$). The wavelet transform allows information on the location of events to be retained, and so any assumptions of stationarity in the mean and variance are not necessary. Wavelet analysis, however, does assume that the series is linear. Wavelet transforms (similar to Fourier transforms) decompose the series of data onto a predefined set of orthonormal basis functions, which relate to specified intervals of frequency. Any sequence $u(x)$, where $x = 1, 2, \ldots, n$, may be decomposed onto a set of basis functions. When we decompose our data using these basis functions, however, we assume that the underlying system (of which our data is a realization) is linear and so can be interpreted as a sum of additive components. Natural systems are rarely stationary or linear, although often these assumptions can be made. Huang et al. (1998) state that the hallmark of a non-linear system is 'intra-frequency frequency variation', i.e. the frequency content of one or more of the constituent functions varies as a function of location. Huang et al. (1996, 1998) sort to derive a method for analysing series of data that was not restricted by either assumptions of linearity or stationarity; to that end, they developed the Hilbert-Huang transform (HHT).

To represent a non-linear and nonstationary data series using a set of basis functions, it is necessary for the basis to be adaptive. This means that the basis depends on the data and so cannot be defined a priori (Huang and Shen 2005), unlike spectral and wavelet methods which use well-defined functions. The HHT comprises (i) an empirical mode decomposition (EMD) and (ii) a Hilbert spectral analysis. The EMD is used to partition the series of data into a number of components, which are known as intrinsic mode functions (IMFs). Similar to the wavelet MRA, these functions sum together to give the original data series. An IMFs' amplitude and frequency can vary over the series. The Hilbert spectral analysis is used to evaluate the 'instantaneous frequency' of the IMFs; in other words, we can calculate a local measure of frequency for each mode function.

The HHT is, in some ways, more intuitive than other methods of analysis that rely on basis functions that are derived a priori and do not obviously related to the data they aim to describe. The HHT has only been tested empirically, however, and mathematical validation still remains an open problem (see chapter 1 of Huang and Shen 2005). Since its development, the HHT has been used in many application areas (see Huang and Shen 2005, and references therein). Most notably Biswas and Si (2011) introduced this method of analysis to soil science. In what follows, we describe some of the developments made on the HHT and apply them to our case study. All the analysis was done using programs based on the MATLAB code provided by Dr Wu at http://rcada.ncu.edu.tw/research1_clip_program.htm.

15.4.1 Empirical Mode Decomposition

Empirical mode decomposition is used to derive the basis functions for a given series of data. As we discussed above, these basis functions (known as IMFs) are calculated from the data. The derivation of the IMFs is based on the assumptions that (i) each IMF oscillates about zero and the number of extrema (n_E) may differ from the number of times it crosses zero (n_Z) by one at most, i.e.

$$|n_Z - n_E| \leq 1 \tag{15.21}$$

and (ii) the IMFs' oscillations are symmetric about zero.

The steps taken to derive the IMFs for a series of data are as follows. First we must identify all of the local minima and maxima in the series. An upper envelope, $E_u(x)$, is derived by using a cubic spline to join all of the local maxima. Similarly a lower envelope, $E_l(x)$, is derived from the local minima (see Fig. 15.8). The local mean of the two envelopes is calculated by $m(x) = (E_l(x) + E_u(x))/2$ and is subtracted from the data series, $u(x)$, to give the first candidate IMF $h_1(x) = u(x) - m_1(x)$. This process is known as *sifting* and it both eliminates riding waves and makes the wave profile more symmetric (Huang and Shen 2005). If this candidate IMF, $h_1(x)$, satisfies the stoppage criteria (defined below), then it becomes the first IMF $I_1(x)$. If not, then the processes start again, but this time, $h_1(x)$ is treated as the data and the candidate IMF is given by $h_2(x) = h_1(x) - m_2(x)$, where $m_2(x)$ is the mean of the upper and lower envelope of $h_1(x)$. The process is repeated until the candidate IMF, $h_k(x)$, satisfies the stoppage criteria. Then the first IMF, $I_1(x)$ is set equal to $h_k(x)$.

Locally, the first IMF, $I_1(x)$, contains the highest frequency information. The second IMF, $I_2(x)$, is calculated in the same way as the first, but this time, the residual $r_1(x) = u(x) - I_1(x)$ is used in place of the original data series. The process continues until the Jth level residual, defined

$$r_j(x) = u(x) - \sum_{i=j}^{J} I_j(x), \tag{15.22}$$

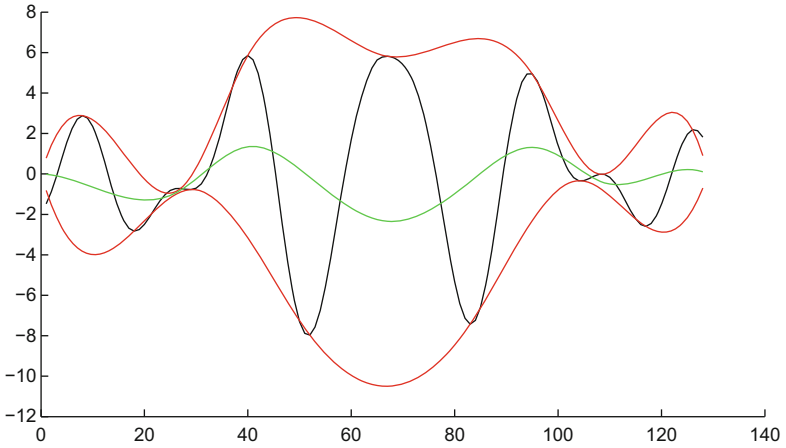


Fig. 15.8 An illustration of the calculation of the mean envelope of a series, which is part of the *sifting* process. The series, $u(x)$, is shown by the *black line*. The *upper envelope* $E_u(x)$ is defined by using a cubic spline to join the local maxima; similarly the *lower envelope* $E_l(x)$ is defined using the local minima. The envelop is shown in *red*. The mean envelop $m(x) = (E_l(x) + E_u(x))/2$ is calculated (shown in *green*). Later the mean envelope is subtracted from the data series ($u(x)$) to give the first candidate IMF $h_1(x) = u(x) - m_1(x)$

is either a monotonic function or a function with only one extremum. The final residual is comparable to the smooth component of the wavelet MRA, which captures the trend in the data. Similar to the MRA, the data series $u(x)$ is equal to the sum of the IMFs plus the final residual,

$$u(x) = \sum_{j=1}^{J} I_j(x) + r_J(x). \tag{15.23}$$

At least two different stoppage criteria have been suggested (Huang and Shen 2005). The first, suggested by Huang et al. (1998), uses a Cauchy-type convergence test. It requires that

$$\frac{\sum_{x=1}^{n} |h_{k-1}(x) - h_k(x)|^2}{\sum_{x=1}^{n} \{h_{k-1}(x)\}^2} < \varepsilon, \tag{15.24}$$

where n is the number of points in the series and ε is a predetermined value. This criterion has two shortfalls: firstly it does not guarantee that Eq. (15.21) is satisfied, and secondly, the value of ε must be defined. The description of the method given by Biswas and Si (2011) overcomes the former problem by making the condition $|n_E - n_Z| \leq 1$ an additional stopping criterion, and they also suggest that ε lay between 0.2 and 0.3.

Huang et al. (1999, 2003) offer an alternative criterion, that is, that $|n_E - n_Z| \leq 1$ and that the values of n_E and n_Z have stayed the same S consecutive times. The choice of S is still arbitrary, although Huang et al. (2003) do give some guidance and established that for optimal sifting, S should be set between 4 and 8.

Both of the stoppage criteria described above require a parameter to be arbitrarily set. In the first method, we described the value of ε had to be set and in the second the value of S. The choice of the value for this stoppage parameter can affect the decomposition (i.e. different sets of IMFs might be produced by varying the value of the stoppage parameter). It is therefore important to gauge the sensitivity of our IMFs to the choice of stoppage parameter.

Huang et al. (2003) proposed a method for calculating confidence intervals for the IMFs based on exploring the sensitivity of the IMFs to changes in the stoppage parameter. Decompositions are calculated for a range of values of the stoppage parameter. Each decomposition produces a set of IMFs, $I_{k,j}(x)$ where $j = 1, 2, \ldots, J$ and k denotes that the decomposition was done with stoppage parameter $S(k)$ (or $\varepsilon(k)$). From these, the mean IMF can be calculated

$$I_{\hat{\mu},j}(x) = \frac{1}{K} \sum_{k=1}^{K} I_{k,j}(x), \tag{15.25}$$

and standard deviation

$$I_{\hat{\sigma},j}(x) = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} \left\{ I_{k,j}(x) - I_{\hat{\mu},j}(x) \right\}^2}, \tag{15.26}$$

from which confidence intervals can be deduced using standard statistical formulae.

Figure 15.9 shows the mean IMFs with their 95% confidence intervals for the log CO_2 data using EMD. The EMD method we used had a stoppage criterion defined by Eq. (15.24) and the additional condition given by Eq. (15.21). We allowed ϵ to take values between 0.01 and 1. There is a strong similarity between the residual component and the smooth component of the MRA. This is also true of the three coarser-scale MRA components and IMF 35. In IMF 2, we can see inter-frequency modulation most clearly. For example, around position 200, component oscillates with lower frequency than it does from around position 220 to the end of the series.

15.4.2 Analysis of Instantaneous Frequencies

15.4.2.1 Hilbert Spectral Analysis

Theoretically, the instantaneous frequency $\omega(x)$ of each IMF can be calculated using Hilbert spectral analysis. Firstly the complex conjugate $\tilde{I}_j(x)$ of the IMF $I_j(x)$ is calculated by

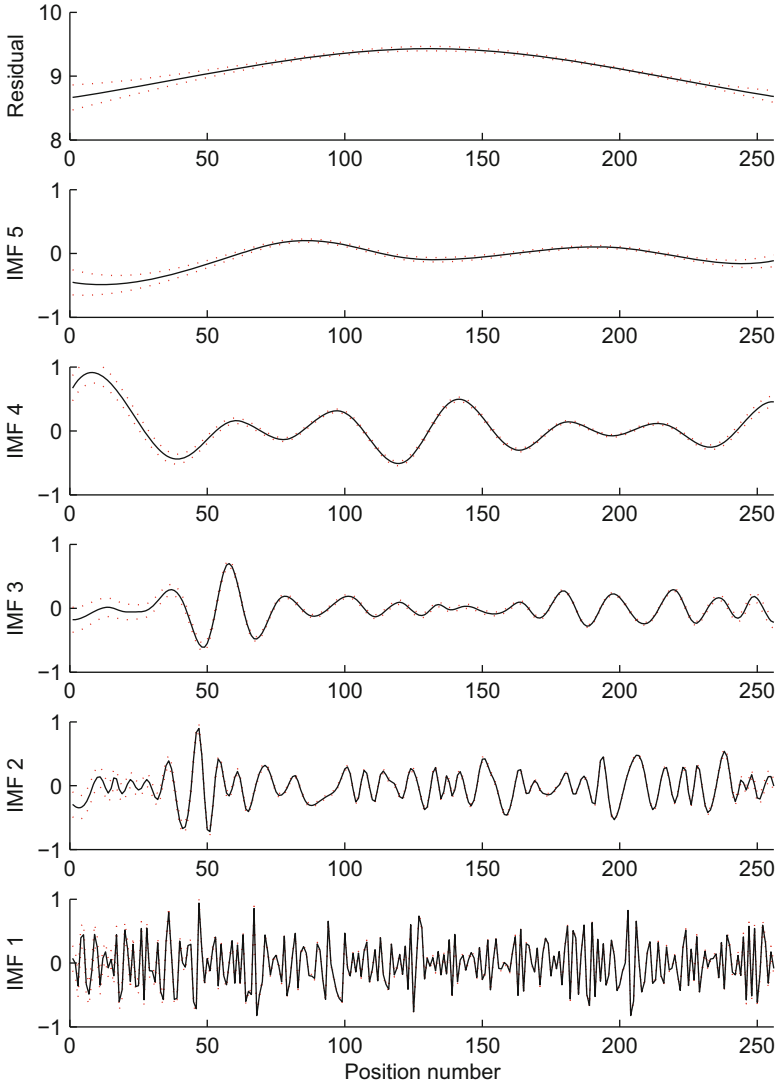


Fig. 15.9 The IMFs for log CO_2. The IMFs were calculated using EMD with a stoppage criterion defined by Eq. (15.24) and the additional condition given by Eq. (15.21). The IMFs were calculated 100 times with the value of ϵ set to $0.01, 0.02, \ldots, 1.0$. The mean and variance for each IMF were estimated. The mean is shown in *black* and the 95% confidence interval, which was calculated using the variance, is shown in *red*

$$\tilde{I}_j(x) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{I_j(\tau)}{x - \tau} \mathrm{d}\tau, \tag{15.27}$$

where P is the principal value of the singular integral (also known as the Cauchy principal value). The analytical series is then defined:

$$z_j(x) = I_j(x) + i\tilde{I}_j(x), \tag{15.28}$$

which can be written in polar coordinates as

$$z_j(x) = a_j(x) \exp\{i\theta_j(x)\}, \tag{15.29}$$

where $i = \sqrt{-1}$, $a_j(x) = \sqrt{\{I_j(x)\}^2 + \{\tilde{I}_j(x)\}^2}$ is the instantaneous amplitude and $\theta_j = \arctan\left(\tilde{I}_j(x)/I_j(x)\right)$. The instantaneous frequency $\omega_j(x)$ is given by

$$\omega_j(x) = \frac{\mathrm{d}\theta_j(x)}{\mathrm{d}x}. \tag{15.30}$$

The original series can be expressed in terms of the polar coordinates by

$$u(x) = \mathrm{Re}\left[\sum_{j=1}^{J} a_j(x) \exp\left\{i \int \omega_j(x)\right\}\right] + r_J(x), \tag{15.31}$$

and the similarity to the Fourier expansion of this data is clear:

$$u(x) = \mathrm{Re}\left[\sum_{j=1}^{J} a_j \exp\{i\omega_j\}\right] + a_0. \tag{15.32}$$

The IMF representation is in fact a generalization of the Fourier expansion with amplitude and frequency given as functions of location.

15.4.2.2 The Normalized Hilbert Transform

The Hilbert transform, given above, exists for L^p function space (Bourbaki 1987); however, the phase function $\theta_j(x)$ does not always give a physically meaningful instantaneous frequency (see Huang and Shen 2005). For this to happen, the IMF function $I_j(x) = a_j(x)\cos\{\theta_j(x)\}$ must satisfy

$$\mathrm{H}\left[a_j(x)\cos\{\theta_j(x)\}\right] = a_j(x)\mathrm{H}\left[\cos\{\theta_j(x)\}\right], \tag{15.33}$$

and for this to hold, the Bedrosian theorem states that the Fourier spectra of $a_j(x)$ and $\cos\{\theta_j(x)\}$ must be totally disjoint in the frequency space and that the frequency range of $\cos\{\theta_j(x)\}$ must be higher than $a_j(x)$. To avoid this problem, the frequency modulation and amplitude modulation parts of the IMF can be separated by a process of normalization (Huang and Wu 2008). The Hilbert transform is then applied to the frequency modulation part, and the problem stated in the Bedrosian theorem no longer applies. This is known as the normalized Hilbert transform.

The normalization of an IMF, $I_j(x)$, is done empirically. Firstly the absolute values of the IMF are calculated and then the local maxima are identified. A spline $e_1(x)$ is fitted through the values, which is unique for a given order of spline. The first approximation to normalization is given by $\psi_1(x) = I_j(x)/e_1(x)$. If $e_1(x)$ is identical to $a_j(x)$, then $\psi_1(x)$ should equal $\cos\{\theta_j(x)\}$ and so $|\psi_1(x)| \leq 1$. This is often not the case and so the process of normalization is done again, this time on $\psi_1(x)$ instead of $I_j(x)$. This iterative processes stops at the mth iteration when $|\psi_m(x)| \leq 1$. The frequency modulation part of the IMF and the amplitude modulation part of the IMF are thus empirically defined by $\psi_m(x)$ and $I_j(x)/\psi_m(x)$, respectively.

The instantaneous frequencies for IMFs 1 to 5 of the log CO_2 data are shown in Fig. 15.10. At the start of the transect, the variation is greatest at wavelengths of 1500 m and more. Another notable feature is that at wavelengths less than 750 m there is an increase in variation around position number 50. This coincides with the significant changes in variation quantified by the wavelet analysis.

15.4.3 The Direct Quadrature Method

The estimates of instantaneous frequency made using the normalized IMFs improve greatly on those calculated using IMFs that have not been normalized. However, the true frequency of fluctuations might still be underestimated. More recently, Huang et al. (2009) have proposed abandoning the Hilbert transform and instead proposed using their *direct quadrature* method. This method simply computes the phase function by using the arc cosine of the frequency modulation part of the IMF (Huang et al. 2009).

15.4.4 Intrinsic Mode Function Variance

The spectral density function, used in Fourier analysis, decomposes the variance of a series across frequencies, and similarly Percival and Walden (2000) show us that the wavelet variance decomposes the variance of a series with respect to scale. No similar quantitative breakdown of variance exists for the HHT. The EMD decomposes the series into a sequence of IMFs, but the sum of the variance of these

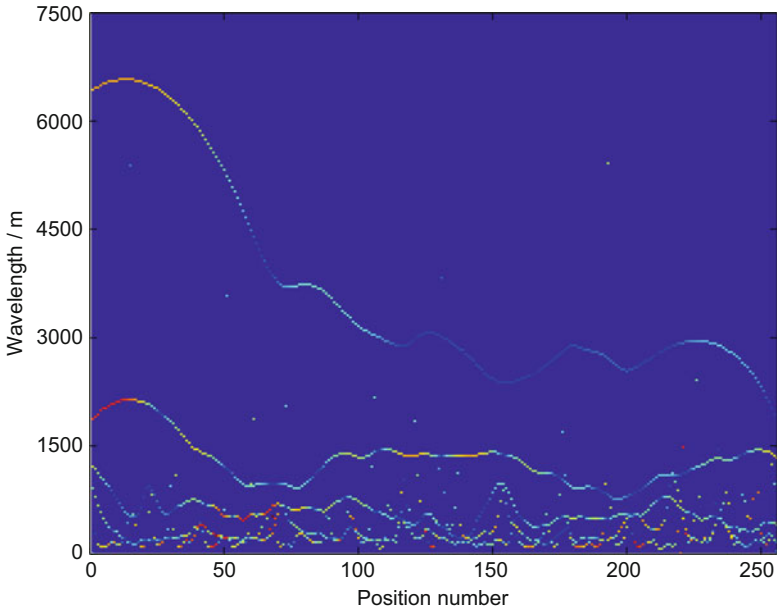


Fig. 15.10 The magnitude of the variations plotted against position number and instantaneous frequency for the IMFs of log CO_2. The IMFs were calculated using empirical mode decomposition

Table 15.1 The percentage of the sum of the variances of each IMF

Component	Percentage of total variance
IMF 1	35.83
IMF 2	17.07
IMF 3	9.56
IMF 4	29.55
IMF 5	0.08

individual components does not in general equal the variance of the original signal. Biswas and Si (2011) investigated the partition of the variance between the IMFs by calculating the variance of each IMF and reporting these as a proportion of the sum of the IMF variances. Table 15.1 shows the contribution to the variance of each of the IMFs for log CO_2. The partition of variation is broadly similar to that seen in the wavelet results. The largest contribution is related to the highest frequency component (IMF 1) and the smallest at the lowest (IMF 5).

15.4.5 Intrinsic Mode Function Correlations and Ensemble Empirical Mode Decomposition

Biswas and Si (2011) considered the correlation between individual IMFs of soil water storage and several variables that were thought to influence soil water storage. The driving variables were not decomposed into IMFs themselves. Therefore, they estimated correlations between a component associated with a given scale interval of one variable and a second variable which contained information on a much larger interval of scale. In their example, they found that there was a strong negative correlation between the IMF associated with wavelengths of 80 to 100 m and elevation. The correlation between the raw data had been weak.

Huang and Wu (2008) considered data on the Southern Oscillation Index (SOI) and the cold tongue index (CTI). They decomposed both sets of data into nine IMFs and a residual and estimated the correlations between the respective pairs of IMFs. They found that the correlations between the pairs were smaller than the overall correlation between the two data sets and attributed this to 'mode mixing'. Mode mixing occurs when either a single IMF comprises signals of widely disparate scales or a signal of similar scale resides in two or more IMFs. Ensemble EMD (EEMD) was developed to overcome the scale separation problem. In this method, white noise with finite amplitude is added to the original data and the IMFs are computed. The processes are repeated many times. The average over the sets of IMFs defines the IMFs of the EEMD. The concept is that only the true signal persists and that the white noise is removed in the averaging processes. Huang and Wu (2008) found that using the EEMD to produce the IMFs for the SOI and CTI data greatly improved their results.

The EEMD requires that the number of IMFs that the data is decomposed into must be set a priori; otherwise different numbers of IMF may be generated at each iteration resulting in a set that cannot be averaged. The total number of IMFs should be set to $\log_2 N$ or fewer, where N is the number of data in the series.

We used the EEMD to produce IMFs for log CO_2 and log OC and then estimated the correlation between the pairs for each level. We set the standard deviation of the white noise to be equal to 0.1 of the standard deviation of the measurements of each variable, and we set the number of IMFs to six (as calculated in the EMD above). Figure 15.11 shows the IMFs produced. The black line shows the IMFs for the log CO_2 data and the black for the log OC data. The correlations between the IMF pairs and the raw data are shown in Fig. 15.12. The correlation between the raw data is shown as IMF 0. Similar to the wavelet results, the IMF pairs associated with shorter wavelengths (IMFs 1 and 2) are weakly correlated to one another, whereas the IMF associated with the longest wavelengths (IMF 5) was more strongly correlated than the raw data sets were.

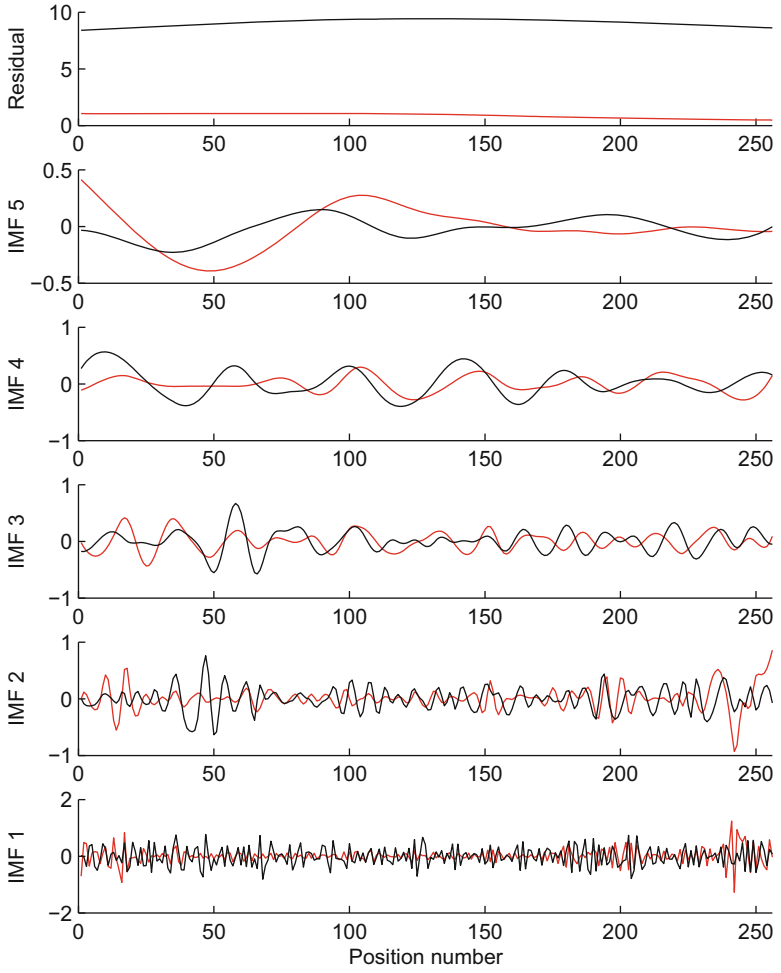


Fig. 15.11 The IMFs of log CO_2 (*black*) and log OC (*red*) calculated using ensemble empirical mode decomposition (EEMD). The number of IMFs was fixed to 5

15.4.6 Comparison of Methods

Both the Hilbert-Huang transform and the wavelet transform offer power methods for analysing data. The wavelet transform has an underlying mathematical theory, whereas the Hilbert-Huang transform is empirical and lacks any underpinning mathematical basis. The advantage of the Hilbert-Huang transform over wavelets is that it is better suited to data series from non-linear processes. Huang and Wu (2008) illustrate this using Duffing's equation

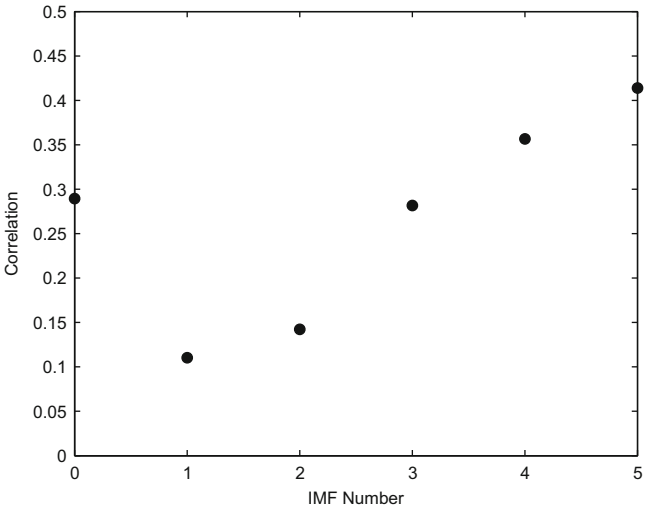


Fig. 15.12 The correlations between IMFs for log CO_2 and log OC. The correlation between the original measurements is plotted against IMF 0

$$\frac{\mathrm{d}x}{\mathrm{d}t^2} + x + \epsilon x^3 = \gamma \cos \alpha t \tag{15.34}$$

where ϵ, γ and α are model parameters. The cubic term causes the non-linearity and so the frequency of x changes within a period. This is clearly represented in the Hilbert spectrum that results from a Hilbert-Huang analysis, whereas other methods that assume the system is linear capture this feature less clearly in the harmonics (Huang et al. 1998).

The examples presented in this chapter, however, show similar results from both sets of methods. The patterns in the IMFs are similar to the MRA, the partitioning of the variance of the IMFs is similar to that from the wavelet variance, and the wavelet correlations are broadly similar to the correlations calculated with the EEMDs.

References

Biswas A, Si BC (2011) Revealing the controls of soil water storage at different scales in a hummocky landscape. Soil Phys 75:1295–1306

Bourbaki N (1987) Topological vector spaces, elements of mathematics. Springer, Berlin

Daubechies I (1988) Orthonormal bases of compactly supported wavelets. Commun Pure Appl Math 41:909–996

Daubechies I (1992) Ten lectures on wavelets. Society for Industrial and Applied Mathematics (SIAM), Philadelphia

Goovaerts P, Webster R (1994) Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur J Soil Sci 45:79–95

Haskard KA, Welham SJ, Lark RM (2010) Spectral tempering to model non-stationary covariance of nitrous oxide emissions from soil using continuous or categorical explanatory variables at a landscape scale. Geoderma 159:358–370

Huang NE, Shen SS (2005) Hilbert-Huang transform and its application. World Scientific Publishing Co, Pte. Ltd., Singapore

Huang NE, Wu Z (2008) A review on Hilbert-Huang transform: method and its applications to geophysical studies. Rev Geophys 46:1–23

Huang NE, Long SR, Shen Z (1996) The mechanism for frequency downshift in nonlinear wave evolution. Adv Appl Mech 32:59–111

Huang NE, Shen Z, Long SR (1999) A new view of nonlinear water waves-the Hilbert spectrum. Annu Rev Fluid Mech 31:417–457

Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc R Soc Lond Ser A 454:903–993

Huang NE, Wu Z, Long SR, Arnold KC, Blank K, Liu TW (2009) On instantaneous frequency. Adv Adapt Data Anal 1(2):177–229

Huang NE, Wu M-LC, Long SR, Shen SSP, Qu W, Gloersen P, Fan KL (2003) A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proc R Soc Lond A 459:3217–2345

Inclán C, Tiao GC (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. J Am Stat Assoc 427:913–923

Lark RM, Webster R (1999) Analysis and elucidation of soil variation using wavelets. Eur J Soil Sci 50:185–206

Lark RM, Webster R (2001) Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet transform. Eur J Soil Sci 52:547–562

Lark RM, Milne AE, Addiscott TM, Goulding KWT, Webster CP, O'Flaherty S (2004) Scale- and location-dependent correlation of nitrous oxide emissions with soil properties: an analysis using wavelets. Eur J Soil Sci 55:601–610

Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11:674–693

Milne AE, Webster R, Lark RM (2010) Spectral and wavelet analysis of gilgai patterns from air photography. Aust J Soil Res 48:309–325

Milne AE, Haskard KA, Webster CP, Truan IA, Goulding KWT, Lark RM (2011) Wavelet analysis of the correlations between soil properties and potential nitrous oxide emission at farm and landscape scales. Eur J Soil Sci 62:467–478

Percival DP (1995) On estimation of the wavelet variance. Biometrika 82:619–631

Percival DB, Guttorp P (1994) Long-memory processes, the Allan variance and wavelets. In: Foufoula-Georgiou E, Kumar P (eds) Wavelets in geophysics. Academic, New York, pp 325–344

Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge

Si BC (2003) Scale and location dependent soil hydraulic properties in a hummocky landscape: a wavelet approach. In: Pachepsky Y, Radcliffe D, Magdi Selim H (eds) Scaling methods in soil physics. CRC Press, Boca Raton, pp 169–187

Si BC, Farrell RE (2004) Scale-dependent relationship between wheat yield and topographic indices: a wavelet approach. Soil Sci Soc Am J 68:577–587

Webster R (1977) Spectral analysis of gilgai soil. Aust J Soil Res 15:191–204

Whitcher BJ, Guttorp P, Percival DB (2000) Wavelet analysis of covariance with application to atmospheric time series. J Geophys Res Atmos 105(D11):14941–14962

Youden WJ, Mehlich A (1937) Selection of efficient methods for soil sampling. Contrib Boyce Thompson Inst Plant Res 9:59–70

Chapter 16
Pedodiversity

Mario Fajardo and Alex. B. McBratney

"Como la tierra
eres
necesaria."

"As the earth
You are
necessary."

Pablo Neruda
Oda a la Alegría

16.1 Introduction

Pedodiversity deals with the analysis of the number and complexity of pre-classified soil entities and/or their properties in the landscape (Fridland 1974; Jacuchno 1976; Linkeš et al. 1983; Ibañez et al. 1987, 1995; McBratney 1992; McBratney and Minasny 2007; Ibáñez and Bockheim 2013). If we consider that soil directly and/or indirectly affects every single biotic structure, it is easy to see that the study of its distribution needs to be a priority when trying to preserve the biodiversity of our planet. Whereas the pedologist might regard pedodiversity a great boon, an agronomist might regard the exact same situation a nuisance.

A generally accepted definition of pedodiversity can be inherited from the concepts behind biological and ecological diversity (Magurran 1988, 2013) involving the quantification of (a) the *richness* of individuals in a determined geographical area and/or timeframe, e.g. how many different soil types exist in a region, (b) the *abundance* or the sum of soil entities of each type and (c) the relative distribution or the *evenness* of those soil individuals, i.e. how separate sites are similar in their abundances.

M. Fajardo (✉) • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: mario.fajardo@sydney.edu.au; alex.mcbratney@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_16

Pedodiversity inherits most of the notions and procedures of its counterparts, e.g. biodiversity. It is therefore possible to also successfully apply most of the established methods to studies of pedodiversity. However, and due to the dissimilar nature of the object of study (*soil [taxonomic] object* vs *species*), it is necessary to outline some ideas that may appear conflicting at first.

16.1.1 Continuous and Discrete Classes

The main difference between conventional biological diversity, which relies on the concept of *species*, and pedodiversity is that the soil can be considered as a continuous body, most of the time. Historically, this duality has caused somehow *different schools* of pedodiversity, similar to the different working groups in pedology and pedometrics.

Accepting the fact that the soil can behave as continuous, we understand that it also can be discretised by a rather taxonomical approach (USDA Soil and Conservation Service 1999; Isbell 2002; Jahn et al. 2006) and/or different numerical techniques (Campbell et al. 1970; Triantafilis et al. 2001; Hughes et al. 2014). As other authors have already noticed (McBratney and Minasny 2007; Feoli et al. 2013), it is possible to apply the notions of distance dissimilarity between classes, dealing with the apparent dichotomy of continuous and discrete classes (refer to Chaps. 8 and 9).

In a *crisp or clear* partition equivalent to a conventional taxonomic classification where the distance between soil classes is not numerically assigned, the dissimilarity within a class will be assumed 0 and between classes will be equal to 1. On the other hand, in the case of a numerical classification or taxonomic classes with an assigned distance, e.g. as shown in McBratney and Minasny (2007), the average dissimilarity within and between classes can range between 0 and 1 (Feoli et al. 2013). In this chapter we will not prefer a particular approach, since both have their own benefits and drawbacks.

16.1.2 The Object of Study

We have seen that the soil continuum can be systematically stratified and classified by its attributes. In the same way, if this stratification (*crisp* or *fuzzy*) is spatially performed, we obtain a delimited soil body which is homogeneous enough to be considered as a physical separate entity or *soil object*. As outlined in Chap. 8, a generally accepted notion of this elementary soil unit is the "point representation", i.e. a more flexible view of the concept of *pedon* (also see NRCS 1993; Holmgren 1988).

The characterization of a *point representation* is based on the information obtained from soil profiles or, in their absence, soil cores in a determined geographical location. Each point representation will be classified based on a (most of the time) vertical sequence of soil horizons and/or soil layers and its observed and/or measured attributes. However, in order to fully (taxonomically) classify a

physical pedological unit, greater detail needs to be considered, i.e. lateral variation of horizons, shape of boundaries, etc.

As noticed by Campbell and Edmonds (1984), as soon as we apply taxonomic criteria into a geographic representation, we will not necessarily obtain a homogeneous or well-defined landscape unit, but most probably a highly discontinuous sum of patches. Butler (1981) called this the taxonomic chop. Many attempts have been made in the past to define this basic structure of soil which has both a clear spatial and taxonomical representation. Some of them are, for example, the elementary soil areal (ESA) defined in Fridland (1965), which is a homogeneous body of soil bordered on all sides by other ESAs. The same author revised some other structures or heterogeneous ESAs, e.g. sporadically spotted and regular-cyclic areal (Fridland 1974; Saldaña 2013). By their definition, homogeneous ESAs greatly resemble the concept of the *polypedon* (Johnson 1963; United States Department of Agriculture Soil and Conservation Service 1975), which is currently the most widely accepted definition of the *soil individual*.

Assuming that ESAs or polypedons correspond to the basic soil units, then it is possible to establish some analogies with biology. For example, Johnson (1963) compared the polypedon to individual pine trees, individual fish or men. The concept of *soil series* can then be regarded as a population of soil profiles with more or less fixed properties, which somehow corresponds to the concept of *species* in biology (Krasil'nikov 2009). However, here, we need to keep in mind that this comparison only refers to the status of soil as an elementary classification unit.

From here, higher taxonomical-geographical units can be identified. Conventional biodiversity studies recognize higher geographical units such as *assemblages*, *guilds* or *communities of species* for describing the diversity of biological ecosystems; however, they cannot be "translated" straightforwardly, and some distinctions have to be considered. Undoubtedly, the spatial (and temporal) subject of a pedodiversity study will depend on the purpose of each investigation; however, it is important – at least in the context of this chapter – to briefly mention some of the commonly used *soil geographical units* of pedodiversity studies.

As previously mentioned, conventional diversity studies consider different units of species aggregation (or ecological sets) defined after shared *resources*, *geography* and *phylogeny* (Magurran 1988, 2013). In the case of pedodiversity, these units will have a close relation with Jenny's well-known soil-forming factors (Chap. 18). The subdiscipline of ecology which studies some of these soil homogeneous zones is that of landscape ecology, defined as a "sub-discipline of ecology that emphasizes the interaction between spatial patterns and ecological processes, that is, the causes and consequences of spatial heterogeneity across a range of scales" (Turner 1989; Saldaña 2013).

Some of the first approaches to define *aggregation units* are, for example, those detailed in Fridland (1974) as *soil combinations*, namely, *complexes*, *catenas*, *mosaics*, *spottinesses*, *variations* and *tachets*. Newer approaches of these units are, for example, the so-called map units defined by the NRCS (1993), namely, *consociations*, *associations*, *complexes* or *undifferentiated groups*.

In the same way, regular or arbitrary units can be considered, e.g. regular areas. The pedodiversity of those will be calculated based on the richness or abundance

of the amount of soil units contained within them (of different hierarchical levels) (Saldaña and Ibáñez 2004).

16.1.3 Digital Soil Mapping and Pedodiversity

With the development of digital soil mapping (DSM) (Chap. 12), the representation of soil types in a map acquires a different perspective. The elemental soil unit can be represented by a grid of *points*, *pixels (2d)* or *voxels (3d)*, which have a spatial and a taxonomic value. Higher hierarchical units (e.g. analogue to associations) are then known as polygons with conventional soil units, i.e. clustered or simply regular areas such as an arbitrary 10 ha area.

An interesting feature of DSM products is the *associated membership* value for each pixel, for example, in maps where each pixel has a probable soil order denomination (e.g. *fuzzy* classes). No research has explored yet the use of the pixel value of classification memberships in pedodiversity calculations. In this sense, it should be relatively easy to plug the membership values into the pedodiversity calculations – an extension of methodology using taxonomic distance (see Sect. 16.2.3 on distance-dependant indices).

16.1.4 Parametric and Non-parametric Estimators

A commonly used qualifier in diversity indices is the term parametric or non-parametric. If we assume that the distribution of a group of observations – such as the abundance of soil suborders in a specific geographical area – follows a particular function which can be fitted to the observed data, then we can use *parameters* to describe the diversity of that group of observations. Some of the most popular are inter alia Fisher's alpha and Preston's lambda, which will be reviewed in the next sections. On the other hand, non-parametric methods do not necessarily assume any predefined abundance distribution. It is not unusual, however, to sometimes find indices that are classified as both by different authors, such as Shannon's index of diversity which has been classified as parametric (Feoli et al. 2013) and non-parametric (Magurran 2013).

16.1.5 Goals

Having first clarified some key aspects of pedodiversity studies and how they compare with conventional methods of ecology, the goal of this chapter is to present some of the state-of-the-art research related to the characterization of the diversity of soils in space. Consequently, our goal here is to present the most popular methodologies in the area using concrete examples and emphasizing quantitative over qualitative procedures, as quantitative ones have a closer connection to the framework of this book, i.e. pedometrics.

16.2 Taxonomical Pedodiversity

Conventional diversity studies make a clear distinction of at least three types of diversity; these are *alpha*, *beta* and *gamma* diversity (Whittaker 1972). Simply, *alpha* diversity characterizes the local diversity, e.g. different species present at a site; *beta* diversity characterizes how different sites relate to each other in terms of their species composition or relative abundances, and *gamma* diversity characterizes the diversity of a complete region, e.g. an island or a continent.

As mentioned in the introductory section, soil has both a discrete and a continuous nature; hence the concept of *beta* diversity will always be present to some degree (after the creation of rather taxonomical or fuzzy classes). In order to avoid unnecessary confusion, the following sections will not include the concepts of *alpha-beta* diversity, and it will be explicitly mentioned when the distance between soil units is considered and when it is not.

16.2.1 Abundance

Having defined the concept of the *individual soil entity*, we can now characterize it based on its *abundance* in the landscape, i.e. if a particular type of soil is common or rare. Let us look at our first example. Figure 16.1 shows the predicted soil suborders of the Hunter Valley Wine Country Private Irrigation District (HWCPID) in NSW,

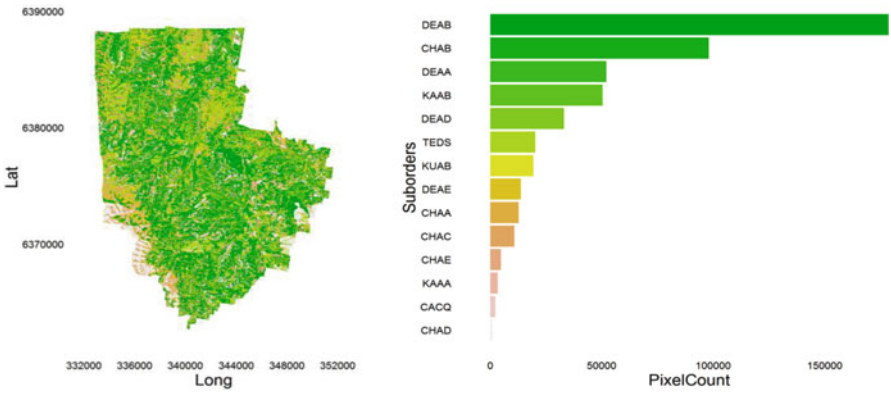


Fig. 16.1 DEAB, Brown Dermosol; CHAB, Brown Chromosol; DEAA, Red Dermosol; KAAB, Brown Kurosol; DEAD, Grey Dremosol; TEDS, Orthic Tenosol; KUAB, Brown Kurosol; DEAE, Black Dermosol; CHAA, Red Chromosol; CHAC, Yellow Chromosol; CHAE, Black Chromosol; KAAA, Red Kandosol; CACQ, Hypercalcic Chromosol; CHAD, Grey Chromosol. These results are a reflection of the underlying soil formation factors. Brown Dermosols are probably the most ubiquitous soil types along the NSW coastline, covering a wide spectrum of soils with a lack of a clear or abrupt textural contrast between the surface and subsurface horizon

Australia, based on the Australian soil classification scheme (Isbell 2002). Here, the most common soil types (*pixels)* are DEAB (Brown Dermosols), followed by CHAB (Brown Chromosols), with CHAD (Grey Chromosols) and CACQ (Hypercalcic Calcarosols) the rarest soil types in the region.

It is common that the relative distribution of different soil types in an area will generally follow a power law where the most abundant taxa are exponentially higher than the rarer ones. The first approach that explained this behaviour in mathematical terms was by Fisher et al. (1943). Fisher et al. (1943) described the much used Fisher's *logarithmic series* or Fisher's *log-series* in their 1943 paper which is still considered one of the key papers of modern ecology.

More specifically, Fisher et al. (1943) observed that in a totally randomized collection of individuals, the number of representatives of each class (soil types in our case), when ordered from the least to the most frequent, will form a *frequency series* with the form

$$n_1, \frac{n_1 x}{2}, \frac{n_1 x^2}{3}, \frac{n_1 x^3}{4}, \tag{16.1}$$

where n_1 is the number of classes with only one individual and x is a number less than 1. Let us apply this concept to the example shown in Fig. 16.1. If we collect a totally random 30,000 pixel sample (without replacement), the number of classes with only 1 individual (n_1) will be 14, and as observed by Fisher, the number of representatives of each class when sorted from the least to the most frequent will decrease exponentially (Fig. 16.2).

Following up on the Fisher's log-series publication, Frank Preston noticed that one of the datasets used did not fulfil the requirements of a random sample (Preston 1948). He, then, proposed a new way of representing abundance data, i.e. as *octaves* or intervals of abundance in an exponential scale (base 2).

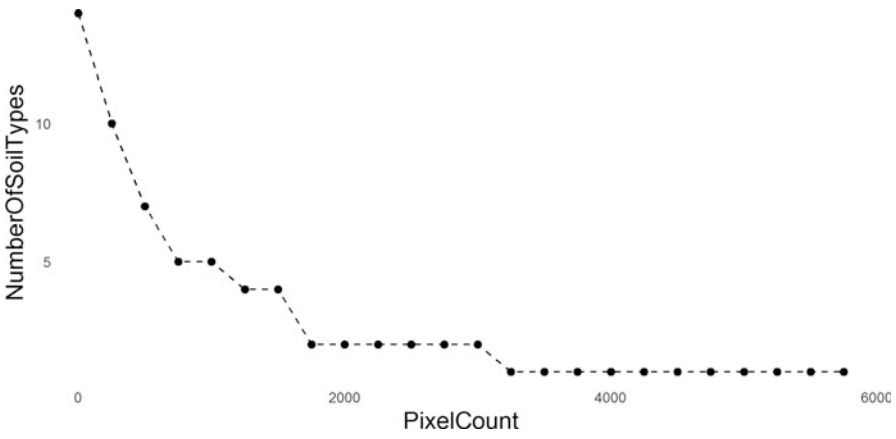


Fig. 16.2 Random sample of 30,000 pixel hollow curve

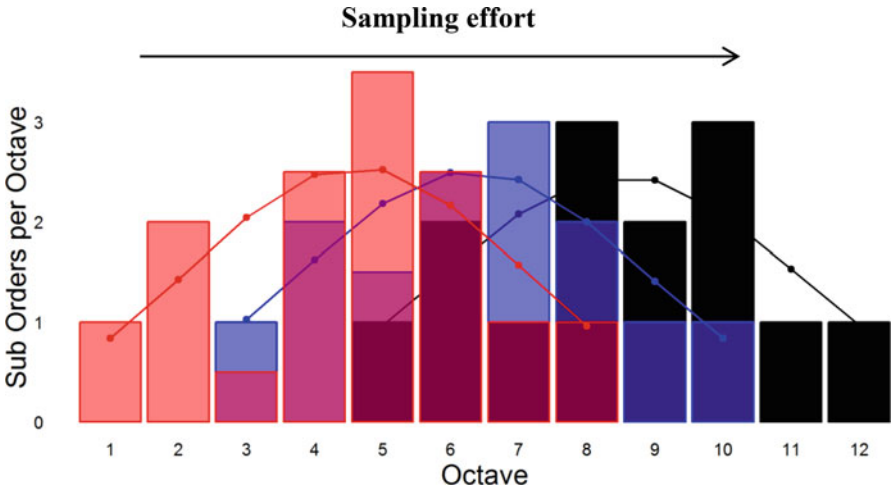


Fig. 16.3 Preston's graph of three random samples of 1000 (*red*), 3000 (*blue*) and 15,000 (*black*) pixels. *Lines* represent fitted values following Preston's approach

Shown in the next example, if we select a random sample of pixels, those soil suborders that have 1–2 pixels will fall in the first octave, those that have between 8 and 16 pixels will fall in the fourth octave ($2^3 = 8$–$2^4 = 16$) and so on (Fig. 16.3). This way of representing abundance data was then named the *log-normal* approach (after modelling the abundance data with a Gaussian curve on a geometric base).

Having observed that most of the communities followed this behaviour, and based on the fact that some very rare species will not be represented in a small sample, Preston proposed the concept of the *veil line*, i.e. the line which divides the unsampled universe with the observed samples.

In this way, and as observed by Preston, the *shape* of the sample distribution will be a good approximation of the universe, and the *veil line* will be uncovered as a more representative sample of the universe.

16.2.1.1 Abundance Models

One of the goals of representing the abundance of a population (e.g. different soil types within a geographical area) is to compare it with another one, as exemplified in Fig. 16.4.

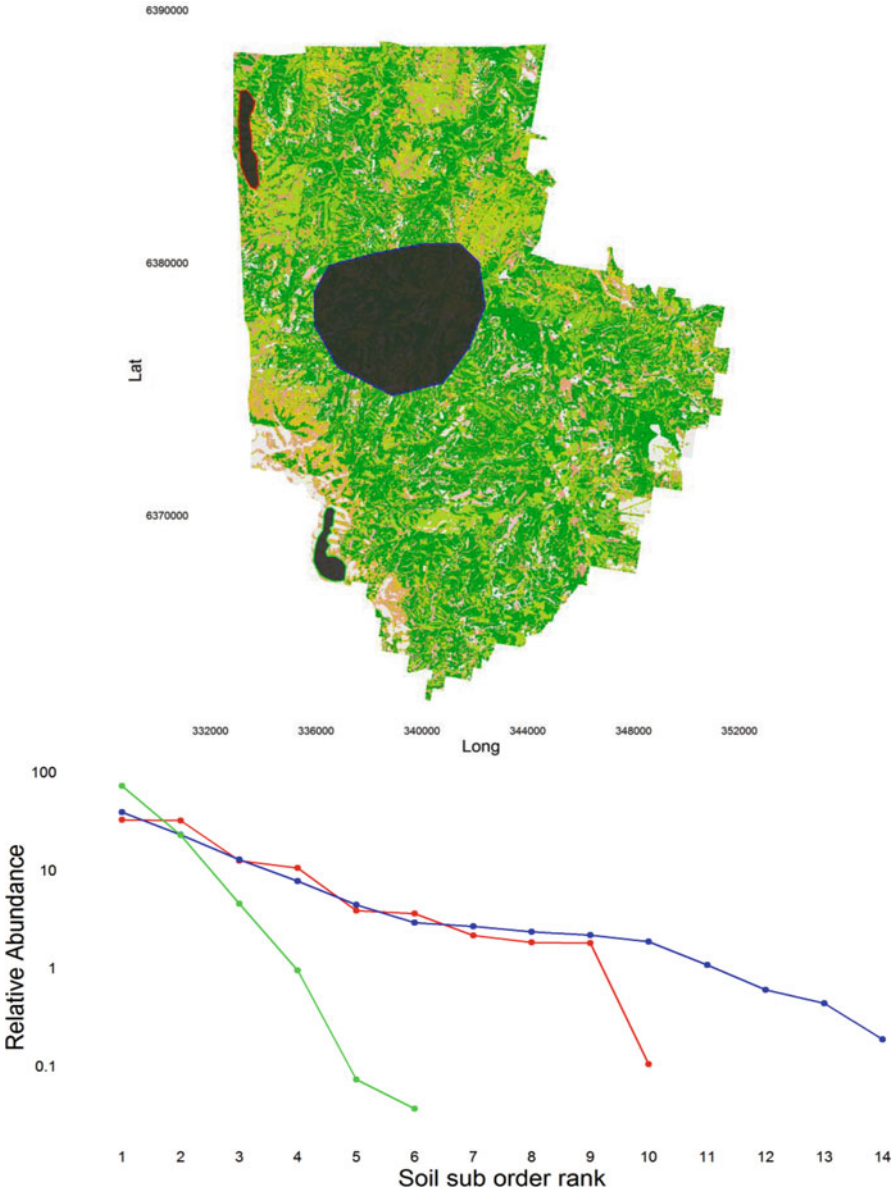


Fig. 16.4 Example of different relative abundance distributions of soil types in the Hunter Valley. The *red*, *green* and *blue lines* represent abundance distributions of soil suborder types that correspond to the areas mapped out in the areal map of the Hunter Valley (*red*, *green* and *blue* edged polygons). Note that the ordinate is presented on a log scale

A revision of the existing models of abundance distribution (or how an ordered group of classes is organized) tells us that there are at least 20 ways (after Magurran 2013) of representing abundance. Some of the most popular are among others the *negative binomial* series (Bliss and Fisher 1953), the *particulate niche* series (MacArthur 1957), the *Zipf-Mandelbrot* series (Zipf 1949) and the *dominance pre-emption* model (Tokeshi 1990).

The main reasons for the continuous update of the existing abundance distribution models are the disputed practical implications of the different statistical approaches (namely, *Fisher's log-series* and *Preston's log-normal*) and the so-called niche-oriented models (*geometric* and *broken stick*).

With respect to the previous, the literature dealing with pedodiversity does not have an apparent conflict with the existing models. In fact, the most popular models used in pedodiversity studies are the oldest ones, namely, the *logarithmic* (Fisher et al. 1943), *log-normal* (Preston 1948), *geometric* (Motomura 1932) and the *broken stick* models (MacArthur 1957) (Fig. 16.5).

The *geometric* model (Motomura 1932) is inspired by the idea that a particular species will *pre-empt* a proportion of the available resources and a second one will pre-empt the same proportion of the remainder and so on. If this behaviour is repeated, then the ordered abundance distribution will follow a geometric series of the following form:

$$n_i = \mathrm{NC}_k k(1-k)^{i-1} \tag{16.2}$$

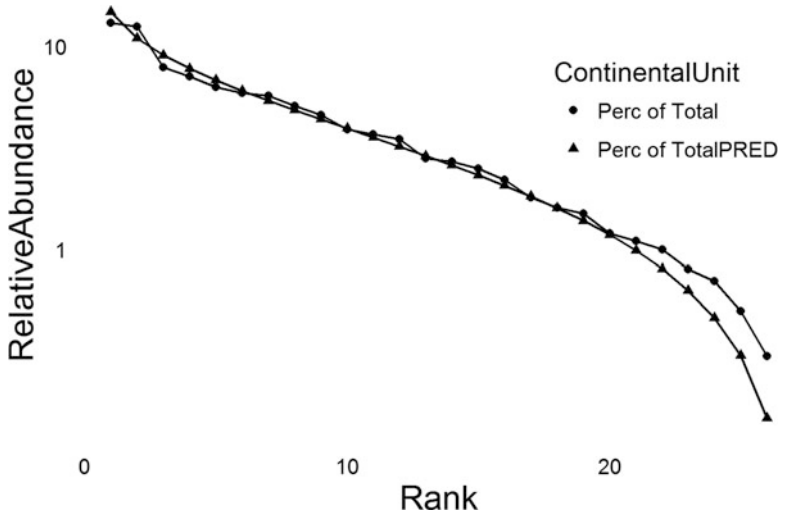


Fig. 16.5 World relative abundance occupied by the major soil groups in accordance with FAO (1988). Perc of total: relative area occupied by the major soil groups, data obtained from Table 1 in Ibañez et al. (1998). Perc of total PRED: MacArthur's broken stick model adjusted using the R-package "sads" (Prado et al. 2016). Note that the ordinate is presented on a log scale

with n_i is the total number of individuals in i^{th} species (class), N the total number of individuals and k the proportion of the remaining space occupied by each successive species (class).

A similar approach underlies the so-called *broken stick* model (McArthur 1957). The idea behind this model is that a specific area can be considered as a stick that once broken is simultaneously distributed into S parts (or classes). Figure 16.5 shows the broken stick model fitted to the data presented in Ibáñez et al. (1998). Ibáñez et al. (1998) showed that the relative abundance of the major soil groups of the world follows this behaviour.

It could be argued that not one single model can be used to describe a community; in fact many models can be adjusted to the same dataset. In this regard, Ibàñez and García-Alvarez (2002) showed that observed data can be successfully fitted to different distributions depending on the software employed, the goodness of fit tests and how the data are grouped. Other authors have highlighted the same (Tokeshi 1990; Magurran 2013).

16.2.2 Richness

The richness of a specific area can be simply defined as the total number of classes of a particular hierarchical level contained in it. As an example, let us use the map of soil types of Denmark, produced by Jacobsen (1984). This map contains a total of 14 different soil types classified following the FAO-Unesco system (global soil map of the world); hence its richness in terms of different soil units is simply 14 (Fig. 16.6). So clearly taxonomic level is an important consideration when calculating richness (especially of small areas).

At first sight, the simplicity of the richness of an area can be considered as an unnecessary indicator, even more so, considering that its value is directly related to a classification scheme (a global or local taxonomical system). As Magurran (2013) clearly explains, the purpose of describing (and attempting to model) the richness of a place is to be able to predict the increase in species richness after additional sampling efforts or the increase of the area sampled, rather than just to estimate the total number of classes.

The study of the richness of a community has been recognized and used by botanists for almost 100 years. In 1921, Olof Arrhenius wrote his classic paper named *Species and Area* (Arrhenius 1921), developing the idea of a previous study (Arrhenius 1920). In his 1921 paper, he states that there is a clear relation between the number of different species and the area sampled:

$$\frac{y}{y_1} = \left(\frac{x}{x_1}\right)^n \tag{16.3}$$

where x is the number of different species found in area y, x_1 those found in y_1, and n a constant varying between 2 and 12.5 (in the Arrhenius dataset), with n depending on the sampled community. From here, Arrhenius (1921) found that when summing

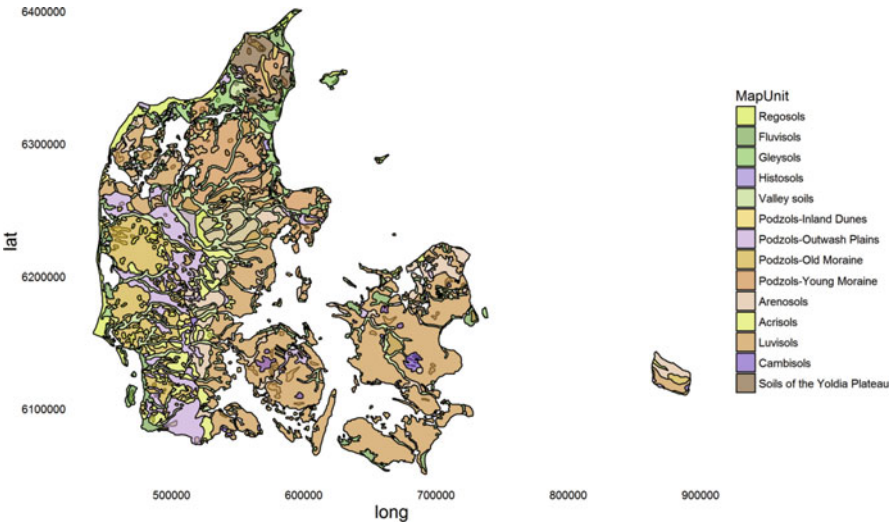


Fig. 16.6 Soil map of Denmark as published in Jacobsen (1984) (Data kindly provided by Anders Moller, Aarhus University)

up all the individual (species) probabilities of occurrence in an area, the probable number of species in an area could be expressed as a geometric series as seen in Eq. 16.4 (this equation was not explicitly written in the original paper but mentioned in words):

$$S = \mathrm{cA}^{z} \tag{16.4}$$

with S the total number of species, A the area sampled and c and z constants.

Two years later, Henry Gleason (1922), disapproving of Arrhenius' first formula, proposed the function that together with the exponential approach (Eq. 16.4) represents one of the most widely accepted approximations of the total number of species in an area:

$$S = d + k \log\ A \tag{16.5}$$

with k and d positive constants.

Even though it is possible to successfully fit any of these functions to a pedodiversity dataset (in fact there are several studies that have successfully fitted power-law series to soil richness as reviewed by Ibáñez and Bockheim 2013), it is interesting to notice that every taxonomical system has a predefined maximum of different classes; hence in a hypothetical maximum area, the richness will necessarily lead to an asymptote.

16.2.2.1 Parametric Models

As revised by Colwell and Coddington (1994), the main difference between parametric and non-parametric methods for estimating richness is the type of information required. In the case of parametric methods, a previous fitting to a known abundance distribution is required, e.g. *log-series*, *log-normal*, *broken stick*, etc. Hence, abundance data are required.

In the case of non-parametric estimators, only absence-presence data are required (yet abundance data as input are not excluded in non-parametric models, as, for example, in the case of *Chao1* and *ACE* which will be explained in the following sections).

Two of the previously mentioned abundance models (Fisher's *log-series* and Preston's *log-normal*) can be directly used as parametric methods for estimating richness from abundance data, where, for example, Fisher's total number of species (S) can be estimated through

$$S = \alpha \ln\left(1 + \frac{N}{\alpha}\right) \tag{16.6}$$

with N the total number of individuals and α fitted using a representative sample with observed S and N.

In the case of Preston's *log-normal* approach, the calculation of S is straightforward. S basically equals the summation of the value of each *octave* (once the distribution is *unveiled*) or, in a continuous way, the integration of the fitted *log-normal* distribution (see Colwell and Coddington 1994 and Fig. 16.3).

16.2.2.2 Non-parametric Models

Out of the multiple non-parametric approaches of estimating richness, the most cited in the literature are the *Michaelis-Menten index* (usually not classified as non-parametric nor parametric (Michaelis and Menten 1913)), *Chao1 and Chao2* (Chao 1984, 2005), *bootstrap* (Smith and van Belle 1984), *jackknife* (Heltshe and Forrester 1983), the a*bundance-based coverage estimate* (ACE) and the *incidence-based coverage estimate* (ICE) (Chao and Mark 1993; Chao et al. 2004).

The *Michaelis-Menten* model is one of the best known asymptotic approaches and was used initially for enzymatic reactions. This model was applied in several biodiversity studies (e.g. Colwell and Coddington 1994; Chazdon et al. 1998; Toti et al. 2000) and requires abundance data (e.g. counts of different soil types). This model can be written as

$$S(n) = \frac{S_{\max} n}{B + n} \tag{16.7}$$

where $S(n)$ is the number of classes (soil orders, suborders, etc.) observed in n samples, S_{max} is the total number of classes in the assemblage and B is the sampling effort required to detect 50% of S_{max}.

The *Chao1* index is an asymptotic model, which similar to the *Michaelis-Menten* model uses abundance data. This index has been widely adopted by ecologists and also by pedologists in pedodiversity studies (Colwell and Coddington 1994; Ibáñez et al. 2003). This index is based on the proportion of singletons and doubletons (classes which are represented by one and two individuals, respectively) in a geographical area, and it can be easily calculated by

$$S_{\text{Chao1}} = S_{\text{obs}} + \frac{F_1^2}{F_2} \tag{16.8}$$

with S_{Chao1} the number of species in a sample, F_1 the number of singletons and F_2 the number of doubletons.

Anne Chao and collaborators (Chao and Lee 1992; Chao and Mark 1993) devised a set of more sophisticated estimators, as detailed in Magurran (2013). The a*bundance-based coverage estimate* (ACE), and the *incidence-based coverage estimate* (ICE). The ACE is written as follows:

$$S_{\text{ACE}} = S_{\text{abund}} + \frac{S_{\text{rare}}}{C_{\text{ace}}} + \frac{F_1}{C_{\text{ace}}}\gamma^2_{\text{ACE}} \tag{16.9}$$

where S_{rare} are the number of rare classes ($\leq$10 individuals) and S_{abund} are the number of abundant classes (>10 individuals). F_1 is the number of singletons; C_{ACE} is the relation between singletons and the number of rare classes:

$$C_{\text{ACE}} = 1 - \frac{F_1}{N_{\text{rare}}} \tag{16.10}$$

with N_{rare} the number of individuals in rare classes and γ^2_{ACE} the coefficient of variation of F_i's, expressed as

$$\gamma^2_{\text{ACE}} = \max\left\{\frac{S_{\text{rare}}}{C_{\text{ace}}}\frac{\sum_{i=1}^{10} i\,(1-1)\,F_i}{(N_{\text{rare}})\,(N_{\text{rare}}-1)} - 1, 0\right\} \tag{16.11}$$

with F_i the number of classes with i individuals.

The *bootstrap* method was first used as a richness estimator by Smith and van Belle (1984) and uses absence/presence data as follows:

$$S_{\text{boot}} = S_{\text{obs}} + \sum_{k=1}^{S_{\text{obs}}} (1 - p_k)^n \tag{16.12}$$

with p_k the number of quadrats where k class is present and n is the number of samples.

Chao2 uses the same principles of its predecessor; however, unlike *Chao1* and similar to the *bootstrap* method, it uses absence/presence data. The index uses the proportion of the number of classes (or soil types) that occur in one location only (Q_1) (unique classes), to the number of classes that occur in two locations (Q_2), and can be written as

$$S_{\text{Chao2}} = S_{\text{obs}} + \frac{Q_1^2}{2Q_2} \tag{16.13}$$

The last index (among many others not presented here) is the *jackknife* index, which has two versions. The first one (S_{Jack1}) uses only the number of classes found in a single sample, while the second one (S_{Jack2}) uses both, the classes found in one location (Q_1) and in exactly two locations (Q_2):

$$S_{\text{Jack1}} = S_{\text{obs}} + Q_1\left(\frac{n-1}{n}\right) \tag{16.14}$$

$$S_{\text{Jack2}} = S_{\text{obs}} + \left[Q_1\left(\frac{2n-3}{n}\right) + Q_2\frac{(n-2)^2}{n\,(n-1)}\right] \tag{16.15}$$

Figure 16.7 shows the performance of six of the previously described indices using the information from Fig. 16.6. On the left-hand side of Fig. 16.7 those indices are shown that were designed for abundance data and on the right-hand side indices that were designed to use absence/presence information.

Even though all of these indices have been extensively used in ecology and biological sciences, their use in soil science is limited, and except for a few studies, they are hardly reported in the literature. Undoubtedly, once a soil map is finished (both conventional and digital), the total number of classes is known a priori; however, the study of the relation between sampling efforts or area sampled and richness becomes valuable information when used in a conservationist approach (Ibáñez and Bockheim 2013).

The previous examples are known as *species accumulation curves*, in our case *classes accumulation curves*. Even though classes can be represented using richness information versus samples as shown in Fig. 16.7, the *area-richness approach* seems to be more useful for soil sciences.

Figure 16.8 shows four maps made by using the soil types presented in Fig. 16.6. Each map is presented with four different pixel resolutions (area), i.e. 4, 25, 225 and 2500 km^2, respectively. The scale of each map indicates the amount of different soil types contained on each pixel; in this way as the area of the pixel increases, the total amount of different soil types belonging to that particular pixel will increase accordingly.

As a rough exercise, if we decide to preserve the largest number of soil types (at the order taxonomical level) in a specific area, we can plot the richness per area increase as presented in Fig. 16.9. As other authors have observed, the *area-richness* relation follows the same pattern that was explored in previous sections (Phillips and Marion 2005; Toomanian et al. 2006; Ibáñez and Bockheim 2013).

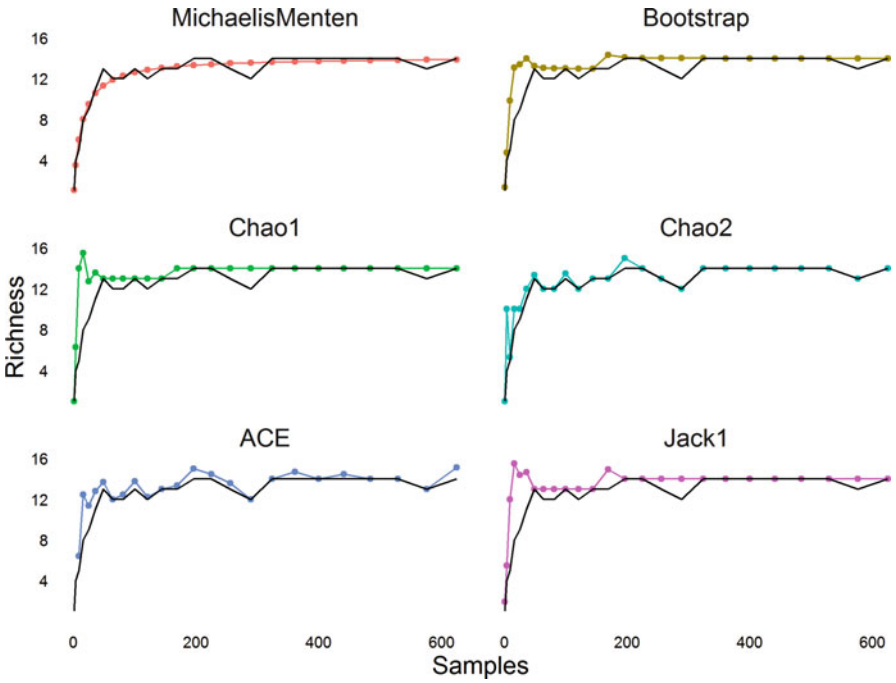


Fig. 16.7 Six different approaches of sampling effort/richness modelling. On the left-hand side, indices are shown that were designed for abundance data, and on the right-hand side, indices are shown that were designed to use absence/presence information

16.2.3 Diversity Indices

The most challenging way of describing the diversity of a community (or any environmental system) is to be able to represent diversity in a numerical manner. Hence, all aspects of diversity need to be represented, i.e. the *abundance* and *richness* of species in a community, the arrangement of its components (*evenness*) and the *distance* between individuals.

Since there is a myriad of diversity indices, a word of caution needs to be made here: as the reader may guess, every index will have a particular sensitivity to a particular aspect of diversity. Some of the indices will put more emphasis or weight on the different aspects of a community, i.e. *richness*, *even* or *uneven* communities, *rare* or *common* classes within a community, *similarity* between classes, etc.

For example, there are indices that are especially designed to reflect the *richness* of a system, e.g. *Margalef's* index (Eq. 16.16):

$$D_{\mathrm{Mg}} = \left(\frac{S-1}{\ln N}\right) \tag{16.16}$$

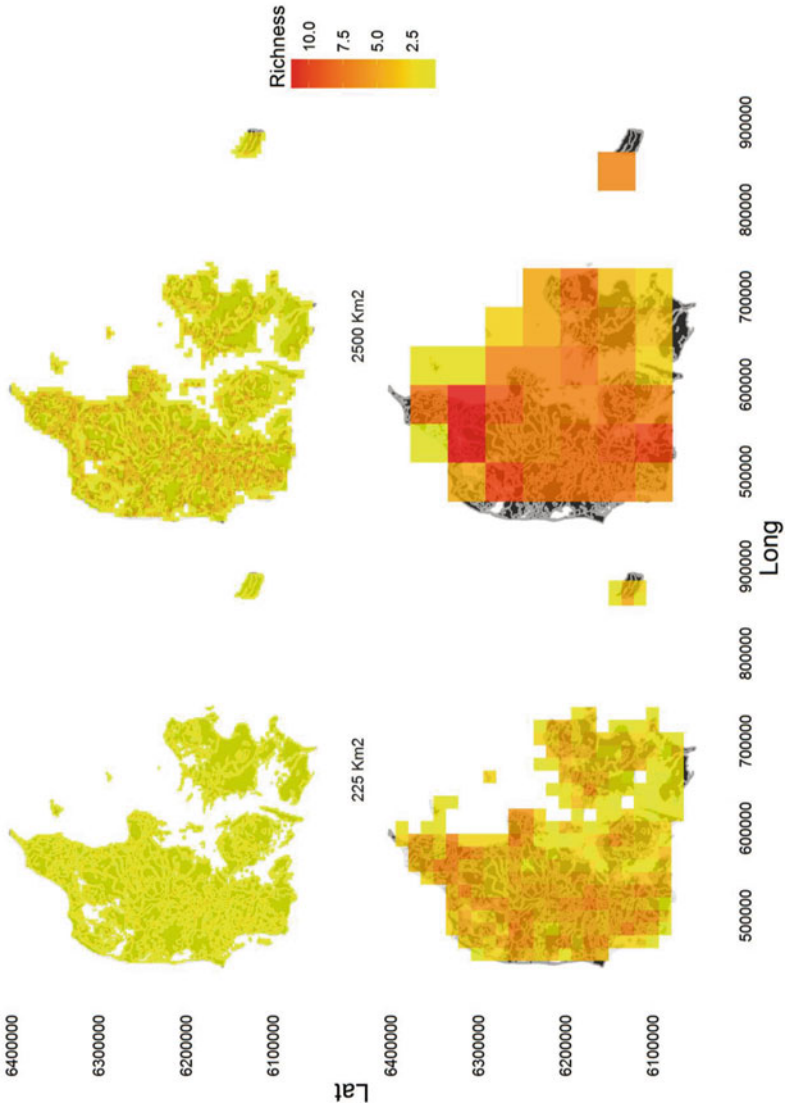


Fig. 16.8 Soil richness of Danish soil types as specified in Jacobsen (1984). Increasing scales represent the size of the pixel in km^2. The colour scale represents the amount of different soil types belonging to each pixel

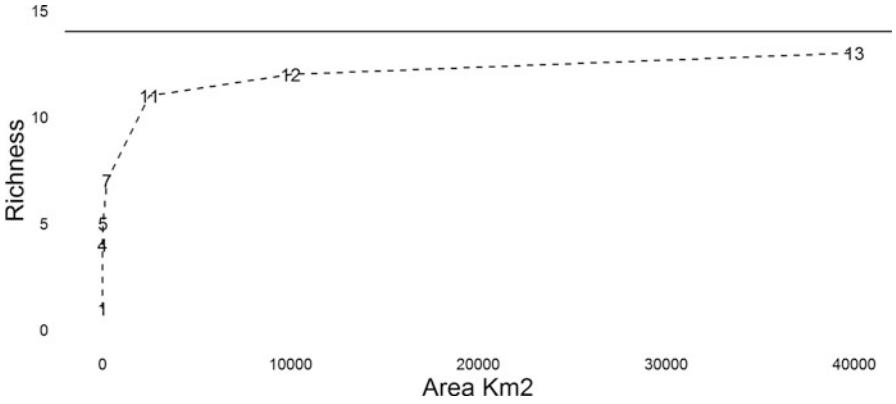


Fig. 16.9 *Dashed line* shows Arrhenius/Gleason curve showing the increase of richness versus area observed. *Numbers* represent the amount of different classes contained in a single pixel of size x (area km^2). The *continuous line* represents the total amount of soil types present in Denmark as detailed in Jacobsen (1984)

and *Menhinick's* index (Eq. 16.17):

$$D_{\mathrm{Mh}} = \frac{S}{\sqrt[2]{N}} \tag{16.17}$$

which calculate a relation between the number of classes S and the number of individuals in each class N. They need little information to be calculated; however, these indices do not take the relative abundance within the classes into consideration.

One of the first indices that considered both richness and relative abundance in a single value is the alpha (α) parameter of Fisher's *logarithmic series* that we explained in Sect. 16.2.1 (Fisher et al. 1943), where the number of species is a function of the number of individuals (Eq. 16.6).

In a similar manner, Preston's *log-normal* series' standard deviation (see Sect. 16.2.1) can be used as an index of diversity. The benefits of using these are that they consider both the richness and the distribution of the abundance of the classes of a community (with more weight given to richness as specified in Ibáñez et al. 1995).

Even though these indices have been repeatedly applied in biodiversity, as stated in Magurran (2013), their use has not been reported in pedodiversity studies except for some rare cases (Ibáñez et al. 1998).

By far, the two most popular indices in pedodiversity studies are *Shannon's* index or *Shannon's* entropy (Shannon 1948) and the *Simpson's* index (Simpson 1949).

Shannon's index was first published in *A Mathematical Theory of Communication* by Claude Shannon (1948) and a year after materialized in the book written by the same author and Warren Weaver: *The Mathematical Theory of Communication* (Shannon and Weaver 1949). It is therefore also sometimes called the *Shannon-Weaver* index. It was originally used in information theory and applied later on to

several areas of environmental sciences (Lowitz 1984; Ibáñez et al. 1995; Mendicino and Sole 1997; Bishop et al. 2001; Minasny et al. 2010; Tamames et al. 2010; Legendre and Legendre 2012). The index takes the following form:

$$H_c = -\sum_{c=1}^{k} p_{\mathrm{ic}} \ln\ p_{\mathrm{ic}} \tag{16.18}$$

where p_{ic} is the proportional abundance of individuals of class c at site i relative to the total amount of individuals. Even though this index is frequently presented by the natural logarithm, the original version was expressed in logarithm with base 2 because it was based on the idea of bits of information.

The *Simpson* index can be written as (Simpson 1949)

$$E_c = \sum_{c=1}^{k} {p_{\mathrm{ic}}}^2 \tag{16.19}$$

This index was first derived by Gini in a slightly different form (after Ceriani and Verme 2012):

$$D_c = 1 - \sum_{c=1}^{k} {p_{\mathrm{ic}}}^2 \tag{16.20}$$

The Gini index is therefore also now known as the *Gini-Simpson* index or *Simpson's index of diversity*. In Eqs. 16.19 and 16.20, p_{ic} is the proportional abundance of individuals of class c at site i, relative to the total amount of individuals found. The *Gini-Simpson's* index is regarded to be more intuitive as its values range from 0 to 1, with values closer to 1 equal to being more diverse. Interestingly, in cases with a high level of partitioning (e.g. many classes are present), the *Gini-Simpson* index will generate values closer to unity (1), and any change in class abundance (e.g. in soil sciences a permanent land-use modification which could lead to a reclassification) will have no significant impact on the actual index value.

In order to create a mathematical generalization, Alfréd Rényi wrote his own formula, now known as *Rényi's* entropy (originally described by a logarithm of base 2 in Rényi 1961):

$$H_\alpha = \frac{\ln \sum_i p_i^\alpha}{1 - \alpha} \tag{16.21}$$

where α is a *sensitivity* parameter (after Feoli et al. 2013; Leinster and Cobbold 2012). Note that if $\alpha \to 0$, the natural logarithm of the richness is obtained (lnS), and if $\alpha \to 1$, we obtain *Shannon's* index (Eq. 16.18). On the other hand, if $\alpha \to 2$, then –ln of *Simpson's* index is obtained (Eq. 16.19).

As Guiasu and Guiasu (2012) noticed, Hill (1973) claimed that by modifying *Rényi's* entropy, a new biodiversity index (often called the Hill numbers) could be expressed:

$$N_a = \left(\sum_i p_i^a \right)^{\frac{1}{1-a}} \tag{16.22}$$

when a $\rightarrow$ 0, we obtain S (richness), when a $\rightarrow$ 1, we obtain the exponential of *Shannon's* index:

$$N_{a \rightarrow 1} = e^{-\sum_i p_i \ln\ p_i} \tag{16.23}$$

and when a $\rightarrow$ 2, we obtain the reciprocal of *Simpson's* index or 1/*Simpson's* index. According to some authors (Jost 2006; Feoli et al. 2013), the fact that this index reaches the maximum number of species S is considered as advantageous.

Later on, Patil and Taillie (1976) proposed their own equation:

$$H_\beta = \frac{\left(1 - \sum_i p_i^{\beta+1}\right)}{\beta} \tag{16.24}$$

when $\beta \rightarrow -1$, the richness minus one is obtained ($S-1$), for $\beta \rightarrow 0$ *Shannon's* index (Eq. 16.18), and *Gini's* index is obtained for $\beta \rightarrow 1$ (Eq. 16.20).

It becomes evident now why some indices (e.g. *Shannon's* and *Simpson's* index) are persistently appearing in almost every pedodiversity study (Ibáñez et al. 1995, 1990; Ibàñez and García-Alvarez 2002; Guo et al. 2003; Phillips and Marion 2005; Toomanian and Esfandiarpoor 2010; Ibáñez and Bockheim 2013).

In this regard, Feoli et al. (2013) used the approach of Leinster and Cobbold (2012) and showed that indices can complement themselves and in this way create a complete picture of the diversity of an area. They showed that while some communities differ greatly based on one index, they can have similarities when using another. Feoli et al. (2013) therefore proposed to use *diversity profiles* where, for example, as shown in Eq. 16.21, the values of the *sensitivity parameter* α, the diversity of a *profile* can be created, evaluating the *sensitivity parameter* with different values, i.e. $\alpha \rightarrow 0$ (ln$S$), $\alpha \rightarrow 1$ (*Shannon*) and $\alpha \rightarrow 2$ ($-$ln of *Simpson*).

Thus, so far a basic aspect in the calculation of the *abundance*, *richness* and the *diversity* of classes is the lack of consideration of how different the classes are. In fact, as Peet (1974) clearly states, one of the assumptions of any diversity study is:

> *All species or classes are assumed to be equally different.*

However, as nicely summarized by Leinster and Cobbold (2012):

> *Nonspecialists are amazed to learn that a community of six dramatically different species is said to be no more diverse than a community of six species of barnacle.*

These simple statements roughly summarize the motivation for an extensive area of diversity research, which is taking into consideration the similarity (or dissimilarity) of classes when considering the diversity of a community. One of the first that attempted to create a link between diversity of classes and distance between them was C. R. Rao (1982, 2010) who created the well-known *Rao's* quadratic entropy:

$$Q = \sum_i \sum_j p_i d_{\mathrm{ij}} p_j \tag{16.25}$$

where p_i is the proportional abundance of individuals in class i and j relative to the total amount of individuals and d_{ij} is the distance matrix between all pairs of p_i and p_j (matrix of dissimilarities). In a hypothetical case where all classes are equally different (i.e. if $d_{\mathrm{ij}} = 1$ for all $\mathrm{i} \neq \mathrm{j}$ and $d_{ii} = 0$), then *Rao's* entropy is simply *Gini's* index (Eq. 16.20), and if a matrix of similarity is used instead of dissimilarity (i.e. $d_{ij} = 0$ and $d_{ii} = 1$ instead of $d_{\mathrm{ij}} = 1$ and $d_{ii} = 0$), then Simpson's index is obtained (Eq. 16.19).

As explained by Feoli et al. (2013), it is possible to transfer *Gini's* to *Rényi's* formulae by applying a dissimilarity matrix, and the same can be applied to *Shannon's* index by introducing a dissimilarity matrix using the equation proposed by Ricotta and Szeidl (2006):

$$H = -\sum_i p_i \ln \left(1 - \sum_{i \neq j} d_{\mathrm{ij}} p_{\mathrm{ij}} \right) \tag{16.26}$$

McBratney and Minasny (2007) found that, as shown in a previous paper by Guo et al. (2003), when using *Shannon's* index and the observed richness S in the calculations of the pedodiversity at different taxonomical levels (order, suborder, great group, sub-group family and series), the diversity increased as the taxonomical levels increased (from orders to series). This phenomenon was demonstrated for the soils of the United States of America. The authors then hypothesized that, since the same land cover was used in the calculations, the level of taxonomy should not significantly affect the diversity. In order to test this hypothesis, McBratney and Minasny (2007) used *Ricotta and Szeidl's* approach (Eq. 16.26) along with *Rao's* quadratic entropy approach (Eq. 16.25) and concluded that while *Shannon's* diversity is greatly affected by the number of classes (see Eq. 16.18), distance-based metrics do not increase linearly according to the number of different classes.

Following on, Minasny et al. (2010) used these concepts to calculate pedodiversity at a global level. They showed that while *Shannon's* diversity was linearly related to the area sampled, this relation was not as evident with distance-based metrics (in this example *Rao's* index (Eq. 16.25) was used). Still, as observed by other authors, there was a consistent relation between these distance and non-distance-based metrics indices. A recent example of this relationship is shown in Figs. 16.10 and 16.11, where *Shannon's* and *Ricotta-Szeidl's* metrics were

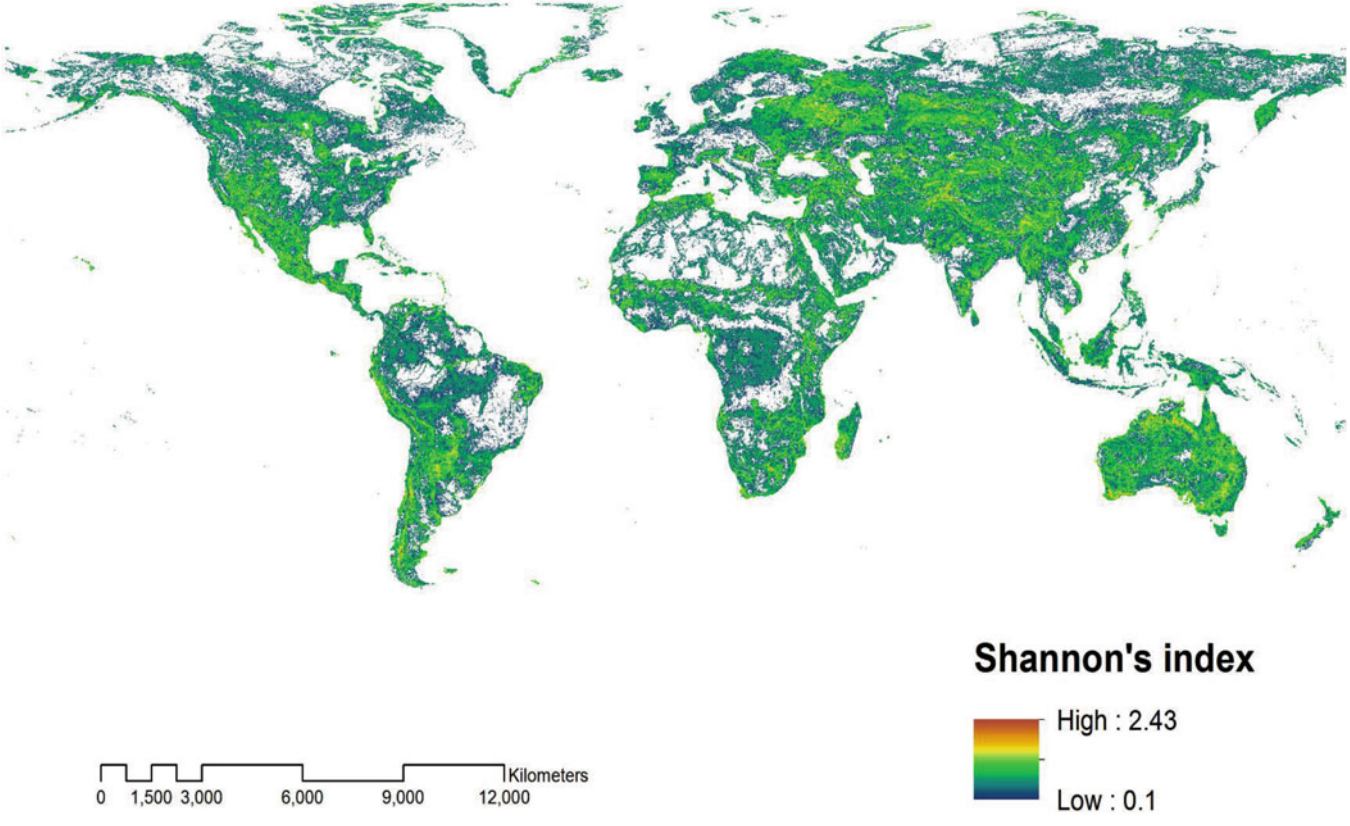


Fig. 16.10 Pedodiversity on a global scale using the *Shannon* index. *White areas* represent an index value of 0

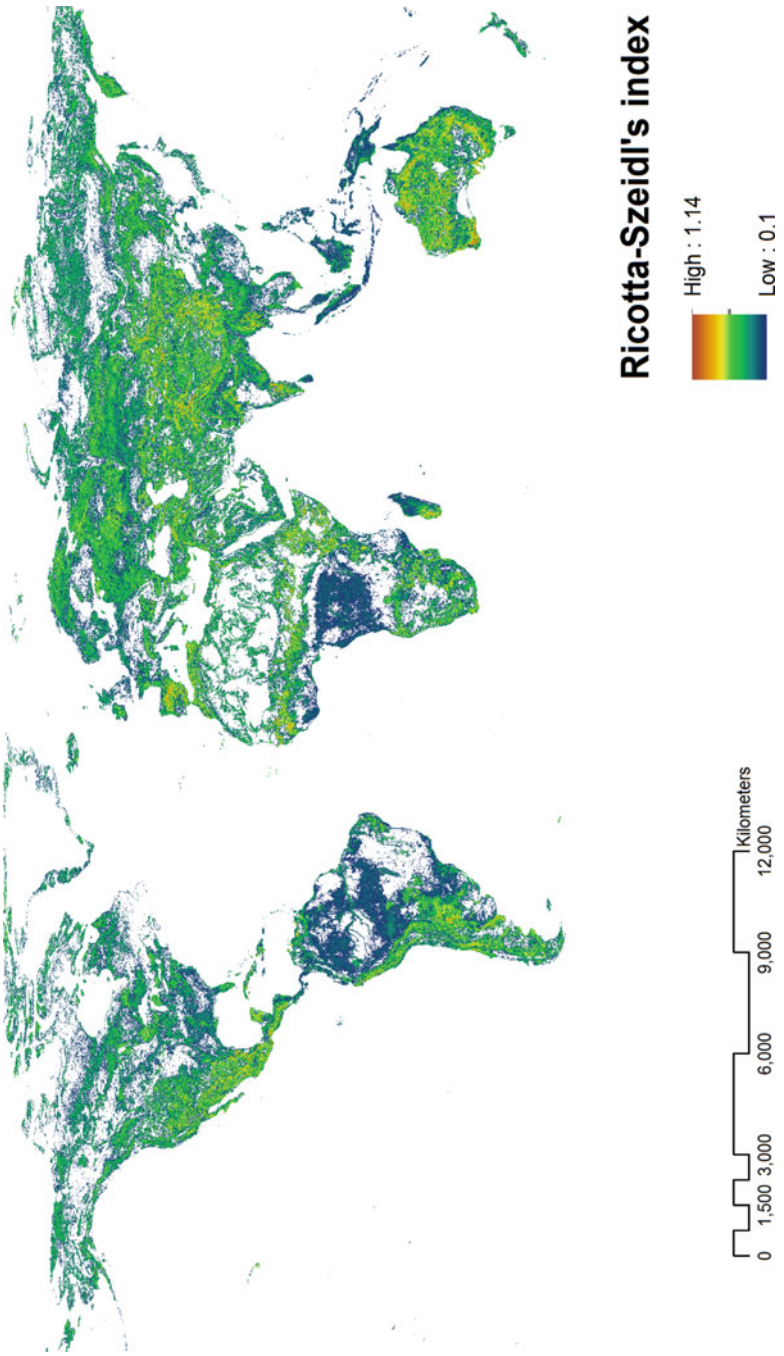


Fig. 16.11 Pedodiversity on a global scale using the *Ricotta-Szeidl* index. *White areas* represent an index of value 0

calculated on the global extent. The input information, a global soil class map based on the World Reference Base soil classification scheme, was sourced from Hengl et al. (2014). The *Shannon* and *Ricotta-Szeidl* indices were calculated for an approximately 10 km by 10 km moving window. Equation 16.26 was then used to calculate the distance matrix which is the same measure that was used by Minasny et al. (2010) (refer to Table 1 in Minasny et al. 2010).

Furthermore, Minasny et al. (2010) found that while diversity maps show good reliability, they are highly sensitive to coverage and density of the soil information used. An example of this behaviour can be seen in Fig. 16.12.

If we consider the input data used to predict the WRB soil classes by Hengl et al. (2014), e.g. MODIS satellite images, a digital elevation model (DEM) and its derivatives among others, it is relatively sensible to expect high pedodiversity associated with highly complex landscapes such as the Andes mountain range in South America, as seen in Fig. 16.12. However, some artefacts related to the creation of discrete classes are also evident in the map created. Large areas that correspond to values of 0 pedodiversity shown in white are examples of this, especially those that fall within the Amazon region which is considered one of the world regions with a high biodiversity. This phenomenon then raises the question of "Do we need more classes in our soil classification systems?" It is known, for example, that as we go to the lower levels of hierarchical (soil) classification systems, the richness (i.e. the number of classes at that level) will increase. This does of course have an impact on both the *Shannon* index and to some degree also on the *Ricotta-Szeidl* index, which was explained in previous sections. Furthermore, considering the fact that maps are just a two-dimensional representation of multiple factors of soil formation, another question arises of "Can we use various soil properties as input information to calculate pedodiversity?" which will be discussed in the following section.

16.3 Soil Property Variation and Pedodiversity

The relation between soil property variation and pedodiversity has had a central place in how pedodiversity is perceived and hence has been calculated for the last two decades. The earliest attempts to link the concepts of soil property variation and pedodiversity were made by McBratney (1992). McBratney (1992) argued that because soil properties fluctuate in space (and time), the soil's diversity is a representation of this variation. On the contrary, Saldaña et al. (1998) and Saldaña and Ibáñez (2007) observed that independent soil properties do not necessarily follow the same evolutionary path. They found, for example, that along a soil chronosequence (i.e. a sequence of different evolutionary stages of soils sharing the same parent material), soil evolution can simultaneously lead to a homogenization of soil properties and on the other hand to processes of horizonation, which translates to an increase of pedodiversity. Toomanian and Esfandiarpoor (2010) revisited this subject and emphasized it as a major challenge of pedodiversity (refer to Sect. 1 "Soil Spatial Variability Versus Pedodiversity" in Toomanian and

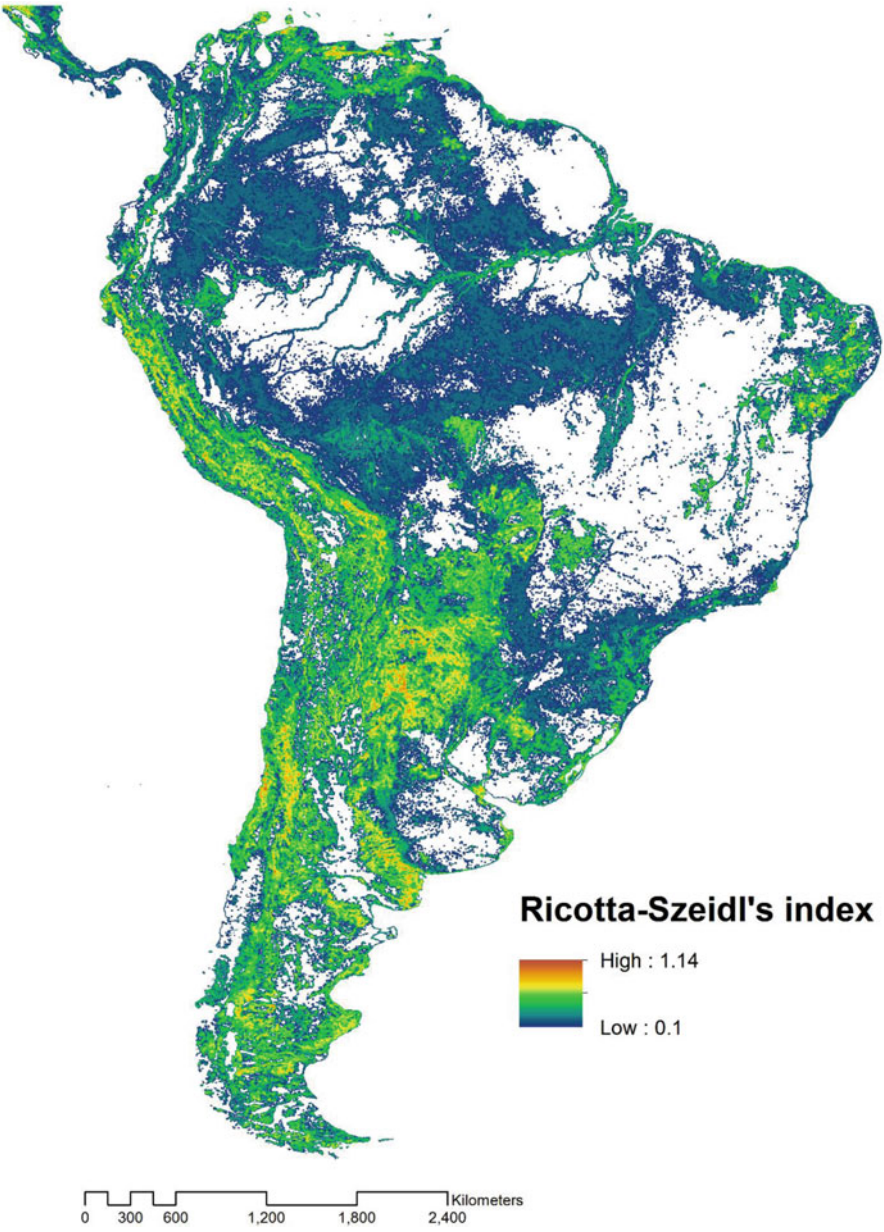


Fig. 16.12 Pedodiversity in South America calculated using the *Ricotta-Szeidl* index. *White areas* represent an index value of 0

Esfandiarpoor (2010)). The authors proposed that if a relation between soil property variation and pedodiversity is to be made, then a multidimensional analysis should be employed (i.e. to consider the joint variation of several soil properties at the

same time). Following on, Petersen et al. (2010) presented two new approaches for the calculation of pedodiversity considering the variation of soil properties as an indicator of pedodiversity (namely, pH, EC, C and S, TOC and the magnitude of rooting space). The authors used this analysis to establish relationships with biodiversity of higher plants in South Africa. They observed that the highest correlation between pedo- and biodiversity was found using what they termed the *variable space* (VS) *pedodiversity*. This method uses the area (or volume) of a convex hull formed by the extreme values of observations projected in a n-dimensional space and created with a set of (normalized) selected variables. The same procedure was used in an earlier study by Islam et al. (2005) in order to estimate soil variability using the soil's spectroscopic information as the selected variables (for more details about the method, refer to Petersen et al. 2010 and Islam et al. 2005).

Petersen et al. (2010) concluded that the VS method was highly advantageous when compared to taxonomic classification systems, as it avoids the definition of classes and thus does not lose information about quantitative soil properties. But they also showed that taxonomic classification of pedodiversity is still a valuable tool as it is sensitive enough to represent the pedodiversity of a site even at a 1 km^2 extent.

Following on, Fajardo et al. (2017) used a similar VS approach (Islam et al. 2005; Petersen et al. 2010) finding the resemblance of these methods with what is known as *functional diversity*, previously defined by Tilman (2001): "A branch or a subset of biological diversity, that deals with the components that influence how an ecosystem operates and functions, and that can be measured by the values (and ranges in the values), of the organismal 'traits' that influence one or more aspects of the functioning of an ecosystem". Fajardo et al. (2017) concluded that by using a large set of variables (in their case spectroscopic information), the pedodiversity and the multidimensional variation of soil properties forms a linear relationship, as proposed by Toomanian and Esfandiarpoor (2010).

16.4 Final Remarks

From the concepts discussed in this chapter, we can conclude that the characterization of pedodiversity involves a variety of different facets and also requires the understanding of elementary concepts of biodiversity. As discussed in the introductory section, this chapter only covered a fraction of the advances made in pedodiversity. Our intention was to create a simplified view of the underlying concepts. In this regard, we have focused on the basic ideas rather than on their applications.

Even though it may appear that pedodiversity studies are behind its ecological cousins, the nature of soil, e.g. its geographical and taxonomical continuity, resulted in a deeper understanding of the ways of describing its variation. The study of the soil's diversity is increasingly recognized by the scientific community and in

fact plays an essential role in the studies of ecosystems. Several recent studies are recognizing the close relationship between the soil's distribution and functions and the presence and diversity of micro- and macro-biological communities (e.g. Ibañez et al. 2012; Vries et al. 2013).

Moreover, the ever-increasing computing power capabilities can through big data analysis unveil ecological patterns that once were hardly understood. Many questions are still to be answered. Acquiring a deeper understanding of how the soil resource is distributed and impacts on the environment is therefore yet to be discovered.

References

Arrhenius O (1920) Öcologische studien in den Stockholmer Schären, Stockholm, Sweden
Arrhenius O (1921) Species and area. J Ecol 9(1):95–99
Bishop T, McBratney A, Whelan B (2001) Measuring the quality of digital soil maps using information criteria. Geoderma 103(1):95–111
Bliss CI, Fisher RA (1953) Fitting the negative binomial distribution to biological data. Biometrics 9(2):176–200
Butler BE (1981) Soil classification for soil survey. Oxford University Press, Oxford
Campbell JB, Edmonds WJ (1984) The missing geographic dimension to soil taxonomy. Ann Assoc Am Geogr 74(1):83–97
Campbell N, Mulcahy M, Mcarthur W (1970) Numerical classification of soil profiles on the basis of field morphological properties. Soil Res 8(1):43–58
Ceriani L, Verme P (2012) The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. J Econ Inequal 10(3):421–443
Chao A (1984) Nonparametric estimation of the number of classes in a population. Scand J Stat 11(4):265–270
Chao A (2005) Species estimation and applications. In: Balakrishnan N, Read CB, Vidakovic B (eds) Encyclopedia of statistical sciences, 2nd edn, vol 12. Wiley, New York, pp 7907–7916
Chao A, Lee S-M (1992) Estimating the number of classes via sample coverage. J Am Stat Assoc 87(417):210–217
Chao A, Mark CKY (1993) Stopping rules and estimation for recapture debugging with unequal failure rates. Biometrika 80(1):193–201
Chao A, Chazdon RL, Colwell RK, Shen T-J (2004) A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol Lett 8(2):148–159
Chazdon RL, Colwell RK, Denslow JS, Guariguata MR (1998) Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of northeastern Costa Rica
Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc B Biol Sci 345(1311):101–118
de Vries FT, Thébault E, Liiri M, Birkhofer K, Tsiafouli MA, Bjørnlund L, Bracht Jørgensen H, Brady MV, Christensen S, de Ruiter PC, d'Hertefeldt T, Frouz J, Hedlund K, Hemerik L, Hol WHG, Hotes S, Mortimer SR, Setälä H, Sgardelis SP, Uteseny K, van der Putten WH, Wolters V, Bardgett RD (2013) Soil food web properties explain ecosystem services across European land use systems. Proc Natl Acad Sci 110:14296–14301
Fajardo MP, McBratney AB, Minasny B (2017) Measuring functional pedodiversity using spectroscopic information. Catena 152:103–114
Feoli E, Ganis P, Ricotta C (2013) Measuring diversity on environmental systems. In: Ibáñez JJ, Bockheim JG (eds) Pedodiversity. CRC Press, Boca Raton

Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. J Anim Ecol 12(1):42–58
Fridland V (1965) Make-up of the soil cover. Sov Soil Sci 4:343–354
Fridland VM (1974) Special issue soil science in the U.S.S.R structure of the soil mantle. Geoderma 12(1):35–41
Gleason HA (1922) On the relation between species and area. Ecology 3(2):158–162
Guiasu RC, Guiasu S (2012) The weighted Gini-Simpson index: revitalizing an old index of biodiversity. Int J Ecol 2012:1–10
Guo Y, Gong P, Amundson R (2003) Pedodiversity in the United States of America. Geoderma 117(1–2):99–115
Heltshe JF, Forrester NE (1983) Estimating species richness using the jackknife procedure. Biometrics 39(1):1–11
Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GB, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JG, Walsh MG (2014) SoilGrids1km—global soil information based on automated mapping. PLoS One 9(8):e105992
Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. Ecology 54(2):427–432
Holmgren GG (1988) The point representation of soil. Soil Sci Soc Am J 52(3):712–716
Hughes PA, McBratney AB, Minasny B, Campbell S (2014) End members, end points and extragrades in numerical soil classification. Geoderma 226–227:365–375
Ibáñez JJ, Bockheim JG (2013) Pedodiversity. CRC Press, Boca Raton
Ibàñez JJ, García-Alvarez A (2002) Diversidad: biodiversidad edáica y geodiversidad. Edafología 9(3):329–385
Ibañez J, García A, Monturiol F, El Hombre, C.E.d.P (1987) Heterogeneidad edáfica inducida por el adehesamiento del bosque mediterráneo. MAB, Instituto Nacional de Edafología y Agrobiología "Jose María Albareda"
Ibañez JJ, Jiménez Ballesta R, García Alvarez A (1990) Soil landscapes and drainage basins in Mediterranean mountain areas. Catena 17(6):573–583
Ibáñez JJ, De-Albs S, Bermúdez FF, García-Álvarez A (1995) Pedodiversity: concepts and measures. Catena 24(3):215–232
Ibañez JJ, De-Alba S, Lobo A, Zucarello V (1998) Pedodiversity and global soil patterns at coarse scales (with discussion). Geoderma 83(3):171–192
Ibáñez J, García-Álvarez A, Saldaña A, Recatalá L (2003) Scientific rationality, quantitative criteria and practical implications in the design of soil reserves networks: their role in soil biodiversity and soil quality studies. Preserving soil quality and soil biodiversity—the role of surrogate indicators. IMIA-CSIC, Zaragoza, pp 191–274
Ibáñez JJ, Krasilnikov PV, Saldaña A (2012) REVIEW: archive and refugia of soil organisms: applying a pedodiversity framework for the conservation of biological and non-biological heritages. J Appl Ecol 49(6):1267–1277
Isbell RF (2002) The Australian soil classification. In: Isbell RF (ed) Australian soil and land survey handbook, vol 4. CSIRO Publishing, Collingwood
Islam K, McBratney A, Singh B (2005) Rapid estimation of soil variability from the convex hull biplot area of topsoil ultra-violet, visible and near-infrared diffuse reflectance spectra. Geoderma 128(3–4):249–257
Jacobsen NK (1984) Soil map of Denmark according to the FAO-UNESCO legend. Geogr Tidsskr Dan J Geogr 84(1):93–98
Jacuchno V (1976) K voprosu opredelenia raznoobrazija struktury pocvennogo pokrova. Tezisy dokl. III. VSES Sov. Po structure pocv. Pokrova. Vaschnil, Moskvao
Jahn R, Blume HP, Asio VB, Spaargaren O, Schad P (2006) Guidelines for soil description, 4th edn. FAO, Rome
Johnson WM (1963) The pedon and the polypedon. Soil Sci Soc Am J 27(2):212–215
Jost L (2006) Entropy and diversity. Oikos 113(2):363–375
Krasil'nikov PV (2009) A handbook of soil terminology, correlation and classification. Earthscan, London

Legendre P, Legendre LF (2012) Numerical ecology, vol 24. Elsevier, Amsterdam
Leinster T, Cobbold CA (2012) Measuring diversity: the importance of species similarity. Ecology 93(3):477–489
Linkeš V, Simonyová S, Tobrman D (1983) Zložitosť štruktúry pôdneho pokryvu hodnotena pomocou entropie. Ved. Pr. VU podozalec. Vysivy rastl. Bratislave, pp 133–145
Lowitz GE (1984) Mapping the local information content of a spatial image. Pattern Recogn 17(5):545–550
MacArthur RH (1957) On the relative abundance of bird species. Proc Natl Acad Sci 43(3): 293–295
Magurran AE (1988) Why diversity? Ecological diversity and its measurement. Springer, Berlin, pp 1–5
Magurran AE (2013) Measuring biological diversity. Wiley, Malden
McBratney A (1992) On variation, uncertainty and informatics in environmental soil management. Soil Res 30(6):913–935
McBratney A, Minasny B (2007) On measuring pedodiversity. Geoderma 141(1–2):149–154
Mendicino G, Sole A (1997) The information content theory for the estimation of the topographic index distribution used in TOPMODEL. Hydrol Process 11:1099–1114
Michaelis L, Menten ML (1913) Die kinetik der invertinwirkung. Biochemist 49(333–369):352
Minasny B, McBratney AB, Hartemink AE (2010) Global pedodiversity, taxonomic distance, and the world reference base. Geoderma 155(3):132–139
Motomura I (1932) On the statistical treatment of communities. Zool Mag 44:379–383
NRCS, U (1993). Soil Survey Division Staff (1993) Soil survey manual. Soil conservation service. US Department of Agriculture Handbook 18:315
Patil G, Taillie C (1976) Ecological diversity: concepts, indices and applications. Proceedings of the international biometric conference. Biometric Soc, vol 20. pp 383–411
Peet RK (1974) The measurement of species diversity. Ann Rev Ecol Syst 5:285–307
Petersen A, Gröngröft A, Miehlich G (2010) Methods to quantify the pedodiversity of 1 km2 areas – results from southern African drylands. Geoderma 155(3–4):140–146
Phillips JD, Marion DA (2005) Biomechanical effects, lithological variations, and local pedodiversity in some forest soils of Arkansas. Geoderma 124(1–2):73–89
Prado P, Dantas-Miranda M, Chalom A (2016) R-package sads: maximum likelihood models for species abundance distributions
Preston FW (1948) The commonness, and rarity, of species. Ecology 29(3):254–283
Rao CR (1982) Diversity and dissimilarity coefficients: a unified approach. Theor Popul Biol 21(1):24–43
Rao CR (2010) Quadratic entropy and analysis of diversity. Sankhya A 72(1):70–80
Rényi A (1961) On measures of entropy and information. Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, pp 547–561
Ricotta C, Szeidl L (2006) Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. Theor Popul Biol 70(3):237–243
Saldaña A (2013) Pedodiversity and landscape ecology, pedodiversity. CRC Press, Boca Raton, pp 105–132
Saldaña A, Ibáñez JJ (2004) Pedodiversity analysis at large scales: an example of three fluvial terraces of the Henares River (central Spain). Geomorphology 62(1–2):123–138
Saldaña A, Ibáñez JJ (2007) Pedodiversity, connectance and spatial variability of soil properties, what is the relationship? Ecol Model 208(2–4):342–352
Saldaña A, Stein A, Zinck JA (1998) Spatial variability of soil properties at different scales within three terraces of the Henares River (Spain). Catena 33(3–4):139–153
Shannon C (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
Shannon C, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana, p 29
Simpson EH (1949) Measurement of diversity. Nature 163:688
Smith EP, van Belle G (1984) Nonparametric estimation of species richness. Biometrics 40(1): 119–129

Tamames J, Abellán JJ, Pignatelli M, Camacho A, Moya A (2010) Environmental distribution of prokaryotic taxa. BMC Microbiol 10(1):85
Tilman D (2001) Functional diversity. Encycl Biodivers 3(1):109–120
Tokeshi M (1990) Niche apportionment or random assortment: species abundance patterns revisited. J Anim Ecol 59(3):1129–1146
Toomanian N, Esfandiarpoor I (2010) Challenges of pedodiversity in soil science. Eurasian Soil Sci 43(13):1486–1502
Toomanian N, Jalalian A, Khademi H, Eghbal MK, Papritz A (2006) Pedodiversity and pedogenesis in Zayandeh-rud Valley, Central Iran. Geomorphology 81(3–4):376–393
Toti DS, Coyle FA, Miller JA (2000) A structured inventory of Appalachian grass bald and heath bald spider assemblages and a test of species richness estimator performance. J Arachnol 28(3):329–345
Triantafilis J, Ward WT, Odeh IOA, McBratney AB (2001) Creation and interpolation of continuous soil layer classes in the lower Namoi valley. Soil Sci Soc Am J 65(2):403–413
Turner MG (1989) Landscape ecology: the effect of pattern on process. Annu Rev Ecol Syst 20(1):171–197
United States Department of Agriculture Soil, Conservation Service (1975) Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. US Government Printing Office
USDA Soil, Conservation Service (1999) Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. 2nd Edition. US Department of Agriculture, Soil Conservation Service
Whittaker RH (1972) Evolution and measurement of species diversity. Taxon 21(2/3):213–251
Zipf GK (1949) Human behavior and the principle of least effort; an introduction to human ecology. Addison-Wesley Press, Cambridge, MA

Chapter 17
Pedometric Valuation of the Soil Resource

David G. Rossiter, Allan E. Hewitt, and Estelle J. Dominati

> *"[W]hen you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind".*
>
> William Thompson (Lord Kelvin)
> – Lecture to the Institution of Civil Engineers, 3 May 1883

Soil forms the thin skin of the Earth and is the site of many ecological processes, transformations, and fluxes. It forms the substrate for most of the activities that take place at the Earth's surface, including almost all food production and human occupation, and underpins both natural and managed ecosystems. Soils differ in their structure, composition, and ability to function under a use. Soil is a multifunctional resource that affects human well-being both directly (e.g., food provision) and indirectly (e.g., surface and groundwater supplies) and that affects all near-land surface ecological processes. Clearly, soil is "valuable" as that term is understood in common language. The pedometric program as outlined in this book, i.e., the development of "quantitative methods for the study of soil distribution ... as a sustainable resource," should therefore include an attempt to quantify this value. Chapter 1 of the present book lists as the third of four items on the pedometric

D.G. Rossiter (✉)
School of Integrative Plant Science, Soil & Crop Sciences Section, Cornell University, Ithaca, NY 14850, USA

ISRIC-World Soil Information, PO Box 353, 6700 AJ Wageningen, The Netherlands
e-mail: d.g.rossiter@cornell.edu

A.E. Hewitt
Manaaki Whenua – Landcare Research, PO Box 6940, Lincoln 7640, New Zealand
e-mail: HewittA@landcareresearch.co.nz

E.J. Dominati
Farm Systems and Environment, Land and Environment, AgResearch Grasslands, Palmerston North, New Zealand
e-mail: Estelle.Dominati@agresearch.co.nz

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_17

agenda "evaluating the utility and quality of soil," and it is in this sense that we attempt in this chapter to define and quantify the value of the soil resource. This process is referred to as "valuation."

As the review of Robinson et al. (2014) on the value of the soil resource states, "Common to all valuation is the initial and fundamental question, what is the valuation for? There must be a clearly defined policy objective or management purpose for valuation." In this chapter we consider that the fundamental reason to value the soil resource is to include a fair representation of the multifunctionality of the soil resource in any discussion about the use of natural resources and value-driven trade-offs in resource management discussions. Examples of synchronic and short-term resource management issues are taxation of agricultural lands and fair value in land swaps, e.g., in land consolidation programs. A diachronic longer-term example is the capability of soils to perform under different land uses and different management intensities, compared to one-off uses such as foundation for construction (urbanization) or as a mineable resource.

We proceed as follows. First, we define the concept of value and in what terms it can be measured. Second, we describe pedometric approaches to internal valuation, i.e., comparing soils to each other, such as land indices. Third, we link the concept of external value, i.e., comparing the value of the soil resource to other goods in monetary terms, to FAO-style land evaluation. We then expand the discussion of value by describing the multiple contributions of the soil resource to human well-being via the concepts of natural capital and ecosystem services. Finally, we propose an approach to measuring the value of the soil using these concepts.

17.1 Concepts of Value

The noun "value" has a number of meanings in both common and technical English. The simplest definition is "quality of an object that permits measurability and therefore comparability" (Robertson 2012). In this sense soils can be described by any number of measureable attributes such as effective rooting depth or available water capacity, and pedometrics provides tools for quantifying these. However, in this chapter we consider "value" in the broader sense of the word, namely, value as a suitably defined utility.

The Concise Oxford Dictionary gives the fundamental definition of the term "value" as "worth, desirability, utility, [and] qualities on which these depend," which begs the question "worth, etc. to whom?" If to humans viewed as economic actors, this is economic value, in the sense of the neoclassical economic theory of value. If to humans viewed as social animals, this is sociocultural value. If to ecological systems through interactions between its components, this is ecological value, in the sense of the ecological theory of interactions. If to humans viewed as economic actors through the contribution of natural ecosystems to the value of final economic

goods and services, this is contributory value (Ulanowicz 1991), in the sense of the ecological economic theory of value as explained below. The same object will have different values depending on which framework is chosen.

All of these concepts of value are explicitly anthropocentric, which only says that humans are the ones to assign value, including value to other organisms, the ecosystem, or even the planet as a whole. We humans are the only ones in this conversation, and we can choose to include or exclude what we perceive to be value to other actors.

There is also a verb "to value," which has several meanings, including "to appreciate, prize," but in the sense of the noun "value" as used here, it means "to assign a value [the noun] to something."

There are two types of measures of value: internal and external with respect to the resource being assessed. Internal values use a ratio scale, i.e., with a natural zero representing no value, and with units of equal value, to compare two or more objects of the same type, in our case equal areas of soil. The scale is defined so that a given change in a measure attribute of the object represents a specific change in utility. This relation is not necessarily linear. By contrast, external values use a ratio scale to compare a unit area of soil to any other goods. External measures must use a scale that is commensurate with other "valuable" things. The obvious external measure of value is money. McBratney et al. (2014) state that only by placing a monetary value on something is it possible to include that object in an accounting procedure. Hewitt et al. (2015) respond that the importance of soil can be quantified in other terms, e.g., as contributors to ecosystem services provision, for use in policy discussions, without assigning a monetary value. We return later to this discussion.

In the second half of the twentieth century, some economists began to analyze environmental problems in economic terms in order to point out the dependence of human societies on natural ecosystems (de Groot 1992). They stressed that the undervaluation of the contributions of ecosystems to public welfare and economic growth was due in part to the fact that many of the critical nonmarketed contributions of ecosystems underpinning human economies were not adequately quantified in terms that could be compared to economic indicators (Braat and de Groot 2012; Costanza 1997). The resulting discipline of ecological economics (Costanza and Waigner 1991) sees global economies as a subsystem of the larger finite global ecosystem. Ecological economists question the sustainability of the current economy that does not internalize environmental impacts and does not see raw material and energy as finite resources. Ecological economics uses concepts from conventional neoclassical or welfare economics and expands them to include environmental impacts, ecological limits, finite natural resources, and issues of equity and scale as necessary requirements for increasing the sustainability of human activities (Martinez-Alier 2002). These economists have extensively discussed valuation (e.g., Farber et al. 2002; de Groot 2002; Gómez-Baggethun and Ruiz-Pérez 2011). Ecological economics emphasizes the interdependence of economic and social systems. Foundational concepts include natural capital and ecosystem services. These concepts will be extensively used in the following discussion and so are introduced next.

17.2 Ecosystem Approach to Value as Contributions to Well-Being

17.2.1 Natural Capital and Soil Stocks

Traditional methods to value soil, including land indices and FAO-style land evaluation (see below), are restricted to the value of soil as a medium of production, usually for agricultural production within locally important land use systems. The concepts of ecological economics are much broader and correspondingly more complex to apply in practice. In this section we define natural capital, in the following ecosystem services, and then we discuss how to use these concepts to value the soil resource.

Ecological economists define the concept of *natural capital* as "stocks of natural assets … that yield a flow of valuable ecosystem goods or services into the future" (Costanza and Daly 1992). Another definition is "the living and nonliving components of ecosystems – other than people and what they manufacture – that contribute to the generation of goods and services of value for people" (Guerry et al. 2015). The concept of natural capital comes from trying to frame the contribution to the economy of natural resources alongside manufactured capital (factories, buildings, tools), human capital (labor, skills), and social capital (education, culture, knowledge). This is explicitly an anthropocentric viewpoint to illustrate the dependence of human societies on natural ecosystems.

Natural capital can be separated into renewable natural capital (RNC) and nonrenewable natural capital (NNC). Ecological economists hold that sustainable economic activity is based on sustainable income coming from all capital types; this is termed sustainable ecosystem services provision, which requires constant natural capital (Costanza and Daly 1992). In agricultural systems as land use intensity increases, NNC is often reduced (e.g., topsoil is lost; K reserves are depleted), so that ecosystem services flows, and therefore incomes, decrease in the absence of investment in RNC. To keep income constant, total natural capital needs to be maintained, which requires that some of the income coming from nonrenewable resources be reinvested into renewable natural capital, for example, adding mineral nutrients from external fertilizer sources to replace those removed.

Natural capital stocks can be increased or decreased by management. For example, the fertilization strategies in the Brazilian Cerrado (Goedert 1983) use large doses of lime to neutralize the pH-dependent charge of these highly weathered soils dominated by Al; this also reduces P-fixation capacity so that P fertilizers are effective. These effects last for several years, and at large enough doses, the subsoil becomes semipermanently changed. Thus, the natural capital for crop production is enhanced. On the other hand, Noble et al. (2000) present a sobering example of permanent land degradation after 37 years of continuous cropping on an acrisol in northeast Thailand: organic matter decreased dramatically; the exchange complex became almost saturated with Al, which was backed by a large Al reserve from soil minerals. The cation exchange capacity was so reduced that cations released during

mineralization of organic matter could not compete with Al and thus were leached out of the reach of plant roots. The natural capital for crop production was severely depleted.

The concept of changing natural capital by management is close to the concepts of McBratney et al. (2014) of *soil condition* relative to a reference state they call the *capability*: "an optimal capacity of the soil to which the current condition of the soil can be compared." Capability can be permanently reduced by erosion, soil sealing, or mining for raw materials. Less permanent changes in natural capital are changes in condition only, which may be reversed.

Soil natural capital stocks include inherent stocks that vary over long timescales (e.g., clay content) and manageable dynamic stocks that vary over short timescales (e.g., soil water content) (Dominati et al. 2010). Soil carbon stocks are familiar for their use in carbon inventory and as proxies for overall soil condition. Soil stocks also include nonmaterial soil properties such as energy (e.g., stored heat) and soil fabric (e.g., total porosity). These directly relate to the mass, energy, and organizational components of soil natural capital identified by Robinson et al. (2009). Topsoil stocks are generally more dynamic than subsoil stocks, since they respond more rapidly to management.

17.2.2 Ecosystem Services

An *ecosystem service* is defined as "the direct and indirect contributions of ecosystems to human wellbeing" (TEEB 2008). This concept is an explicitly anthropocentric viewpoint and focuses on what is important to humans. This is not as restrictive as it might first appear, because there is ample evidence that a healthy ecosystem benefits humans. The interconnectedness of the ecosystem means that some aspects that might at first seem to be unimportant for the provision of benefits fulfilling human needs, e.g., crop production, are in fact necessary components of the provisioning mechanisms. For example, soil organisms form a complex food web; the direct service of, e.g., waste recycling or pest population regulation cannot be isolated from this web, so that the health and biodiversity of the web have value. The concept of ecosystem services has led to a large literature, including an eponymous journal, to examine and extend it.

Ecosystem services can be grouped into three types: provisioning, regulating, and cultural (TEEB 2008; Dominati et al. 2010). These groups contain specific services that can be quantified for a specific land use. Provisioning services include provision of physical support; of food, wood, fiber, and other plant and animal products; and of raw materials. Regulating services include flood mitigation, regulation of greenhouse gases, and control of pest populations. Cultural services include landscape preservation. Figure 17.1 is taken from Dominati et al. (2010) and illustrates the relation between natural capital, ecosystem services, and human needs.

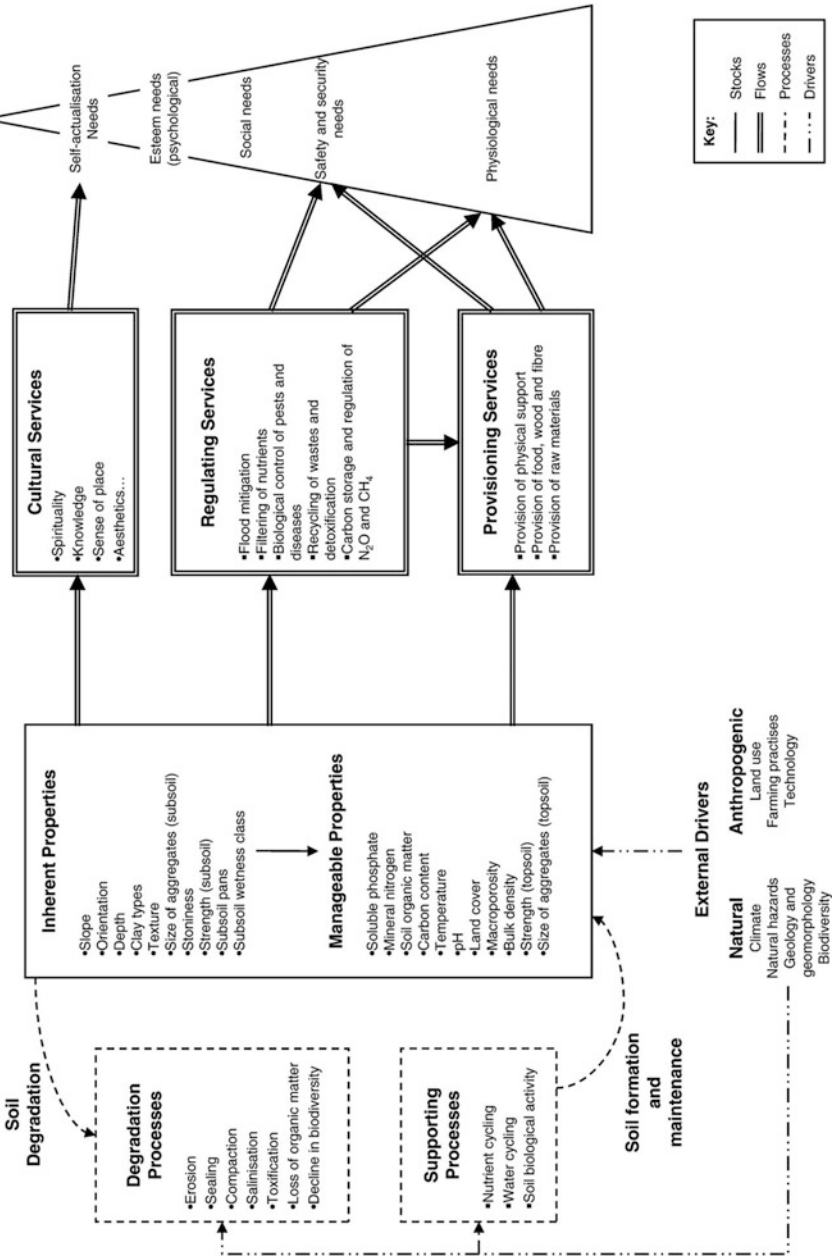


Fig. 17.1 Framework for the provision of ecosystem services from soil natural capital (From Dominati et al. 2010, Figure 2)

Earlier literature such as MEA (2005) and Haygarth and Ritz (2009) also refers to another category: supporting services. These are the ecosystem processes which support the provision of other services. These include primary production, soil formation, and nutrient cycling. These processes, not being directly of use to humans, are not referred to as services any longer to avoid double counting (Boyd and Banzhaf 2007).

The provision of ecosystem services, from a combination of land type and land use, can be quantified using information coming from a range of disciplines from ecology, to soil science, agronomy, or social science, as long as appropriate metrics have been defined for each service. These quantifications can then be used to put economic values on ecosystem services.

Economic valuation in monetary terms of environmental goods and services provides quantified information that can be used in a benefit-cost analysis (BCA). It has been extensively used for resource management and decision-making. The aim is to promote sustainable development by ensuring that policies fully account for the costs and benefits of development proposals on the natural environment. For policy-making in the context of agroecosystems and resource management, the more relevant application of economic valuation is to compare management options or assess an investment in either built (e.g., irrigation) or ecological (e.g., soil conservation) infrastructure that increase natural capital.

BCA is usually based on existing information, e.g., prices, as well as assumptions about future prices, regulations, and discount rates, and thus is not as objective as one would hope. However, failure to include ecosystem services in benefit-cost calculations implicitly assigns them a value of zero. To date this has been the norm, and this lack of accounting is accused of contributing to the depletion of natural capital stocks and increasing environmental problems (MEA 2005).

Traditionally the only soil contributions to ecosystem services that have been valued are those related to food provision, since that can be approximated by a farm budget. The other soil contributions to services provision are either unknown or considered too difficult to value. We propose to include all services in the valuation approach proposed here, as recommended by Dominati et al. (2016). This is intended to provide a holistic representation of the sustainability of land uses and the pivotal role the pedosphere plays in human well-being.

17.3 Soil vs. Land

Our aim is to value the soil resource. However, the soil is only one aspect of land as defined by the FAO (1976): “an area of the earth’s surface, the characteristics of which embrace all reasonably stable, or predictably cyclic, attributes of the biosphere vertically above and below this area including those of the atmosphere, the soil and underlying geology, the hydrology, the plant and animal populations, and the results of past and present human activity, to the extent that these attributes

exert a significant influence on present and future uses of the land by [humans]." Further, each land area has a definite location, which greatly influences its use and therefore value.

Soils occur at definite landscape positions with combinations of soil-forming factors, notably climate and relief, not independently of these land factors. Therefore, we propose to use the term "soil" to refer to the land resource including those soil-forming factors that are intimately bound to pedogenesis as discussed in Part VI of this book. Further, we accept the concept of USDA soil taxonomy (Soil Survey Staff 1999), which considers soil moisture and temperature regimes as soil characteristics. We exclude certain aspects of land, notably location, legal status, and water resources other than water that naturally affects the soil profile, and details of atmospheric climate that are not reflected in the more general concept of moisture and temperature regimes. For example, access to irrigation water affects land value, but the soil value for irrigated land uses only includes properties such as infiltration rate, water holding capacity, effective rooting depth, and soluble salts.

One method to disaggregate the effect of location from other land characteristics, including soil properties, is hedonic valuation within a single economic and market context (Rosen 1974). The hedonic hypothesis is that the value of an object (e.g., an area of soil) is equivalent to the utility received from its use, as revealed by differential prices associated with different levels of the attributes (e.g., rooting volume, available water capacity) that influence utility. Hedonic valuation estimates differences in value based on a set of attributes, typically in a spatial auto-regressive regression (SAR) formulation (Anselin and Bera 1998). The coefficients of the fitted model are interpreted as the elasticity of value due to changes in that attribute. Samarasinghe and Greenhalgh (2013) used this method to compare land values among 4477 farms in $a \approx 6000$ ha catchment in New Zealand. The type of current land use, distance to towns and roads, administrative unit, and their interactions all affected land value, as was expected. Soil characteristics including profile available water, rooting depth, gravel class, and drainage class also affected land value. Their model is a typical SAR model (Eq. 17.1):

$$\ln\ V_i = \beta_0 + \beta_S S_i + \beta_A A_i + \beta_C C_i + \beta_T T_i + \beta_G G_i + \beta_L L_i + \lambda W_i u_i + \varepsilon_i \quad (17.1)$$

where V is the value (as measured) at site i, other capital letters represent factors that influence land price (Soil, surface Area, Climate, Topography, Geography, Location), β are the weights to be estimated, W is a spatial weights matrix which accounts for spatial autocorrelation in the model residuals, λ is the strength of the autocorrelation, and ε is the random error ("noise"). The partial derivative for S gives the hedonic value of the chosen soil property. This is most easily estimated when the other factors are constant, but the equation can be solved even if every observation has different combinations of factors. Since we consider soil climate and landscape position to be part of the definition of "soil," we would remove T and C and include them under S in this equation.

17.4 Pedometric Approaches to Internal Valuation

In internal valuation the aim is to establish a ratio scale by which soils used for a given purpose may be compared, rather than to establish a scale by which soils may be compared with other goods. This has historically been motivated by the need for equitable taxation, compensation for taking, and land swaps for agricultural land, not considering location or water resources, but usually considering permanent land improvements such as drainage or irrigation works. These valuations have used some kind of quantitative *land index* on a ratio scale (typically 0–100), which can then be used for direct comparison in land swaps or converted to valuation for taxation by their proportion of the total tax base. These approaches are pedometric in that they use measured soil properties, either on a continuous or classified scale, and relate these on a linear scale, so that a given change in each property has implicitly the same effect on the soil's value to the user at any position in the measurement scale, and in addition are comparable across properties. A land index could include non-soil land characteristics, e.g., climate and water resources, but the ones we examine in this section do not.

A well-known internal valuation is the Storie index from California (Storie 1933, 1978). This assigns points to attributes of the physical profile (factor A), the surface texture (B), the slope (C), and other factors (X) including drainage, acidity, alkalinity, nutrients, actual erosion, and microrelief. These are multiplied to reach a final rating. The Storie index and its derivatives are additive, multiplicative, or maximum-limitation indices, or some combination, as summarized by Riquier (1974). A number of soil properties that are known to influence agricultural production, e.g., rooting depth, stoniness, and presence of salts, are rated on a ratio scale, e.g., 0–100, and these ratings are combined by arithmetic or geometric averaging. This approach is also used in the former USSR, for example, the Ukraine (Starodubtsev et al. 2011), where it is called soil *bonitas*, from the Latin for "goodness"; this was likely derived from the German *Bonitierung* from the same Latin root. These indices have the following forms, depending on whether the combination is (possibly weighted by weights w_i) additive (see formula 17.2a), multiplicative (17.2b), or geometric (17.2c); in all cases the LC_i are the quantitative land characteristics normalized to a 0.1 range, from "worst" to "best." A Storie-like index S is then computed as

$$S = \sum_{i=1}^{q} \mathrm{LC}_i \cdot w_i \quad \text{where} \sum_{i=1}^{q} w_i = 1 \tag{17.2a}$$

$$S = \prod_{i=1}^{q} \mathrm{LC}_i \tag{17.2b}$$

$$S = \sqrt[q]{\prod_{i=1}^{q} \mathrm{LC}_i} \tag{17.2c}$$

In Germany, agricultural and grazing land is compared by the *Landwirtschaftliche Vergleichszahl* (LVZ), i.e., agricultural comparison number, based partly on the *Bodenschätzung*, i.e., soil valuation, based on a law of October 1934 (Rothkegel 1950). This attempts to separate the intrinsic productivity of the soil from other factors that influence agricultural land value, such as field size and location with respect to main farm buildings and markets.

A problem with land indices is how to assign points to attributes singly and collectively. One approach has been to compare agricultural productivity on each soil, compared to some standard assumed to be without limitation. This implicitly assumes a reference land use system that could in principle be applied on all land areas. For example, in the German system all agricultural soils are compared to the most productive soils in Germany, a Chernozem at a reference location (Eickendorf in the Magdeburger Börde, Saxony-Anhalt), which is assigned full marks (100). Lower marks are estimates of the proportional productivity of common field crops under common management practices. Proportional yields are established by the long-term yields of reference crops at experimental plots (*Musterstücke*) spread throughout the country. Not all soils can be tested this way, so the characteristics of these sites are used to construct a point system that can be applied to any soil, based on its measured characteristics. For agriculture, the "agriculture number" (*Ackerzahl*, abbreviation AZ) is based on general soil texture (*Bodenart*), general type of parent material (*Entstehungsart*), and topsoil thickness (*Zustandsstufe*). For grassland, the "pasture number" (*Grünlandzahl*, abbreviation GZ) is based on general soil texture, topsoil thickness, mean annual air temperature, and drainage conditions (*Wasserverhältnisse*). In both cases, the number is supposed to be a proportional yield and thus is considered a fair basis for taxation or land exchange. If all parties accept the point system as fair, it is de facto the correct valuation for its purpose.

Another way to establish a point system is by regression analysis. For example, Olson and Olson (1986) used multiple regression analysis to estimate average maize yields in New York State (USA) using soil and climatic data, maize being a key indicator crop in that state; this work was later extended to Illinois (Olson et al. 2001). Their rationale was that "states [in the USA] need reasonable yield estimates to determine land value using an income capitalization approach to value for land appraisal and taxation ... our method is an improvement over the existing procedures based primarily on the collective judgment of experts." In New York the yield prediction was based on storage of available water (to counteract short droughts at key periods such as silking) and drainage (to avoid damage from surface ponding); the proportional predicted yields based on these were converted to a relative point system via the multiple regression equation. In this case atmospheric climate did not vary much within the study area (adjusted $R^2 = 0.13$ using only in-season rainfall), so most of the successful model fit (adjusted $R^2 = 0.80$) is attributed to the soil.

For geographically distributed observations, e.g., the reference crop yields and soil properties of the previous example, we cannot assume independence of the linear model residuals. Instead, the model formulation is

$$\mathbf{y} = \mathbf{X}\beta + \eta, \eta \sim \mathcal{N}(0, \mathbf{V}) \tag{17.3}$$

where $\mathbf{y}$ is the response vector, $\mathbf{X}$ is the model design matrix, and $\mathbf{V}$ is a positive-definite variance-covariance matrix of the model residuals, estimated by a covariance function of the separation between observations (see Sect. 3.2). This mixed model (fixed effects, i.e., regression parameters β, and random effects η, i.e., covariance of residuals) can be solved by restricted maximum likelihood (REML) (Lark and Cullis 2004).

Land indices are related to so-called soil quality indices, developed over the past 25 years (Karlen et al. 2003; de Paul Obade and Lal 2016) and the more recent concept of *soil health* (Moebius-Clune et al. 2016), which select soil properties though to be related to sustainable agricultural production and create indices based on their values, especially to highlight the effect of changes in management. However, these indices are not intended as measures of value, rather as measures of soil condition.

A modification of these approaches is to use fuzzy variables, rather than single values, to represent soil properties at a location and fuzzy logic using semantic import functions, rather than Boolean combinations, to combine them (Burrough 1989). This procedure results in a possibility of each rating, i.e., the combination does not give a single value, but rather a fuzzy variable. This avoids the problem of crisp thresholds. Several land evaluations since Burrough's paper have used this approach; a typical example is by Bagherzadeh and Gholizadeh (2016). Burrough gave an example of an FAO-style land evaluation for maize in Kenya, using six land qualities, each with asymmetric membership functions:

$$\begin{aligned} \mu_A(x) &= 1; x \geq c \\ \mu_A(x) &= \left(1 + a(x-c)^2\right)^{-1}; x < c \end{aligned} \tag{17.4}$$

where $\mu_A(x)$ is the membership for an area or site A on 0.1 for the value x of the land quality (here, an ordered class number); a controls the shape of the function, in particular the crossover value where the membership is 0.5 (equally possible and not possible); and c is the land quality value at the central concept, i.e., maximum suitability. The individual memberships were combined with a convex combination (weighted sum) operator to obtain an overall suitability rating μ_A:

$$\mu_A = \sum_j w_j \cdot \mu_{A_j}; \sum_j w_j = 1; w_j > 0 \tag{17.5}$$

See Sect. 5.5.1.1 for more details of fuzzy sets and fuzzy logic.

The first difficulty with any land index is the calibration of the scoring method. In many cases this is based only on expert opinion, due to the lack of sufficient experiments or observations. More fundamentally, these approaches do not consider the land use or management system, except in the most general terms. In the New York example, the indicator crop (maize) is typically produced by best practices on each soil (in particular, fertilization is based on soil testing), so that the land use and management system is implicit. The crop is widely grown and accepted by growers as an indication of agricultural land value, even for fields where it is not currently grown. However, even in New York there is a diversity of management (conventional tillage, no-till, residue management, organic amendments) that is not considered in this relation. This is clearly not appropriate in a more diversified environment and in addition only considers one ecosystem service.

These methods are also purely empirical, i.e., from calibration observations of system productivity to either expert judgment or parametric combinations. Beginning in the 1980s, many semi-physical models of the soil-plant-atmosphere system were built, including the Wageningen models WOFOST and its descendants (van Ittersum et al. 2003), CropSyst (Stockle et al. 2003), and the DSSAT suite (Jones et al. 2003). The supposed advantage is that these models can simulate the effects of changing conditions, including management interventions, based on physical principles. In practice they require extensive parameterization for each location and a very large minimum data set and in addition contain many empirical factors within each "physical" process. They are well suited for understanding the soil-plant-atmosphere-management system, e.g., the response of a system to changing conditions, but much less so for prediction to an accuracy that could be used as a measure of value.

17.5 Simple Pedometric Approaches to External Valuation

External valuation attempts to value the soil in relation to other goods. In this section we examine some cases where this can be fairly easily accomplished, if the soil is only considered a mineable or productive resource, i.e., a marketable good. In the following section, we expand the view of external valuation to ecological economics.

17.5.1 External Valuation of the Soil as a Mineable Resource

Soil can be used as a directly marketable raw material, i.e., a resource to be mined for its bulk properties or constituents. Portions of the profile can be removed for potting mix or for landscaping. In addition, constituents can be separated out from the whole soil, for example, gravel for roads, clay for bricks, or bauxite ore for aluminum. Such uses inevitably destroy some or all of the original resource, in favor

of some immediate economic benefit. As such, external valuation is straightforward: the market value of the mined resource net of direct costs including transportation and any land reclamation. This is the minimum market value of the soil. The value of leaving the soil in place is much greater because of the soil's contribution to the provision of ecosystem services within an ecosystem for a range of potential uses.

Unless the soil is completely removed to bedrock, the degraded soil, i.e., what is left after mining, still contributes to the provision of ecosystem services but to a lesser extent. Dominati et al. (2014a) quantified and valued the services provided by a soil under pastoral agriculture, before and after erosion. The same could be realized for a soil before and after partial extraction of material: the difference in value of services associated with the removed soil material is an expression of the value of that material, in place, under a use.

If soil materials are relocated, e.g., for landscaping, the relocation process affects the services provided both where the soil is removed (loss of natural capital) and where the material is added (gain of natural capital). If substantial portions of the soil profiles are relocated, they are usually mixed during excavation, transport, and relocation. Soil functions rely on the character of the soil horizon sequence and may be changed when soil is disturbed. For example, the morphology that controls soil water dynamics may be radically changed by mixing, bulking, compaction, and by the effects on soil biological and soil organic matter processes. Relocation changes also include the incorporation of foreign materials, as in bricks and mortar, weeds and pests, and the litter of the Anthropocene. It is possible to compare total soil functional status before and after relocation as was done by Dominati et al. (2014b) in the case of erosion. The net market price should account for the extra functionality from the point of view of the landowner where the relocated soil is placed.

17.5.2 External Valuation by FAO-Style Land Evaluation

Land evaluation is the process of predicting the use potential of land on the basis of its attributes (Rossiter 1996). The FAO land evaluation framework (FAO 1976) and subsequent guidelines (e.g., FAO 1985) were important advances over land indices. One of the most important innovations was to make the land use system central to the evaluation. Thus, there is no single index of value, rather a rating for each actual or projected land use system. These may be combined in various ways to produce a single index.

In the FAO method, the evaluator specifies a set of land use systems ("land utilization types," abbreviation LUTs) that may be imposed on a land area and evaluates the suitability of each area for each use. This can be done in terms of limitations (so-called physical evaluations) or in financial terms, by linking limitations to reduced outputs, increased inputs to compensate, or longer term to production (Rossiter 1995). The soil is valued as part of the land, but not considering location; thus outputs and inputs are costed at the farm gate.

The external valuation procedure is conceptually simple but complicated in practice (Rossiter 1995):

1. Select representative LUTs that are feasible in a given socioeconomic-political context.
2. Describe these by their land use requirements (LUR), together with the financial effects of less-than-optimum LUR (lower yields, higher production costs, or longer time to product).
3. Build a model to evaluate the level of output(s) from each LUT, based on levels of the LUR.
4. Build models to evaluate each LUR from measurable land characteristics (LC).
5. Describe the LC of each land area to be evaluated, the land mapping unit (LMU).
6. For each LUT, apply its LUR models to each LMU.
7. For each LUT, combine the results of the LUR models in the LUT model.
8. Compute the financial balance for each LMU and LUT.
9. Compute financial indicators for each LMU, combining all the LUT results.

These combinations express the external value in whatever financial terms are considered most relevant. This could be the maximum return from any LUT ("highest use"), the average of all relevant LUTs ("versatility") or their standard deviation ("security"), or some weighted combination. Already by 1990 the ALES computer program implemented a sophisticated set of methods for these computations (Rossiter 1990), in which yields, time to production, and input amounts can be tied to severity levels of land qualities within a set of contrasting LUT.

This procedure can be applied to LUT where the return is from annual crops, in which case gross margin is the financial measure. It can also be applied to multiyear LUT, in which case the present value of a discounted cash flow is the financial measure; here a discount rate must be selected. These evaluation results can then be used in further planning procedures such as linear programming to reach a constrained optimum in a planning area (FAO 1993). They can also be combined to give a final value to each LMU. This can simply be the value of the maximally valuable LUT ("best use") or some statistic of the returns for the selected LUT, e.g., average (expected value over time) or standard deviation (as a measure of risk). These are suitable for external valuation in the current context, e.g., for land taxation, land taking, or land reallocation.

This approach is useful for comparing land areas in the short to medium term, and with only one ecosystem service in mind, e.g., food and fiber provision, along with any immediate devaluation of ecosystem services resulting from this primary goal, e.g., monetized externalities. It has several problems that must be addressed before it can be used to value the soil resource.

A first problem is that economic land evaluation is for the short to medium term, but land evaluation for multiple outcomes that preserve the soil resource needs to be long term. A net present value approach rapidly discounts the future.

A second problem is the selection of LUTs. In the short term, this can be based on project objectives or current options within the constraints under which

the producer operates (Bouma 2001). However, when considering the value of the soil resource over the long term, it is not possible to anticipate all uses in the face of changing environments, both natural and especially socioeconomic-political; these are known as "option values." For example, an area currently used for crop production, and thus valued for a certain set of LUTs, may later be reassigned for groundwater supply or biodiversity conservation, with completely different LUTs. In addition, new technologies that cannot be anticipated may radically change land use requirements. Examples from the past are LUT based on conservation tillage (minimum disturbance of the soil) or even, if we go back far enough, LUT based on large quantities of manufactured fertilizers. Since the development of the Haber-Bosch process, large extents of sandy soils in northern Europe are now feasible for intensive crop production (which has also led to nitrate pollution of groundwater and streams), which would have been impossible with only organic inputs; this could not have been foreseen prior to 1913.

17.6 Ecological Economic Approaches to External Valuation

This section discusses how to use the concepts of ecosystem services, as provided by natural capital including soil, as the basis for external valuation as an extension of FAO-style land evaluation, as well as fundamental difficulties with this approach.

17.6.1 FAO-Style Land Evaluation and the Ecosystem Service Approach

The value reported by economic land evaluation may be purely financial, e.g., the cash flow to the land user from producing and selling agricultural products. However, it may be economic, including outputs with fictional (nominal) values, for example, the value of preserving an agricultural landscape, estimated by hedonic pricing, and outputs with negative value (externalities), for example, sediment or pollutants delivered off-site. In an economic analysis, shadow prices, representing the opportunity cost to society, may also be used (Huhtala and Marklund 2008).

Two trends emerging from reevaluation of the FAO land evaluation frameworks (FAO 2007) are the recognition of the wider functions and services provided by landscapes and the need for greater stakeholder participation in exploring the balance between economic, environmental, social, and cultural outcomes. Adding an ecosystem services approach to land evaluation would enable the supply of ecosystem services to be directly linked to the performance of a combination of land, land use, and management to deliver outcomes identified by stakeholders. This would provide a more complete picture of the efficiency of use of the natural

resources, assist in defining natural ecosystem boundaries, and give quantitative information on the progress toward economic, environmental, social, and cultural outcomes desired by community (Dominati et al. 2016).

Ecosystem services provision cannot be defined abstractly – this is where a key insight of the FAO Framework comes in. It arises from the combination of land type and land use. As a simple example, consider a reserved area for groundwater recharge to an aquifer used for municipal water supply; this is a common practice in the central Netherlands. The main contribution of soil to ecosystem service here is the quantity and quality of precipitation reaching the groundwater table. However, if the same area is reserved for a production forest, the main contribution of soil to ecosystem services is the supply of water and nutrients to growing trees, as well as physical support. A deep sandy low-humus soil might be ideal for the first use: precipitation filters through a well-aerated sand column with no addition of impurities and with little loss due to evapotranspiration. The same soil is unsuited for the second purpose: the water quickly flows through the root zone leading to drought, and there are few nutrients and little nutrient retention.

17.6.2 Conceptual Framework for Valuation by Ecosystem Services

Dominati et al. (2010) proposed a conceptual framework that places soils into the context of wider ecosystems and human well-being, by showing how soils as natural capital are contributing to the provision of ecosystem services under a use. They proposed an approach to fairly represent soils when quantifying and valuing ecosystem services provision under any land use, in particular from agroecosystems. Their framework considers soil as a form of natural capital, i.e., semipermanent stocks, which provide ecosystem services under a use.

The framework consists of five interconnected components: (1) soil as a form of natural capital stocks; (2) the processes of formation, maintenance, and degradation of this natural capital; (3) the external drivers affecting these processes and thereby natural capital; (4) the provisioning, regulating, and cultural ecosystem services flowing from natural capital stocks under a use; and (5) the human needs fulfilled by these ecosystem services. Depending on which land use the soil resource is under, the services and their level of supply will vary.

This framework requires the analyst to specify one or more land use systems and some way to value each ecosystem service if economic valuation is desired. Samarasinghe et al. (2013) review several approaches to this.

Many ecosystem services benefit the land user (on-site services), but some have their primary benefit from off-site services. For example, the ecosystem service of food production, from the natural capital stock of soil nutrients and water, mainly benefits the farmer first. The ecosystem service of regulating stream and groundwater, from the natural capital stock of soil as a filter and buffer, benefits all

users of the water, as well as stream ecology. These externalities (from the point of view of the land user) are rarely attributed to the landowner, e.g., by a pollution tax, although they are sometimes controlled by regulations, which limit the profitability of a land use (Huhtala and Marklund 2008).

17.6.3 Problems with Economic Valuation of Soil Contributions to Ecosystem Services

There are two aspects to economic valuation of ecosystem services that present difficulties. The first is assigning a value to a level of service, and the second is transforming this into money within the economy. Baveye (2015), using the term "soil services" for what we call the soil contribution to each ecosystem service, outlines these problems succinctly in regard to soils:

> Whereas the ecosystem services idea has been espoused enthusiastically by soil scientists, etc., very little progress has been made to date on the monetization of soil services. This may be due to the fact that it is not straightforward to assign a price to features or processes one does not understand satisfactorily, or the slow progress might be related more to uncertainty and lack of trust about what financiers might do with prices associated to soil services. Nevertheless, significant pressure is currently exerted on soil scientists by national governments and international agencies to engage actively with the ecosystem services framework. The challenge for soil scientists is either to find ways to monetize soil services meaningfully or to demonstrate convincingly (and relatively rapidly) that there are alternative paths that can be followed to preserve soils without necessarily putting price tags on their services.

A first difficulty is that valuation in terms of money is by nature short term. Methods to account for the time value of money, such as net present value, rapidly discount the future. The soil resource, with the exception of its value as a mineable resource, is meant to last "forever" in human terms. In this sense ecological economists argue that there should be no discounting at all, thereby giving equal value to current and future generations. This is well explained by Robinson et al. (2014) as the distinction between *value* and *price*, perhaps taking their cue from Oscar Wilde ("Nowadays people know the price of everything and the value of nothing"). Although both may be measured in money, "[e]conomic value seeks to identify all the final use and non-use, market and nonmarket values, and will often be unrelated to the price that soil commands as a commodity. This is because price only reflects purchase for a single or limited number of uses, whereas economic value tries to identify a combined value for all uses," i.e., the price plus consumer surplus.

In this respect Robinson et al. (2014) attempted to use prices to estimate the value of the topsoil, by calculating its quasi-replacement price: what it would cost the landowner to replace the topsoil, were it to be lost to, for example, erosion. We use the qualifier "quasi" because this price does not account for rebuilding topsoil structure after the (hypothetical) event of purchasing and spreading topsoil. This

sets a floor for the value of the soil in place. In their calculation, to replace 30 cm of English topsoil would cost about \$110 k ha^{-1}. This is not a pedometric calculation, because it is just based on market prices for soil constituents. It does not consider how this is used and performance under that use.

But this is not the most serious problem. Fundamental to all attempts to value ecosystem services is the inescapable fact that markets, including virtual markets revealed by hedonic pricing, contingent valuation, or other indirect neoclassical methods, by definition are set by the present-day population. In so far as they value the future, this can be included, with a suitable discount rate, but humans can only understand relatively short-term outcomes, which is why net present value methods (time value of money) work well for short-term planning and (modestly) delayed gratification. But this does not extend to the services to future generations.

Lastly, Baveye (2015) points out that soil functioning is complex and therefore so are soil contributions to ecosystem services, making it impossible to disaggregate the economic values of ecosystem services to be attributed to specific soil properties. Changes in soil condition (McBratney et al. 2014) can be related to changes in ecosystem services provision and their value under a specific use, as shown by Dominati et al. (2014b). However, those relationships would be different for different land uses, so that any values obtained thus could not be generalized.

Thus, we conclude that a full monetary valuation of the soil resource in terms of its ecosystem services is impossible, in the sense that the future is inherently unknowable. However, it is possible to provide a monetary valuation to decision-makers operating in project (short-term) mode, i.e., a short time frame within which planning occurs. The study of Dominati et al. (2016) does just this for two soils under high-intensity dairy LUT. See Baveye et al. (2013) for an extended discussion of the selection of the time frame for valuation.

Therefore we propose to value the soil resource in terms of its contributions to all ecosystem services which can be envisioned at the site and defer the question as to how these can be expressed in monetary terms, and eventually monetized, to ecological economists. In any case we believe that any such monetization cannot represent the full value of soil over the long term.

17.7 Toward a Pedometric Valuation of the Soil Resource Based on Ecosystem Services

Dominati et al. (2016) argue that land evaluation, thus implicitly soil valuation, needs to be able to inform capability for multiple functions in order to recognize the whole range of ecosystem services provided by landscapes. In this sense, and summarizing the above discussion, we here present the outline of a method to value the soil in terms of its multiple contributions to ecosystem services provision, building on both the stock adequacy approach of Hewitt et al. (2015) and the indicator approach of Calzolari et al. (2016).

A pedometric valuation of the soil resource based on an ecosystem approach must first specify the geographic, political, and socioeconomic context within which it is carried out. This shows explicitly that any conclusions would have to be revisited if the context substantially changes. The first step, selection of LUT, depends completely on context, but so do subsequent steps. For example, if a new dam is to be built, soils upstream must now provide ecosystem services related to water quality and quantity for the reservoir (e.g., prevention of siltation to extend reservoir life or postpone dredging), whereas before this quality was for streams only. The soils now have no role in floodwater regulation downstream of the dam.

In addition, the valuation must be within a defined geographic area with a limited set of soil types. This is because the capability and condition (see below) are normalized to a 0–100 scale and thus require maximum and minimum values of soil properties, which only make sense within a defined area.

The method has the following steps, applied to each soil type to be valued:

1. Select a set of representative actual and potential land utilization types (LUTs), including management, which could be realized on this soil type. This list controls the list of ecosystem services (next step) and so should include enough uses to represent all feasible uses and their services. Note that this is not restricted to uses that are currently feasible, but also those that, should the socio-political-economic-technological context change, become feasible.

   The description of the LUT must be specific as it influences the choice of levels of soil properties required for ecosystem services provision and sustainable management. As in the FAO Framework, land management practices must be included in the definition of LUT.

   For example, high-intensity heavy-animal grazing requires specification of the soil's resistance to treading damage, as this is required for animal health, the health of the soil, and the quality of pasture production, whereas low-intensity grazing has minimal problems with such damage.

   As another example, low- or no-tillage annual field crop systems have less risk of erosion compared to clean-tillage systems. Thus, selection of levels of soil properties for ecosystem services related to surface water quality affected by agricultural runoff will be quite different between these LUTs.
2. For each LUT:

   (a) List the soil-mediated ecosystem services (provisioning, regulating, or cultural) to consider.
   (b) List the critical soil properties, that is, soil natural capital, influencing services provision. These are the *soil stocks*. These may be soil properties measured directly in the field, indirectly by proximal sensing, from laboratory analysis, or derived properties calculated using pedotransfer functions (see Sect. 6). Because of the difficulty of quantifying soil stocks, simpler methods using easily measured indicators may be used to get a first idea of the value of soil stocks (Robinson et al. 2012).

Each service is influenced by one or more soil properties. For example, filtering for groundwater provision is mediated by soil thickness, horizonation, and hydraulic conductivity; these all affect residence time of water in the soil column (Keesstra et al. 2012). It is also affected by the type and quantity of soil organisms.

(c) Determine *soil capability functions* for each service, that is, soil functions delivering levels of that service, in the context of the study area.

  (i) Specify an empirical relation between the identified critical soil properties and the service with pedotransfer functions or process models.
  (ii) For each critical property used in the empirical relation, standardize its range of values to a 0–1 scale relative to its contribution to the service. For example, a soil with rapid saturated hydraulic conductivity will be good (value approaching 1) for flood mitigation, but not good (value approaching 0) for contaminants filtration for groundwater supply. This can be a monotonic, optimal value range or even multimodal curve.
  (iii) Standardize the results of the empirical functions to a continuous 0–1 scale.

(d) For each service, determine the adequacy of the values taken by the soil capability function to the requirements of that land use and management, standardized to a continuous 0–1 scale. This can be done by using agronomic principles or by quantifying with models the provision of each ecosystem service at the extremes of the range of step (4), as well as some intermediate values, and then standardizing these values.

(e) Develop a suitability function, not necessarily linear, to combine the results of the individual services provisions into a single value for ecosystem services provision. This can be evaluated at all values of the soil capability functions.

3. Present suitability ratings for the selected LUTs, both for the individual (step (d)) and overall services (step (e)).

The result is a pedometric suitability rating for the services provision under each LUT carried out on each soil type for which it has been selected. The individual services ratings can then be used directly in multi-criteria evaluations or combined in optimization frameworks; this is beyond our pedometric scope.

Figure 17.2, from Dominati et al. (2016), shows the results of this method for a single LUT (a dairy-based agroecosystem) on two contrasting soil types. This spider diagram shows complete fulfillment of a service at the outer edge (1) and no fulfillment at the center (0). In this diagram there is no attempt to combine the separate services as in step 2(e) of our method; a multi-criteria weighting could be used for this. We clearly see that one soil, Horotiu, has much superior performance for P filtering but much worse than the other soil, Te Kowhai, for net C accumulation; how these are weighted is a social issue.

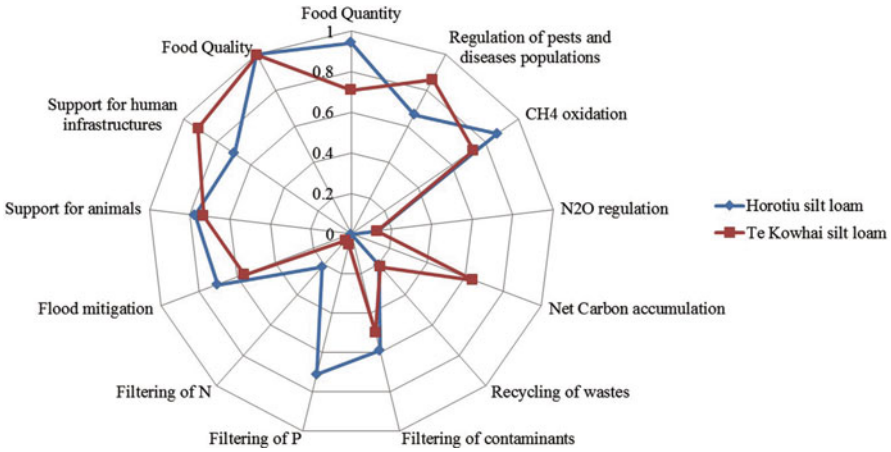


Fig. 17.2 Performance of ecosystem services delivery for two agroecosystems, based on biophysical measures (From Dominati et al. 2016, Figure 2)

17.8 Spatialization

The basic areal unit to be valued is the soil profile (Sect. 8), often called a pedon, in practice some small representative volume, e.g., 1 m^2 surface area to some depth (Hewitt et al. 2015) considered as the point support. The profile comprises a bundle of soil stocks that are the soil properties needed to drive the soil processes that result in soil functions that contribute to the provision of ecosystem services under a use. The profile is considered to represent a grid cell of some desired resolution or a sufficiently homogeneous map unit such as a consociation of a single soil series. The per-profile value can then be multiplied by the number of point support units in a map unit or parcel to give the value of the soil over that area. Conversely, quantities measured over larger land areas, e.g., crop production per hectare, can be normalized to the point support.

However, this concept of the areal unit to be valued assumes that (1) the area of soil to be valued is isolated and (2) ecosystem services are provided vertically (from atmosphere through soil to hydrosphere), so that mapping soil value would be a simple case of valuing at each small area. Of course the soil cover is more or less continuous, so grid cells at any resolution are connected to their neighbors, and many processes operate also horizontally, e.g., throughflow on hillsides and accumulation of materials at seeps and footslopes.

Soil functions may also depend on some minimum contiguous surface area of soils with similar function: a small patch of alluvial soil will hardly provide any floodwater regulation ecosystem service, whereas a floodplain commensurate with maximum overflows will provide this service.

Functions may also depend on some spatial pattern of contrasting soils with different processes. A well-known example is the production function for traditional

grazing of upland soils in the wet season and lowland meadows in the dry season. These two must be combined for successful year-round grazing.

This implies that for some ecosystem functions, we cannot evaluate its capability function for a grid cell or map unit in isolation; the adequacy of a functional unit (e.g., toposequence for stream recharge) must be evaluated as a unit and then allocated across the components of the functional unit.

17.9 Communicating Value

The valuation proposed in this chapter goes far beyond what the public naïvely consider to be the value of soil. If they consider it at all, they would most likely consider the relative value for agricultural production of typical adapted crops under typical current management. Many services are not appreciated, e.g., the regulating services for ground and surface water quality, and in addition these are rarely directly valued back to the soil that contributes to them. In a few high-profile cases such as the New York City water supply (Pires 2004), there may be some public awareness that the natural capital providing the services sustainably (e.g., water purification) is an alternate to high-cost built capital providing the same services. Simply reporting that "scientists" or "experts" have used a complicated method (no need to worry about the details) and have valued a soil at so many dollars per hectare is unlikely to satisfy the general public, let alone policy makers. The ecosystem services that were valued must be listed and their importance explained. Jobstvogt et al. (2014) faced a similar problem in communicating the value of another "hidden" resource, the deep sea. They chose the Ecosystem Principles Approach (Townsend et al. 2011), which presents the valuation in terms of general ecosystem principles that (it is hoped) are easily understandable, e.g., "[f]lora and fauna that filter food or nutrients from the water column and maintain a sedimentary lifestyle have a stabilizing effect on the sediment." For soils this would be something like "deep, well-aerated soils with good structure let rainwater infiltrate and pass at moderate speed through to groundwater."

17.10 Conclusions

Our proposed ecosystem services method has certain drawbacks. It is certainly laborious. A large number of LUT must typically be selected, because management options often have a major influence on ecosystem services types and levels of soil properties that affect them. Then, ecosystem services must be listed and their link to soil properties described. This may reveal a lack of understanding of soil processes or lack of computational models to quantify them. Even if this is successful, there may not be sufficient data on these properties in standard soil survey databases.

However, these deficiencies clearly motivate research into actual and potential LUT, ecosystem services provisions, and models. They also show gaps in soil properties in current databases and thus motivate new survey methods and soil resource inventory.

In the ecosystem services approach, we abandoned any attempt to place an external monetary value on the soil resource over the long term, let alone how to monetize soil services. If an external monetary value is needed, a FAO-style economic land evaluation or the approach of Dominati et al. (2016) can be used to compute a minimum value, under current conditions and for the short term; as we argued above, this cannot represent the long-term value of the soil resource. However, it is hoped that our proposed method can be used in each policy-making situation to provide a relative valuation that is commensurate with valuations of other resources, which can be used for decision-making. In any case, simply pointing out that soil has value and quantifying the many services it provides should raise public awareness of this resource, which is "hidden in plain sight."

References

Anselin L, Bera AK (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DEA (eds) Handbook of applied economic statistics. Marcel Dekker, New York, pp 237–289

Bagherzadeh A, Gholizadeh A (2016) Qualitative land suitability evaluation by parametric and fuzzy approaches for sugar beet crop in Sabzevar plain, northeast of Iran. Agric Res 5(3):277–284. doi:10.1007/s40003-016-0210-1

Baveye PC (2015) Grand challenges in the research on soil processes. Front Environ Sci 3:10. doi:10.3389/fenvs.2015.00010

Baveye PC, Baveye J, Gowdy J (2013) Monetary valuation of ecosystem services: it matters to get the timeline right. Ecol Econ 95:231–235. doi:10.1016/j.ecolecon.2013.09.009

Bouma J (2001) The role of soil science in the land use negotiation process. Soil Use Manag 17:1–6. doi:10.1111/j.1475-2743.2001.tb00001.x

Boyd J, Banzhaf S (2007) What are ecosystem services? The need for standardized environmental accounting units. Ecol Econ 63:616–626

Braat LC, de Groot R (2012) The ecosystem services agenda: bridging the worlds of natural science and economics, conservation and development, and public and private policy. Ecosyst Serv 1:4–15. doi:10.1016/j.ecoser.2012.07.011

Burrough PA (1989) Fuzzy mathematical methods for soil survey and land evaluation. J Soil Sci 40:477–492. doi:10.1111/j.1365-2389.1989.tb01290.x

Calzolari C, Ungaro F, Filippi N et al (2016) A methodological framework to assess the multiple contributions of soils to ecosystem services delivery at regional scale. Geoderma 261:190–203. doi:10.1016/j.geoderma.2015.07.013

Costanza R (ed) (1997) An introduction to ecological economics. St. Lucie Press; International Society for Ecological Economics, Boca Raton

Costanza R, Daly HE (1992) Natural capital and sustainable development. Conserv Biol 6:37–46. doi:10.1046/j.1523-1739.1992.610037.x

Costanza R, Wainger L (eds) (1991) Ecological economics: the science and management of sustainability. Columbia University Press, New York

de Groot RS (1992) Functions of nature : evaluation of nature in environmental planning, management and decision making. Wolters-Noordhoff BV, Groningen

de Groot RS (2002) A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol Econ 41:393–408. doi:10.1016/S0921-8009(02)00089-7

de Paul OV, Lal R (2016) Towards a standard technique for soil quality assessment. Geoderma 265:96–102. doi:10.1016/j.geoderma.2015.11.023

Dominati E, Patterson M, Mackay A (2010) A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol Econ 69:1858–1868. doi:10.1016/j.ecolecon.2010.05.002

Dominati E, Mackay A, Green S, Patterson M (2014a) A soil change-based methodology for the quantification and valuation of ecosystem services from agro-ecosystems: a case study of pastoral agriculture in New Zealand. Ecol Econ 100:119–129. doi:10.1016/j.ecolecon.2014.02.008

Dominati EJ, Mackay A, Lynch B, Heath N, Millner I (2014b) An ecosystem services approach to the quantification of shallow mass movement erosion and the value of soil conservation practices. Ecosyst Serv 9:204–215. Doi:http://dx.Doi.Org/10.1016/j.Ecoser.2014.06.006

Dominati EJ, Mackay AD, Bouma J, Green S (2016) An ecosystems approach to quantify soil performance for multiple outcomes: the future of land evaluation? Soil Sci Soc Am J 80:438–449. doi:10.2136/sssaj2015.07.0266

FAO (1976) A framework for land evaluation, FAO Soils Bulletin 32. Food and Agriculture Organization of the United Nations, Rome

FAO (1985) Guidelines: land evaluation for irrigated agriculture. Food and agriculture Organization of the United Nations, Rome

FAO (1993) Guidelines for land-use planning, FAO Development Series 1. Food and Agriculture Organization of the United Nations, Rome

FAO (2007) Land evaluation: towards a revised framework. Food and Agriculture Organization of the United Nations, Rome

Farber SC, Costanza R, Wilson MA (2002) Economic and ecological concepts for valuing ecosystem services. Ecol Econ 41:375–392. doi:10.1016/S0921-8009(02)00088-5

Goedert WJ (1983) Management of the Cerrado Soils of Brazil – a review. J Soil Sci 34:405–428. doi:10.1111/j.1365-2389.1983.tb01045.x

Gómez-Baggethun E, Ruiz-Pérez M (2011) Economic valuation and the commodification of ecosystem services. Prog Phys Geogr 35:613–628. doi:10.1177/0309133311421708

Guerry AD, Polasky S, Lubchenco J et al (2015) Natural capital and ecosystem services informing decisions: from promise to practice. PNAS 112:7348–7355. doi:10.1073/pnas.1503751112

Haygarth PM, Ritz K (2009) The future of soils and land use in the UK: soil systems for the provision of land-based ecosystem services. Land Use Policy 26:S187–S197. doi:10.1016/j.landusepol.2009.09.016

Hewitt A, Dominati E, Webb T, Cuthill T (2015) Soil natural capital quantification by the stock adequacy method. Geoderma 241:107–114. doi:10.1016/j.geoderma.2014.11.014

Huhtala A, Marklund P-O (2008) Stringency of environmental targets in animal agriculture: shedding light on policy with shadow prices. Eur Rev Agric Econ 35:193–217. doi:10.1093/erae/jbn025

Jobstvogt N, Townsend M, Witte U, Hanley N (2014) How can we identify and communicate the ecological value of deep-sea ecosystem services? PLoS One 9:e100646. doi:10.1371/journal.pone.0100646

Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, ... Ritchie JT (2003). The DSSAT cropping system model. Eur J Agron 18:235–265. doi: 10.1016/S1161-0301(02)00107-7

Karlen DL, Ditzler CA, Andrews SS (2003) Soil quality: why and how? Geoderma 114:145–156

Keesstra S, Geissen V, Mosse K et al (2012) Soil as a filter for groundwater quality. Curr Opin Environ Sustain 4:507–516. doi:10.1016/j.cosust.2012.10.007

Lark RM, Cullis BR (2004) Model based analysis using REML for inference from systematically sampled data on soil. Eur J Soil Sci 55:799–813

Martinez-Alier J (2002) The environmentalism of the poor: a study of ecological conflicts and valuation. Edward Elgar Publishing Ltd, Cheltenham

McBratney A, Field DJ, Koch A (2014) The dimensions of soil security. Geoderma 213:203–213. doi:10.1016/j.geoderma.2013.08.013

MEA – Millennium Ecosystem Assessment Program (2005) Ecosystems and human well-being: synthesis. Island Press, Washington, DC

Moebius-Clune BN, Moebius-Clune DJ, Gugino BK et al (2016) Comprehensive assessment of soil health – the Cornell framework manual, 3rd edn. Cornell University, School of Integrative Plant Sciences, Soil and Crop Sciences section, Ithaca. http://www.css.cornell.edu/extension/soil-health/manual.pdf

Noble AD, Gillman GP, Ruaysoongnern S (2000) A cation exchange index for assessing degradation of acid soil by further acidification under permanent agriculture in the tropics. Eur J Soil Sci 51:233–243. doi:10.1046/j.1365-2389.2000.00313.x

Olson KR, Olson GW (1986) Use of multiple regression analysis to estimate average corn yields using selected soils and climatic data. Agric Syst 20:105–120

Olson KR, Garcia-Paredes JD, Majchkrzah RN, Lang JM (2001) Equations for predicting grain crop yields of Illinois soils using soil properties. Soil Surv Horiz 42:52–64

Pires M (2004) Watershed protection for a world city: the case of New York. Land Use Policy 21:161–175. doi:10.1016/j.landusepol.2003.08.001

Riquier J (1974) A summary of parametric methods of soil and land evaluation. In: Approaches to land classification, Soils Bulletin 22. Food & Agriculture Organization of the United Nations, Rome

Robertson M (2012) Functions, services and values. Wetlandia. Retrieved August 3, 2017, from http://wetlandia.blogspot.com/2012/07/functions-services-and-values.html

Robinson DA, Lebron I, Vereecken H (2009) On the definition of the natural capital of soils: a framework for description, evaluation, and monitoring. Soil Sci Soc Am J 73:1904–1911. doi:10.2136/sssaj2008.0332

Robinson DA, Hockley N, Dominati E et al (2012) Natural capital, ecosystem services, and soil change: why soil science must embrace an ecosystems approach. Vadose Zone J. doi:10.2136/vzj2011.0051

Robinson DA, Fraser I, Dominati EJ et al (2014) On the value of soil resources in the context of natural capital and ecosystem service delivery. Soil Sci Soc Am J 78:685. doi:10.2136/sssaj2014.01.0017

Rosen S (1974) Hedonic prices and implicit markets: product differentiation in pure competition. J Polit Econ 82:34–55. doi:10.2307/1830899

Rossiter DG (1990) ALES: a framework for land evaluation using a microcomputer. Soil Use Manag 6:7–20. doi:10.1111/j.1475-2743.1990.tb00790.x

Rossiter DG (1995) Economic land evaluation: why and how. Soil Use Manag 11:132–140. doi:10.1111/j.1475-2743.1995.tb00511.x

Rossiter DG (1996) A theoretical framework for land evaluation (with discussion). Geoderma 72:165–202. doi:10.1016/0016-7061(96)00031-6

Rothkegel W (1950) Geschichtliche Entwicklung der Bodenbonitierungen, und Wesen und Bedeutung der deutschen Bodenschätzung. E. Ulmer, Stuttgart

Samarasinghe O, Greenhalgh S (2013) Valuing the soil natural capital: a New Zealand case study. Soil Res 51:278–287. doi:10.1071/SR12246

Samarasinghe O, Greenhalgh S, Vesely E-T (2013) Looking at soils through the natural capital and ecosystem services lens, Landcare Research Science Series No. 41. Manaaki Whenua Press, Lincoln. http://www.mwpress.co.nz/science-series/looking-at-soils-through-the-natural-capital-and-ecosystem-services-lens

Soil Survey Staff (1999) Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys, 2nd edn. US Department of Agriculture Soil Conservation Service, Washington, DC

Starodubtsev VM, Petrenko LR, Struk VS (2011) Soil evaluation – materials for lectures and seminars. AGRARMEDIA Group, Kiev

Stockle CO, Donatelli M, Nelson R (2003) CropSyst, a cropping systems simulation model. Eur J Agron 18:289–307. doi:10.1016/S1161-0301(02)00109-0

Storie RE (1933) An index for rating the agricultural value of soils. University of California Agricultural Experiment Station, Berkley

Storie RE (1978) The Storie Index soil rating revised. Special publication 3203. University of California, Division of Agricultural Sciences

TEEB (2008) The economics of ecosystems and biodiversity – an interim report. European Commission, Brussels

Townsend M, Thrush SF, Carbines MJ (2011) Simplifying the complex: an "Ecosystem Principles Approach" to goods and services management in marine coastal ecosystems. Mar Ecol Prog Ser 434:291–301. doi:10.3354/meps09118

Ulanowicz R (1991) Contributory values of ecosystem resources. In: Costanza R (ed) Ecological economics: the science and management of sustainability. Columbia Univ Press, New York

van Ittersum MK, Leffelaar PA, van Keulen H, Kropff MJ, Bastiaans L, Goudriaan J (2003) On approaches and applications of the Wageningen crop models. Eur J Agron 18:201–234. doi:10.1016/S1161-0301(02)00106-5

Part VI
Soil Genesis

"There is nothing in the whole of nature which is more important than or deserves as much attention as the soil. Truly it is the soil which makes the world a friendly environment for mankind. It is the soil which nourishes and provides for the whole of nature; the whole of creation depends on the soil which is the ultimate foundation of our existence."

Friedrich Albert Fallou, 1862

In previous chapters it was shown that soil variability is a function of several soil-forming factors. The pedometric methods explained to describe the soil (spatially), however, can only be used to infer about the present state of the soil resource. This chapter will focus on the concepts and models of soil formation that pedometricians have formalized over the years. Formalizing these has been driven by the need for understanding the complexity and dynamics of soil systems as soil plays a fundamental role in the functioning of ecosystems. Additionally, these efforts have also received increased attention by the need to model soil change through time; for being able to trace the history of soil systems and to predict their future state under differing environmental and climatic conditions.

The importance of soil systems and their evolution over time is reflected in the US National Research Council publication 'Landscapes on the Edge: New Horizons for Research on Earth's Surface' (NRC 2010) that addresses the challenges and opportunities in research on Earth surface processes: "Soil formation is not, however, only of academic interest. Our food comes from plants grown in soil. The rapid rate of soil erosion due to land use relative to the slow rate of transformation of rock into soil endangers soil resources worldwide. The fate of soils, the base of agriculture, is of great concern."

Soil is also recognized as one of the central components of the 'critical zone', a concept that stems from the earth sciences. The critical zone defines "the external terrestrial layer extending from the outer limits of vegetation down to and including the zone of groundwater" and most importantly "sustains most terrestrial life on the planet" (Brantley et al. 2006). Soil is produced through the transformation of bedrock and biomass and acts as an open system which is subject to element gains and losses. Understanding the forces responsible for the formation of this

key component of the critical zone is therefore crucial for being able to understand how physical, chemical and biological processes are working together at the Earth's surface (Anderson et al. 2007).

As discussed in Chap. 1, different types of models have been used in pedometrics to describe the soil resource. In the following, we will present pedogenetic models, starting with models of pedogenesis that describe soil formation qualitatively based on a description of soil evolution in the landscape or based on empirical observations, followed by quantitative, mechanistic models based on empirical equations or detailed differential equations derived from fundamental physics that require a profound understanding of pedogenic processes. These models operate on different spatial and temporal scales. It will also be discussed how processes of pedogenesis can be quantified with field data, to calibrate and validate models of soil genesis. We do not attempt to be exhaustive here; and refer the reader to extensive reviews written on this topic by Hoosbeek and Bryant (1992), Minasny et al. (2008), Stockmann et al. (2011) and Minasny et al. (2015).

References

Anderson SP, von Blanckenburg F, White AF (2007) Physical and chemical controls on the critical zone. Elements 3:315–319

Brantley SL, White TS, White AF, Sparks D, Richter D, Pregitzer K, Derry L, Chorover J, Chadwick O, April R, Anderson S, Amundson R (2006) Frontiers in exploration of the critical zone: a workshop sponsored by the National Science Foundation (NSF). Newark, p 30

Hoosbeek MR, Bryant RB (1992) Towards the quantitative modelling of pedogenesis – a review. Geoderma 55:183–210

Minasny B, McBratney AB, Salvador-Blanes S (2008) Quantitative models for pedogenesis – a review. Geoderma 144:140–157

Minasny B, Finke P, Stockmann U, Vanwalleghem T, McBratney AB (2015) Resolving the integral connection between pedogenesis and landscape evolution. Earth Sci Rev 150:102–120

NRC (2010) Landscapes on the edge. New horizons for research on Earth's surface. The National Academy Press, Washington, DC

Stockmann U, Minasny B, McBratney A (2011) Quantifying processes of pedogenesis. Adv Agron 113:1–74

Chapter 18
Clorpt Functions

Uta Stockmann, Budiman Minasny, and Alex. B. McBratney

> *"Essentially, all life depends upon the soil ... There can be no life without soil and no soil without life; they have evolved together".*
>
> Charles E. Kellogg

The extent of soil formation is dependent on local site characteristics. To model the evolution of soil in the landscape, we therefore need to know which factors and processes are important for describing pedogenesis quantitatively. There are several ways to go about it, and in the following the main approaches of how to formalize soil formation are therefore outlined.

18.1 Factorial

The Russian soil scientist Dokuchaev (1899) is known as one of the first who formalized an equation that linked soil (P) formation to local site characteristics including climate (K), organisms (O), ground or parent rock (G) and time (B):

$$P = f\,(K, O, G)\,B \tag{18.1}$$

Following on, the American soil scientist Shaw (1930) presented a factorial model that included major soil-forming factors which change with local conditions. However, Shaw did not develop his conceptual model any further before he died suddenly in 1930:

U. Stockmann (✉) • B. Minasny • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: uta.stockmann@sydney.edu.au; budiman.minasny@sydney.edu.au;
alex.mcbratney@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_18

$$S = M(C + V)^T + D \tag{18.2}$$

Arguably, the most well-known model of soil formation, however, is the state-factor model by the Swiss/American soil scientist Jenny (1941). In Jenny's model the state factors are independent from the soil system and vary in space and time:

$$S = f\,(cl, o, r, p, t, \ldots) \tag{18.3}$$

where cl is the climate, o are the organisms, r is the topography, p is the parent material and t is the time.

In its original form, the state-factor model is unsolvable. To be solved the indeterminate function f needs to be replaced by certain quantitative relationships. Hence, the *clorpt* equation has been formalized in quantitative ways based on empirical field observations, where a single factor is defined by keeping the other factors constant (Minasny et al. 2008). Empirical models were developed to describe soil formation in the form of quantitative climofunctions, biofunctions, topofunctions, lithofunctions and chronofunctions (refer to Sect. 19.1.1), mostly based on numerically intensive statistical methods (Yaalon 1975; McBratney et al. 2003). Based on the 'clorpt' model of soil formation, McBratney et al. (2003) formulated the *scorpan* model, which indeed applies empirical quantitative relationships to predict soil properties from landscape attributes at specific locations in the landscape. The *scorpan* model is written as:

$$\mathrm{Sc}/\mathrm{Sa} = f\,(s, c, o, r, p, a, n) \tag{18.4}$$

where *Sc* are the soil classes and *Sa* are the soil attributes, s is the soil, c is the climate, o are the organisms, r is the topography, p is the parent material, a is age and n is space or the spatial position. The model is used quite extensively in the field of digital soil mapping to predict the recent state of the soil (soil type as well as soil properties), but is not intended, and cannot be applied, for long-term soil formation predictions (refer to Chap. 12 of this book).

18.2 Processes

Simonson (1959) is known as one of the first soil scientists who conceptualized soil-forming processes by describing the formation of soil as a function of additions (e.g. organic matter), removals (e.g. soluble salts), translocations (e.g. soil organic carbon/sesquioxides) and transformations (e.g. transformation of primary minerals into secondary minerals):

$$s = f\,(\text{addition}, \text{removal}, \text{translocation}, \text{transformation}) \tag{18.5}$$

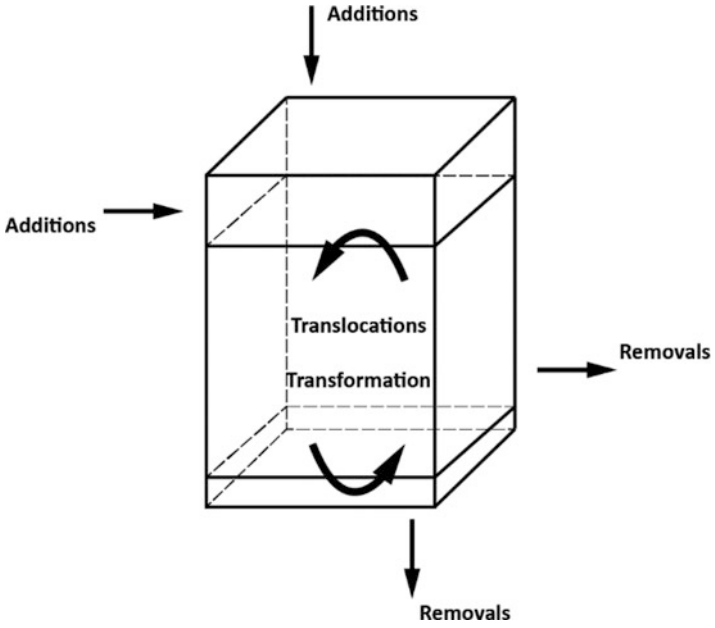


Fig. 18.1 Soil profile evolution as a function of additions, removals, translocations and transformations

Although Simonson's original work is still a qualitative description, his model can be seen as one of the first conceptual frameworks for mechanistic models of pedogenesis (Fig. 18.1).

18.3 Energy

There have been efforts to describe soil formation mechanistically based on the principles of energy or concepts of thermodynamics. The most well-known soil scientist who formulated a soil formation model based on energy is Runge (1973):

$$S = f\,(o, w, t) \tag{18.6}$$

where S is the soil, o is the organic matter production (renewing factor), w is the amount of water available for leaching (developing vector) and t is time. The soil-forming factors of climate and relief are expressed within the vector w. The energy model of Runge (1973) is only useful, however, in a qualitative way as it does not include actual quantitative thermodynamical equations (Hoosbeek and Bryant 1992).

Furthermore, Volubuyev published a variety of papers that linked pedogenic processes with energy laws. In one of his papers, he described the energy spent during soil formation through (Volobuyev 1974):

$$Q = \mathrm{Ra} = \mathrm{Re}^{-1/mK} \tag{18.7}$$

where Q is the expenditure of energy on soil formation, R is the energy of solar radiation, a are the available energy sources, K is the relative wetness and m is a factor expressing the participation of biota in energy exchange. Volobuyev and Ponomarev (1977) explored thermodynamic aspects of soil-forming processes by calculating Gibbs' free energy and entropy values of soil minerals. They were able to show that the thermodynamic characteristics of soil minerals do vary significantly for the soil types studied. Volobuyev (1984) also formulated an energy model to apply Dokuchaev's equation quantitatively (Eq. 18.1):

$$Q = R\,(\pm r)\exp\left[-\frac{P_c w R 0.67}{mP\,(\pm p)}\right] \tag{18.8}$$

where Q is the (annual) expense of energy on soil-forming processes, R is the radiant solar energy, P is the relative wetness, m is the biological activity, r is the radiation balance, p is the atmospheric precipitation, w (chemically bound water of mineral soil components) is the rate of mineral transformations in soils and P_c is water, such as water that is fixed in the mineral, faunal and floral component of soils. Dokuchaev's soil-forming factors, as demonstrated in Eq. 18.1, are represented by R and P and by p and r (climate, K), by m (organisms, O), by P_c and w (parent rock, G).

In Box 18.1 it is discussed if soil exhibits chaotic behaviour.

Box 18.1 Does Soil Exhibit Chaotic Behaviour?

Phillips (1993) argued that soil evolution can be deterministic chaos, where unpredictable behaviours can be obtained from a nonlinear model. He demonstrated that using simple ordinary differential equations, chaotic behaviour can be observed. If we consider soil (S) as a function of progressive (P) and regressive (R) processes with $S = f(P,R)$ (Johnson and Watson-Stegner 1987), then the mathematical form for the state of soil (S) at time t is

$$\mathrm{dS/dt} = \mathrm{dP/dt} – \mathrm{dR/dt} \tag{18.B.1}$$

The form of P and R can be exponential following first-order kinetics:

$$\begin{aligned} \mathrm{dP/dt} &= c_1 \exp(-k_1 S) \\ \mathrm{dR/dt} &= c_2 \exp(-k_1 S) \end{aligned} \tag{18.B.2}$$

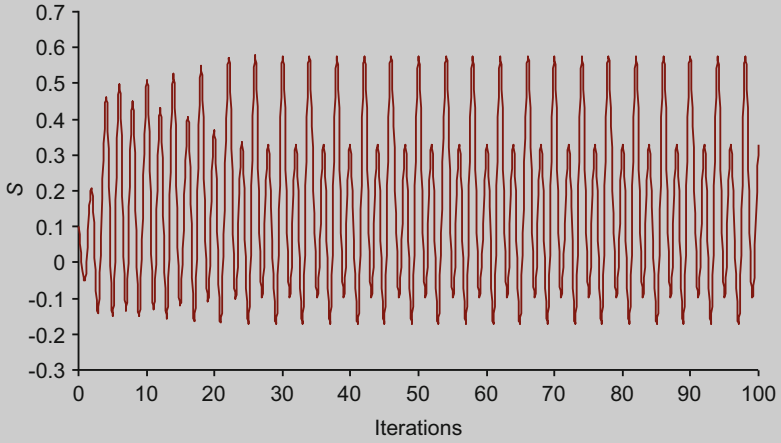
where c refers to some potential soil properties and k is the rate constant with conditions that c_1 should be greater than c_2 for soil and $k_1 > k_2$.

Using finite difference, the form of (18.B.1) and (18.B.2) becomes

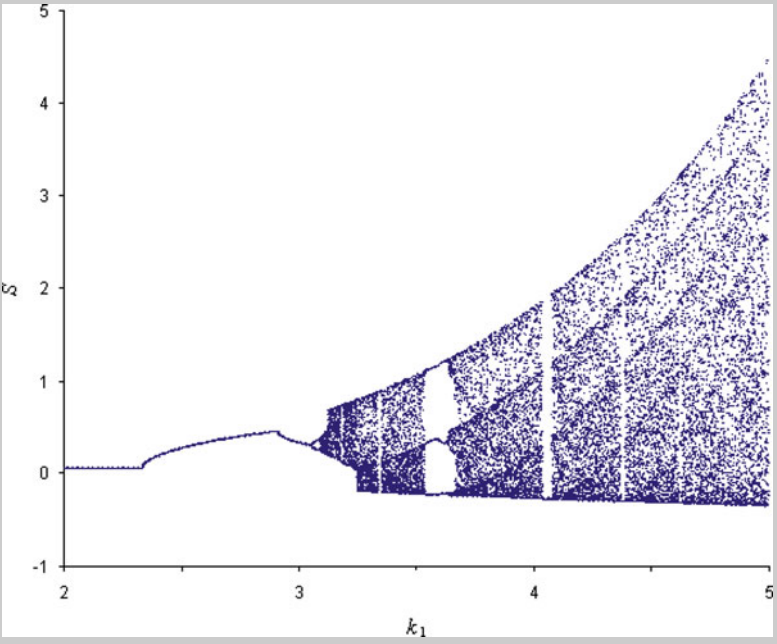
$$\mathrm{S_t} = \mathrm{S_{t-1}} + \Delta\mathrm{P} – \Delta\mathrm{R} \tag{18.B.3}$$

(continued)

It can be easily shown that this equation can produce chaotic behaviour, e.g. for $c_1 = 1$ and $c_2 = 0.9$, with $k_1 = 3$, and $k_2 = 0.1$.



Plotting S (after 100 iterations) as a function of k_1, where the other parameters are held constant, shows the famous bifurcation pattern indicating the instability of the solution. The plot shows that when parameter k_1 is greater than 3.5, the solution starts to show bifurcating results.



(continued)

A more complex model for soil organic matter dynamics and the possible chaotic behaviour can be found in Ryzhova (1996). Furthermore, Minasny and McBratney (2001) demonstrated the chaotic behaviour of a simple soil-landscape model.

The chaotic behaviour is of course dependent on the numerical stability of the differential equation. And in fact, chaotic solutions themselves can only be obtained by numerical solutions (Pachepsky 1998). We may need to ask whether the apparent chaotic behaviour is due to the dynamic laws of nature or is solely the result of an extreme sensitivity to numerical procedures.

References

Dokuchaev, VV (1899) Report to the transcaucasian statistical committee on soil taxonomy in general and especially for the Transcaucasia. Horizontal and vertical soil zones. Office Press of the Civilian Affairs Commander-in-Chief in the Caucasus, Tifilis (in Russian)

Hoosbeek MR, Bryant RB (1992) Towards the quantitative modelling of pedogenesis - A review. Geoderma 55:183–210

Jenny H (1941) Factors of soil formation. A system of quantitative pedology. McGraw-Hill Book Company, New York

Johnson DL, Watson-Stegner D (1987) Evolution model of pedogenesis. Soil Sci 143:349–366

McBratney AB, Mendonca Santos ML, Minasny B (2003) On digital soil mapping. Geoderma 117:3–52

Minasny B, McBratney AB (2001) A rudimentary mechanistic model for soil production and landscape development II; a two-dimensional model incorporating chemical weathering. Geoderma 103(1–2):161–179

Minasny B, McBratney AB, Salvador-Blanes S (2008) Quantitative models for pedogenesis – a review. Geoderma 144:140–157

Pachepsky Y (1998) Discussion of the paper by J.D. Phillips. Geoderma 86:23–42

Phillips JD (1993) Progressive and regressive pedogenesis and complex soil evolution. Quat Res 40:169–176

Runge ECA (1973) Soil development sequences and energy models. Soil Sci 115:183–193

Ryzhova IM (1996) Analysis of the feedback effects of ecosystems produced by changes in carbon-cycling parameters using mathematical models. Eurasian Soil Sci 28:44–52

Shaw CF (1930) Potent factors in soil formation. Ecology 11(XI):239–245

Simonson, R.W. (1959) Modern concepts of soil genesis. Outline of a generalized theory of soil genesis. Soil Sci Soc Proc, 152–156

Volobuyev VR (1974) Main concepts of ecology. Geoderma 12:27–33

Volobuyev VR (1984) Two key solutions of the energetics of soil formation. Soviet Soil Sci 16:1–8

Volobuyev VR, Ponomarev DG (1977) Some thermodynamic characteristics of the mineral associations of soils. translated from Pochvovedeniye 1977, No. 1, 3–13 1, 1–11

Yaalon DH (1975) Conceptual models in pedogenesis: can soil-forming functions be solved? Geoderma 14:189–205.

Chapter 19
One-, Two- and Three-Dimensional Pedogenetic Models

Uta Stockmann, Sebastien Salvador-Blanes, Tom Vanwalleghem, Budiman Minasny, and Alex. B. McBratney

"We know more about the movement of celestial bodies than about the soil underfoot".

Leonardo da Vinci, circa 1500s

19.1 Modelling and Quantifying Pedological Processes

Various methods have been used to measure or estimate pedogenic processes that are responsible for the differentiation of a soil profile. The most important pedogenic processes can be seen in Fig. 19.1, in a simplified form.

In the following, the modelling and quantification of these processes will be reviewed and discussed, in particular transformation processes (soil physical and chemical weathering) and translocation processes (eluviation and illuviation and soil mixing).

U. Stockmann (✉) • B. Minasny • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia
e-mail: uta.stockmann@sydney.edu.au; budiman.minasny@sydney.edu.au; alex.mcbratney@sydney.edu.au

S. Salvador-Blanes
Laboratoire Géo Hydrosystèmes Continentaux, Faculté des Sciences et Techniques, Parc de Grandmont, 37200 Tours, France
e-mail: salvador@univ-tours.fr

T. Vanwalleghem
Departamento de Agronomia, Universidad de Córdoba, Córdoba, Spain
e-mail: ag2vavat@uco.es

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_19

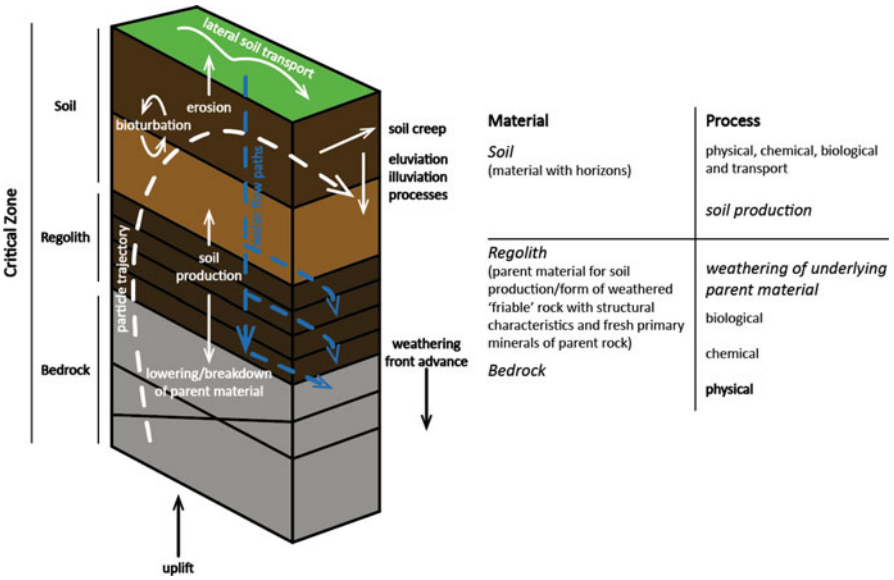


Fig. 19.1 Important pedogenic processes that are responsible for the differentiation of a soil profile (After Stockmann 2010)

19.1.1 Chronofunctions

Jenny's (1941) *clorpt* model introduced in the previous chapter describes the relationship between soil properties and time. The term chronofunction is equated with the mathematical expression of chronosequence data representing a solution to Jenny's state-factor equation (Yaalon 1975; Schaetzel et al. 1994). Jenny (in Stevens and Walker 1970) stated that if we know the ages and properties quantitatively, we have a chronofunction and can fit rate equations to the data; if the ages are relative, we have a chronosequence, from which we can learn a lot about processes and mechanisms, but not necessarily rates. Chronofunctions can perhaps be seen as nonstationary and to represent soil evolution from some non-equilibrium state to an equilibrium state.

Chronosequences are used to investigate and understand the formation of soil profiles, i.e. placing soil profiles developed from surfaces of known or dated age in a chronological order (Huggett 1998; Sauer et al. 2007). They can be used to formalize chronofunctions where soil and landscape properties are plotted against the independent variable time:

$$S\,(\text{Soil}) = f_t\,(\text{time})\,\text{cl}, o, r, p \ldots \tag{19.1}$$

If we believe soil formation is a result of predominantly chemical processes, we may expect the form of the chronofunctions to obey chemical models such

as zero-, first- or second-order kinetics (Sparks 1995). First-order kinetic models show an exponential evolution with time to some asymptotic value. One of the main parameters of this model is a rate constant which can give us estimates of the natural rate of soil formation.

Obtaining data for a chronosequence is not easy; following Jenny's factorial model, it may be impossible to achieve, i.e. finding a site with the same parent material, under ineffectively varying climatic conditions and influence of organisms, and constant relief. However, approximate chronofunctions with various assumptions can be obtained. In several parts of the world, volcanic activity, mudflows or glaciation has created surface materials or soil where the relative age can be determined. An example can be found on the Hawaiian Islands, with a sequence of volcanic surfaces that ranges in substrate age, from a currently active volcano (0 years) to the 4 million-year-old island of Kauai (Kitayama et al. 1997a, b). Kitayama et al. (1997a, b) chose an array of sites formed from basaltic lava parent material, with similar elevation about 1200 m above sea level with minimal topographic relief and a mean annual rainfall of 4000 mm. The island also has a uniform vegetation coverage with rain forests dominated by a single tree species.

Such chronosequences can be utilized in the calculation of the *surface age-profile thickness* (SAST) which represents the long-term average of soil formation (Egli et al. 2014). In a stable environment with minimal processes of erosion and deposition, soil formation rates are calculated from the thickness of the soil divided by the age of the surface soil (in mm/year or if density is considered in $Mg/km^2/year$).

There are different types of mathematical functions that have been commonly used in chronofunctions, to express soil evolution with time, as exemplified in Fig. 19.2, with simple linear and logarithmic functions perhaps being the most frequently used. Jenny (1941) postulated that soil property change over time could be of sigmoidal shape with an initial exponential rate and that the changes gradually become small as they are reaching a steady-state condition (Yaalon 1975). Barrett and Schaetzl (1992) used a single logarithmic model to exemplify the change of the quantity of iron over time during podsolization for a sandy soil near Lake Michigan. Hay (1960) on the other hand found an exponential relationship between clay formation and time from volcanic ash on the island of St. Vincent in the Caribbean as would be expected from first-order kinetics.

19.1.2 Soil Weathering Models and Rates

The importance of soil and its genesis and therefore the quantification of processes of soil weathering are part of at least two of the nine *Grand Challenges in Earth Surface Processes* that has been put forward by the NRC (2010) publication. It specifically states that "The breakdown of bedrock – a major factor in Earth surface processes – is among the least understood of the important geological processes".

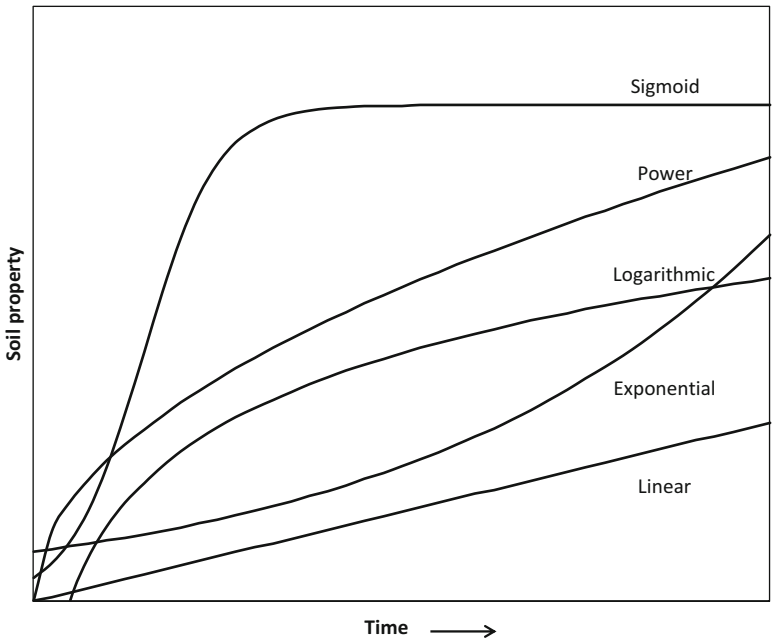


Fig. 19.2 Types of mathematical functions commonly used in chronofunctions. S-shaped or sigmoidal curve, general form of equation, $Y = 1/(a + \text{bexp}(-t))$; power functions, general form of equation, $Y = \text{at}^b$; logarithmic functions, general form of equation, $Y = a + b(\text{logt})$; exponential functions, general form of equation, $Y = \text{aexp(bt)}$; simple linear functions, general form of equation, $Y = a + \text{bt}$ (Stockmann 2010)

Over the years, geomorphologists and pedologists have attempted to formalize rates of soil production from parent materials which led to two assumed main concepts or models of soil weathering with time, the (1) *exponential soil production model* and the (2) *humped soil production model* (visualized in Fig. 19.3). The first states that soil production decreases exponentially with increasing thickness of the overlying soil mantle (Ahnert 1977; Heimsath et al. 1997), whereas the second model explains the conversion of rock into soil using a humped function where soil production is greatest below an incipient soil depth and slower for exposed bedrock or an already thick soil mantle (Gilbert 1877; Humphreys and Wilkinson 2007).

The first concept, the exponential decline of the soil production rate (*SPR*) with increasing soil depth, can be described as (Dietrich et al. 1995; Heimsath et al. 1997):

$$\text{SPR} = P_0 \exp(-bh) \tag{19.2}$$

where P_0 ([L T^{-1}], mm kyr^{-1}) is the rate of weathering of bedrock at zero soil thickness (h ([L], cm)) and b ([L^{-1}], cm^{-1}) is a rate constant, a length scale that

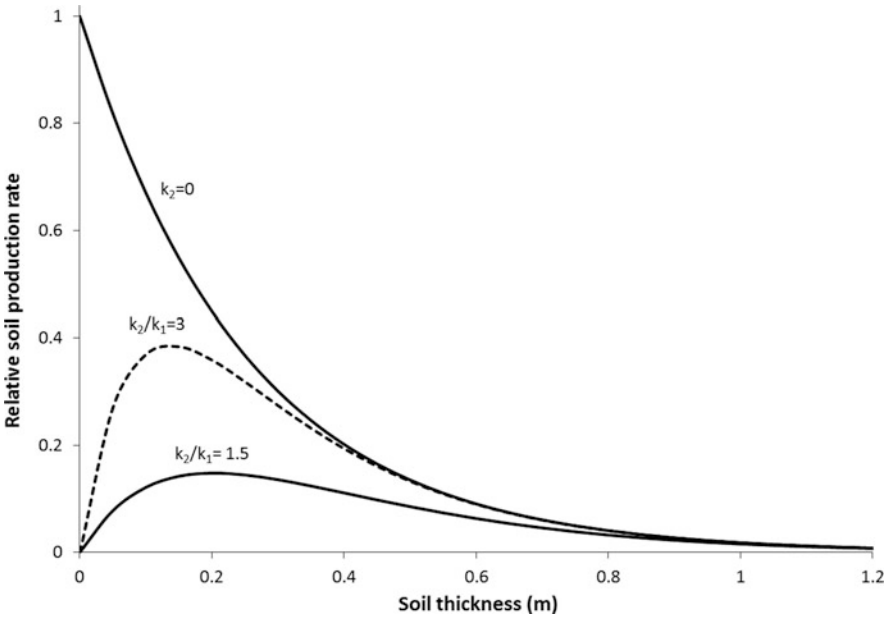


Fig. 19.3 The rate of soil production versus soil thickness (based on Minasny and McBratney 2006; Furbish and Fagherazzi 2001). Here, both the exponential and the humped soil production models are presented. Both axes are dimensionless. Soil production is presented graphically depending on different values of the parameter k_1 and k_2 (also refer to Eqs. 19.3 and 19.4). If $k_2 = 0$, the soil production equals a depth-dependent exponentially decreasing soil production function. If $k_2/k_1 \geq 0$, soil production shows a humped function (changed after Stockmann et al. 2011)

characterizes the decline in soil production with increasing soil thickness. This model of soil production was first verified with field data from the Tennessee Valley in California, USA, by Heimsath et al. (1999), employing terrestrial cosmogenic nuclides (TCN).

The second concept, the humped model of soil production, can for example be formalized as a continuous double exponential function (Minasny and McBratney 2006):

$$\frac{\partial e}{\partial t} = -\left(P_0\left[\exp\left(-k_1 h\right) - \exp\left(-k_2 h\right)\right] + P_a\right) \tag{19.3}$$

where P_0 ([L T^{-1}], mm kyr^{-1}) represents the rate of weathering of bedrock, h ([L], cm) the soil thickness, k_1 the rate of mechanical breakdown of the rock materials and k_2 the rate of chemical weathering and P_a the weathering rate at steady-state condition ([L T^{-1}], mm kyr^{-1}) with condition $k_1 < k_2$. When k_2 equals 0, the humped function is reduced to the depth-dependent exponential soil production function (Eq. 19.2). The critical thickness, h_c, where weathering is at maximum is written as:

$$h_c = \frac{\ln(k_2/k_1)}{k_2 - k_1} \tag{19.4}$$

An empirical parameterization of the humped soil production model is still to be achieved, but this model was used to explain soil formation conceptually in some landscapes. Heimsath et al. (2009), for example, postulated that humped soil production occurred at their study site, Arnhem Land in northern Australia, an outcrop-dominated soil landscape where soil depths of less than the peak in soil production (35 cm) could not be observed. In a soil landscape dominated by humped soil production, it is assumed that soil depths less than the maximum in soil production are unstable and continuously eroded to expose the parent rock material (Dietrich et al. 1995).

19.1.2.1 Physical Weathering

Models of Physical Weathering

Physical weathering processes can be represented using different rock fragmentation models such as symmetric, asymmetric and Whitworth fragmentation and granular disintegration. Parent particles are weathered into daughter particles whilst preserving their mass based on these four different fragmentations which are shown graphically in Fig. 19.4 (Sharmeen and Willgoose 2006). Symmetric fragmentation weathers a parent particle into two daughter particles of equal volume; asymmetric fragmentation on the other hand results in two daughter particles of unequal volumes; Whitworth fragmentation results in a distribution of daughter particles proposed by W.A. Whitworth. Granular surface disintegration is a process used to model the particle breakdown of a thin surface layer which results in several equally sized spherical daughter particles that are of the same diameter as the thickness of the surface layer and one large daughter particle of equal diameter as the parent particle less twice the layer thickness (Wells et al. 2008).

Wells et al. (2008) modelled physical weathering based on these fragmentation models, including the probability of fragmentation to occur in a given time period. They found that the physical weathering rate increased linearly with time, based on the probability of fracture to occur. The Wells' fragmentation model was implemented by Welivitiya et al. (2016) for their soil-landscape model.

Quantifying the Rate of Physical Weathering

Several approaches have been utilized to parameterize weathering rates of parent material to soil from field data. Terrestrial cosmogenic nuclides, predominantly ^{10}Be, have been employed to derive soil production rates of soil (Stockmann et al. 2014). In this method, soil production is interpreted as the physical conversion of bedrock into soil, which is usually expressed in mm of bedrock weathered over time

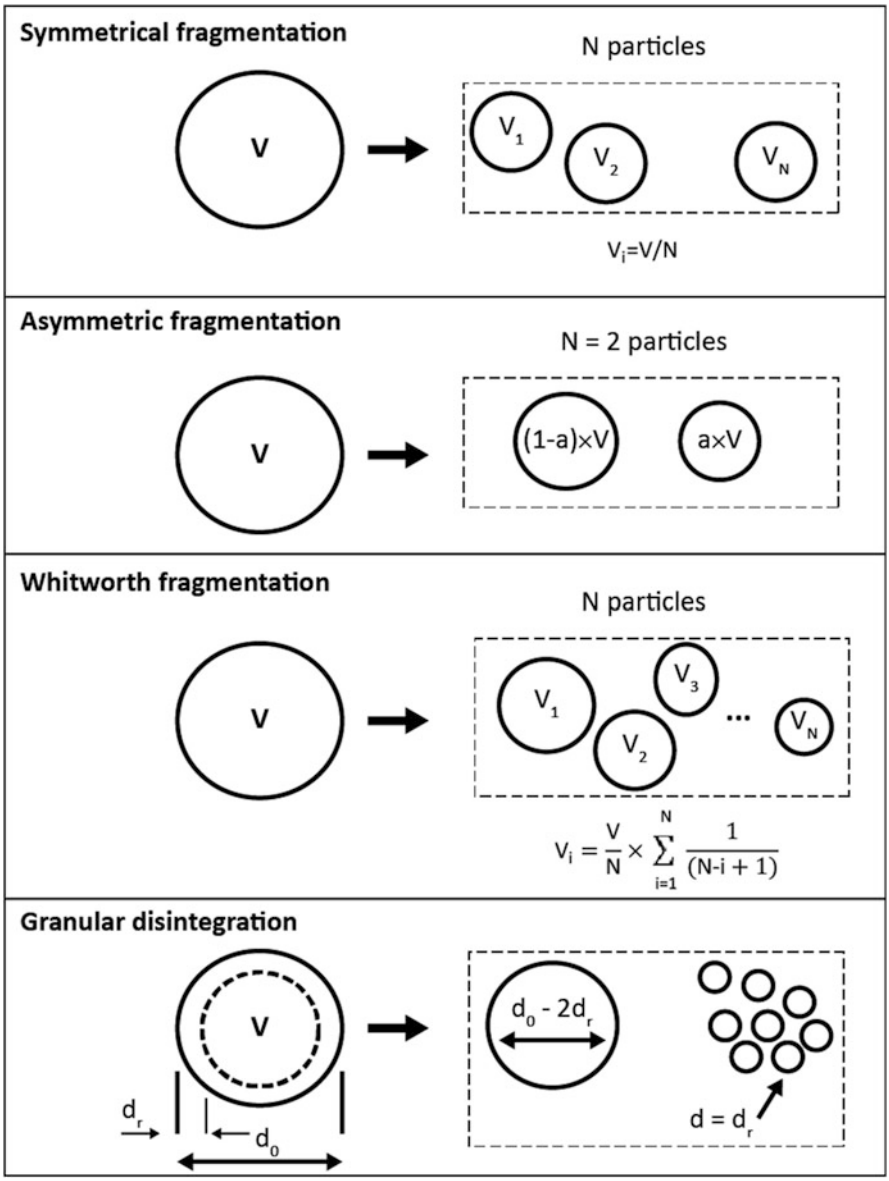


Fig. 19.4 Fragmentation geometries used in the modelling of physical weathering; with V being the volume of a parent particle that fractures into N daughter particles (V_1, V_2, etc.), d the diameter of the parent particle, d_r the surface layer thickness of the parent particle and a the ratio of daughter particle volumes (Redrawn after Wells et al. 2008)

(mm kyr^{-1}). For rate calculations, the rate P of ^{10}Be production in atoms g-quartz^{-1} year^{-1} at depth h and slope θ (Stone 2000), the half-life of ^{10}Be in years (Chmeleff et al. 2009) and the concentration N of ^{10}Be in atoms g-quartz^{-1} in the sample of interest (parent material) are required. Calculations are then based on a model by Nishiizumi et al. (1991) that allows for calculations of the steady-state concentration of ^{10}Be in the sample of interest. This model assumes that ^{10}Be concentrations are controlled by the increase of concentrations with exposure time and the erosion rate of the parent material itself (ε or SPR in cm year^{-1}). Soil production rates in cm year^{-1} or mm kyr^{-1} can then be calculated by solving this model for SPR (ε) (Lal 1991; Heimsath et al. 1997):

$$\mathrm{SPR} = \frac{\Lambda}{\rho_{\mathrm{parent}}} \left(\frac{P\,(h, \theta)}{N} - \lambda \right) \tag{19.5}$$

where ρ_{parent} is the mean density of the parent material in g cm^{-3}, λ is the decay constant of ^{10}Be ($\lambda = \ln 2 / ^{10}$Be half-life) and Λ is the mean attenuation of cosmic rays (g cm^{-2}). Annual production rates of ^{10}Be (P) need to be normalized to the geographical position of the site studied (elevation, latitude and longitude) (Stone 2000); for soil-mantled landscapes, rates also need to be corrected for the soil overburden and shielding by slope (Dunne et al. 1999; Granger and Muzikar 2001). Soil production rates are calculated with the assumption of steady-state conditions where soil erosion and soil production rates are balanced throughout the production of TCN.

This technique was used in field studies situated in soil-mantled landscapes of Australia, North America and South America. Stockmann et al. (2014) compiled these measured rates of soil production which were as low as 0.0001 mm year^{-1} and as high as 0.6 mm year^{-1}, for soils of up to 108 cm of depth. These TCN field data were then used to derive for the first time a quantitative estimate of 'global soil production' with a rate of about 0.2 mm year^{-1} (286 Mg km^{-2} year^{-1}). This rate reflects the potential weathering rate, P_0, for soil-mantled landscapes at zero soil depth (refer to Eq. 19.2). Such estimates are important for modelling global landscape dynamics, as we need to know the rate of soil replenishment from bedrock compared to its loss through erosion. This will be discussed a bit further in the concluding section of this chapter.

Uranium-series isotopes have also been used to estimate soil weathering or production rates (Dosseto et al. 2008). This technique assesses soil weathering throughout a soil profile based on the abundance of the U-series which is considered to be a function of chemical weathering and time and its distribution between primary and secondary minerals. In situ weathering can be identified (increase in weathering from the bottom to the top of a soil profile) through the decrease in the ^{234}U/^{238}U ratios with decreasing soil depth. Soil residence times are then calculated by modelling the U-series activity ratios in a soil profile (Suresh et al. 2013). Research has shown that soil production rates derived with U-series isotopes

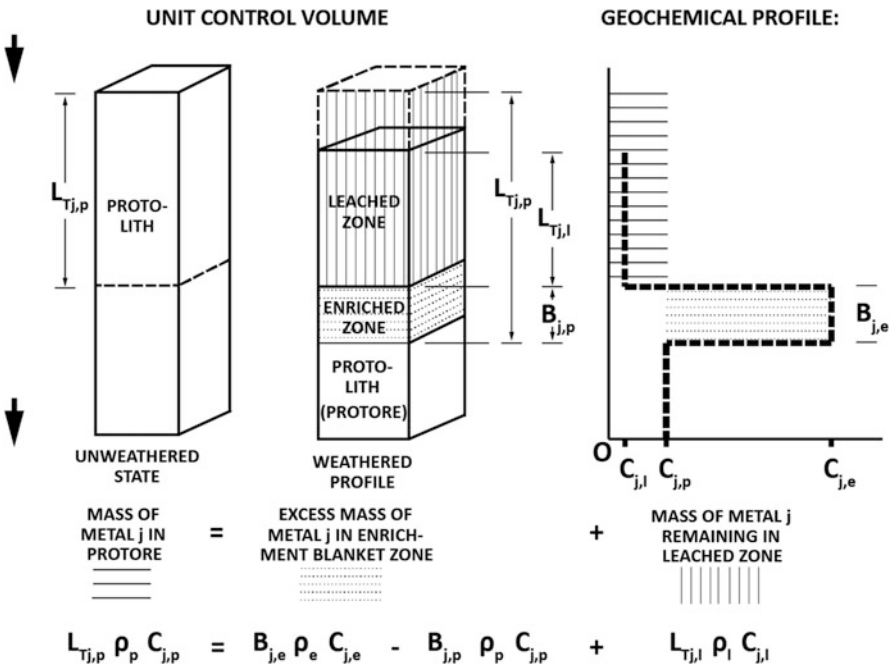


Fig. 19.5 Graphical representation of one-dimensional mass balance for vertical supergene metal transport and secondary enrichment. The mass of element *j* leached is equal to the mass of *j* fixed in the underlying zone of secondary enrichment in this closed system model (with no lateral fluxes of loss of the element *j* through basal discharge) in which a protolith (*p*) volume becomes differentiated into two related parts. The uppermost subsystem is leached in element *j*, and the lower subsystem positioned further along a ground water flow line becomes enriched in *j*. Two strain terms are necessary. In the leached (*l*) zone, $\epsilon_{j,l} = (L_{Tj,l} - L_{Tj,p})/L_{Tj,p}$, and in the enriched (*e*) zone, $\epsilon_{j,e} = (B_{j,e} - B_{j,p})/B_{j,p}$, with *L* describing the thickness of the near-surface zone, *B* the thickness of the lower subset zone, ρ the density and *C* the concentration terms (Redrawn after Brimhall and Dietrich 1987)

are comparable to those derived from TCN (Dosseto et al. 2008; Suresh et al. 2013). Both methods therefore are quite robust to estimate soil production rates for soil-mantled hillslopes.

19.1.2.2 Chemical Weathering

Models of Chemical Weathering

Brimhall and Dietrich (1987) proposed a mass balance model that formally links chemical composition (of bedrock and soil) to bulk density, mineral density, volumetric properties, porosity and amount of deformation (strain). Figure 19.5 shows a graphical representation of such a chemical modelling approach.

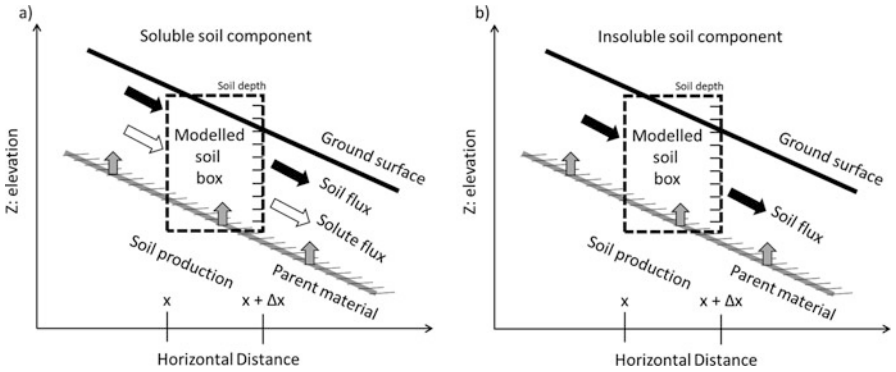


Fig. 19.6 Mass fluxes on hillslope soils: (**a**) mass fluxes of soluble soil components in and out of a modelled soil box, (**b**) mass fluxes of an insoluble soil component in and out of a modelled soil box (Redrawn after Yoo et al. 2007)

Furthermore, Yoo et al. (2007) presented a process-oriented hillslope soil mass balance model that links processes of soil chemical weathering with topographic position (Fig. 19.6). The proposed model explicitly considers the influence of lateral soil transport and soil production from underlying bedrock (physical weathering) on soil chemical weathering processes.

Quantifying the Rate of Chemical Weathering

The intensity of chemical weathering has been estimated from field-based studies through applying a variety of investigative methods. Chemical weathering rates have been estimated based on a mass balance approach, using elemental fluxes (loss and gain) in watersheds and the chemical composition of the parent materials and weathering products studied. Based on a compilation of studies reviewed in Stockmann et al. (2011), chemical weathering rates estimated from catchment-based mass loss of elements range between 0.01 and 0.1 mm year^{-1} and are in general lower than rates of physical weathering (also refer to Fig. 19.7).

For a catchment, chemical fluxes can be estimated using a mass balance approach, by determining the solute discharge flux $Q_{i,dis}$ for a chemical species i:

$$Q_{i,\mathrm{dis}} = C_{i,\mathrm{dis}} \frac{V}{\mathrm{At}} \tag{19.6}$$

where $C_{i,dis}$ is the chemical concentration of a chemical species i, V is the fluid mass, A is the geographic area of the watershed and t is time (White and Blum 1995).

The chemical soil weathering rate (W) can also be calculated in situ from the ratio of a resistant or immobile element (e.g. Zr) in the parent rock ($[\mathrm{Zr}]_{\mathrm{rock}}$) as compared to its amount in the weathered rock (saprolite) or soil ($[\mathrm{Zr}]_{\mathrm{soil}}$) (Riebe et al. 2004a, b):

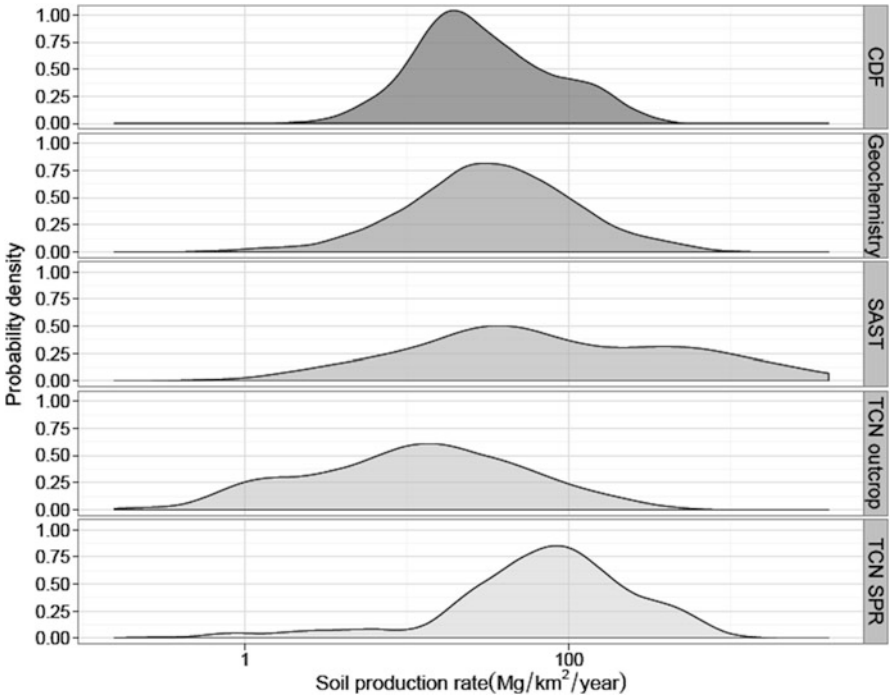


Fig. 19.7 Probability density of soil formation rates in Mg km^{-2} year^{-1} as reported by different methods. Note axes are in logarithmic scale, refer to Table 19.1 for data sources (Source: Minasny et al. 2015)

$$\frac{W}{D} = 1 - \frac{[\mathrm{Zr}]_{\mathrm{rock}}}{[\mathrm{Zr}]_{\mathrm{soil}}} = \mathrm{CDF} \tag{19.7}$$

where CDF is the chemical depletion fraction which equals the ratio of the chemical weathering rate (W) to the total denudation rate (D). Soil weathering rates are often defined with different underlying assumptions. In a range of studies that used weathering ratios, total rates of denudation or weathering (D) were substituted with rates of soil production derived from TCN (Riebe et al. 2003, 2004a, b; Green et al. 2006; Burke et al. 2007, 2009; Yoo et al. 2007). Some studies, however, considered soil production rates determined by TCN as rates of physical weathering only and subsequently added those to chemical weathering rates to calculate total denudation rates (D) (Dixon et al. 2009).

The conservation of mass equation for the chemical weathering rate (Mg km^{-2} year^{-1} or mm year^{-1}) can be written as a fraction of the total denudation rate:

$$W = D\left(1 - \frac{[\mathrm{Zr}]_{\mathrm{rock}}}{[\mathrm{Zr}]_{\mathrm{soil}}}\right) \tag{19.8}$$

The amount of chemical weathering can also be expressed for the loss of a specific element X (e.g. main base cations found in the soil such as Ca, Mg, K and Na) (Riebe et al. 2004a):

$$W_x = \left([X]_{\text{rock}} - \frac{[\text{Zr}]_{\text{rock}}}{[\text{Zr}]_{\text{soil}}} \right) - [X]_{\text{soil}} \tag{19.9}$$

Stockmann et al. (2011) also compiled rates of chemical weathering derived from weathering indices and found that rates vary between 0 and 0.144 mm year^{-1}. Rates derived with this method were relatively similar to rates derived from catchment-based mass balance studies, although overall rates derived from weathering indices were comparatively lower.

19.1.2.3 Conclusions: How Fast Does Soil Form?

As discussed in the previous sections, various field estimates are now available that can provide a quantitative measure of how fast soils form in the landscape. Minasny et al. (2015) compiled such rates in their recent publication that discusses how our world soils shape the landscape (Fig. 19.7). Soil weathering rates from different data sources are compared using the median of the distributions (see Table 19.1 for more details). This statistical measure was used as it is more representative of the distributions which tended to be skewed. Here, weathering rates were reported in units of mass of material over an area over time (Mg km^{-2} year^{-1}). Common bulk densities for rock (2600 kg m^{-3}) and soil (1200 kg m^{-3}) were used to convert volume (mm year^{-1}) to mass for those studies where these were unknown. In the following, the amount of soil produced during weathering is discussed.

Figure 19.7 shows soil weathering rates derived from stable rock outcrops using the TCN technique (global median of 12 Mg km^{-2} year^{-1} or 10 mm kyr^{-1}), and it becomes apparent that those are half to almost one order of magnitude smaller than soil production rates derived with TCN for soil-mantled landscapes (global median of 73 Mg km^{-2} year^{-1} or 60 mm kyr^{-1}). This confirms general assumptions that terrain and environmental factors make a big difference for the intensity of weathering processes. For example, in general the presence of regolith (or partially weathered rock) is a precondition for intense weathering. Regolith or shallow soil mantles form a habitat for fauna and flora, and their presence also enhances the physical and chemical weathering rate. Figure 19.7 also shows that chemical weathering rates based on weathering ratios (global median of 24 Mg km^{-2} year^{-1} or 20 mm kyr^{-1}) and chemical weathering rates based on river geochemistry (global median of 34 Mg km^{-2} year^{-1} or 28 mm kyr^{-1}) are of similar value. Both are about a third of soil production rates derived from TCN.

However, we do not only need to know how fast our soils can form but also how much soil comparatively is lost through processes of erosion. In Fig. 19.8 average soil production rates for soil-mantled landscapes estimated from TCN field data are

Table 19.1 Statistical distribution of soil production and weathering rates (Mg km^{-2} $year^{-1}$) from datasets derived using different measurement techniques

Percentile		SAST[a]	Terrestrial cosmogenic nuclide SPR[b](0–100 cm)	Terrestrial cosmogenic nuclide on rock outcrop[c]	Chemical weathering based on river geo-chemistry[d]	Chemical weather-ing based on CDF[e]
100%	Maximum	4021	934	349	570	263
95%		1606	457	130	197	159
75%	Quartile	338	150	33	72	53
50%	Median	55	73	12	34	24
25%	Quartile	22	35	4	16	15
5%		4	4	1	6	5
0%	Minimum	1	0.3	0	1	0
	n	179	291	530	127	155

Source: Minasny et al. (2015)
[a]*SAST* Surface age-profile thickness (Data from Egli et al. 2014)
[b]*SPR* Soil production rate (Data from Stockmann et al. 2014 combined with Riebe et al. 2004a)
[c]Data from Portenga and Bierman (2011)
[d]Compiled data from Stockmann et al. (2014)
[e]Data from Riebe et al. (2004a) and compilation of Larsen et al. (2014), *CDF* chemical depletion fraction

therefore compared with erosion rates from different sources and scales. It becomes clear that rates and spatial patterns of erosion and deposition depend strongly on the type of erosion processes. For example, in natural environments that exhibit a dense vegetation cover, soil redistribution is mainly driven by mass wasting processes. In this regard, Fig. 19.8 illustrates that soil erosion rates under native vegetation (B) are comparatively low (global median of 0.01 mm $year^{-1}$ equivalent to 0.1 t ha^{-1}) and are in fact in a steady state with soil production rates (*A*) (global median of 0.06 mm $year^{-1}$ equivalent to 0.7 t ha^{-1}). Soil erosion, however, has been highly accelerated by human impact as anthropogenic land use changes and subsequently agricultural management practices have upset the natural balance between soil production and erosion. With a global median of 2 mm $year^{-1}$ (equivalent to 24 t ha^{-1}), soil erosion rates from conventionally managed agricultural soils (*D*) are almost two orders of magnitude higher than soil production rates (*A*). This shows that under humanly managed systems, soil does not seem to be a renewable resource and needs to be managed carefully. Changes in management practices towards more sustainable agricultural systems, where appropriate, however, can narrow the gap between soil production and humanly induced erosion rates. As seen in Fig. 19.8, erosion rates under conservation agricultural practices (*C*) are significantly reduced (global median of 0.1 mm $year^{-1}$ equivalent to 1.2 t ha^{-1}) and are close to a natural balance.

However, not all eroded soil material is actually lost in the streams as lateral transport through soil erosion is also an important mechanism for reshaping the

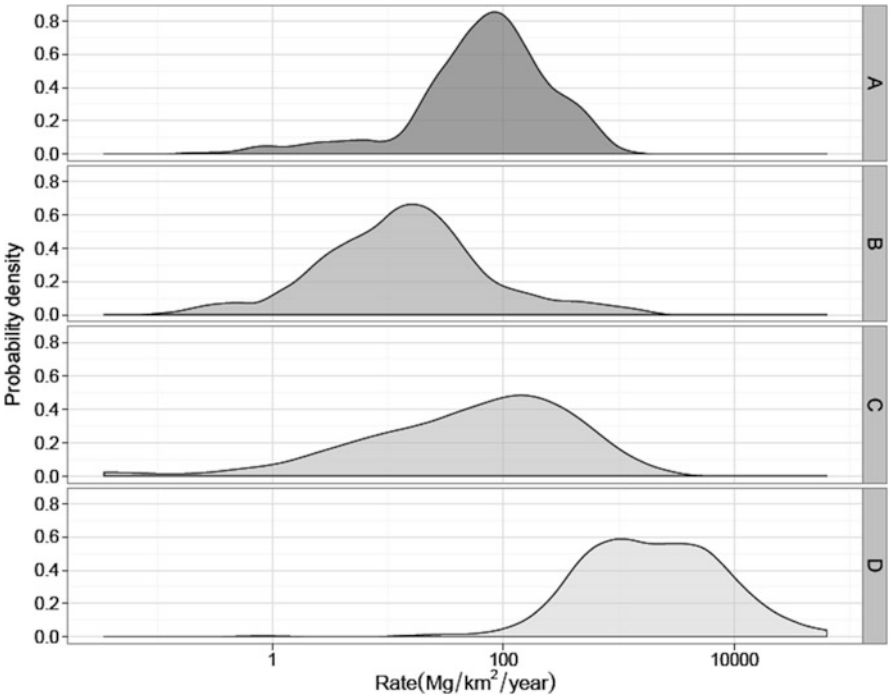


Fig. 19.8 Statistical distribution of average soil production and erosion rates in Mg km^{-2} year^{-1} from different sources and scales. (**a**) Soil production rates determined with TCN, (**b**) Erosion rate of native vegetation, (**c**) Erosion rate of conservation agriculture and (**d**) Erosion rate of conventional agriculture. TCN Soil production rate data are from Stockmann et al. (2014) and Riebe et al. (2004a). Erosion rates from conventional or conservation agriculture and native vegetation are from Montgomery (2007) (Source: Minasny et al. (2015))

Earth's surface through the redistribution of soil material and sediment in the landscape. Estimates of sediment delivery rates to streams, i.e. rates of actual loss of soil to the rivers and oceans, are therefore also needed to assess soil formation and soil transport in the landscape which is still a field that requires more work. Only a few estimates exist in the current literature. The rate of redistribution of soil and sediment to the land surface was estimated, for example, to be of a value of approximately 1.11 mm year^{-1} (Ludwig and Probst 1998). However, much lower estimates of about 0.028 mm year^{-1} (Syvitski et al. 2005) or much higher estimates of about 12.6 mm year^{-1} (Wilkinson and McElroy 2007) have also been reported in the literature on this topic.

19.1.3 Models of Soil Mixing

19.1.3.1 Lessivage

Models of Clay Migration

The works of Van Wambeke in the 1970s are a first attempt to quantify lessivage, the translocation of clay particles during soil profile development (e.g. Van Wambeke 1972, 1976). In his 1976 paper, Van Wambeke (1976) proposed a mathematical model for the differential movement of clay particles which provides a quantitative expression of the pedological process of clay eluviation and illuviation. The quantification of clay migration through a soil profile is based on the assumption that the moving clay fraction which accumulates in the B horizon is associated with fine clay particles and that larger crystals (>0.2 μm) are translocated more slowly through the profile or not at all. These translocation processes alter the ratio of fine clay (0–0.2 μm) to total clay content (0–2 μm) in the A and B horizons which can then be used as quantitative measure for the intensity of clay migration through the soil profile. These assumptions are only true however in a closed system without the removal, destruction, weathering, formation or addition of soil minerals, with constant horizon thicknesses and where no processes of erosion occur.

Quantifying Clay Migration

Radionuclides that are strongly adsorbed to the clay fraction and organic matter can be used to track their movement down the soil profile (He and Walling 1996; Zapata 2003). ^{210}Pb and ^{137}Cs that exhibit a half-life of 22.3 years and 30.2 years, respectively, can be used to investigate short-term vertical processes of soil translocation. These are moving passively down the soil profile using soil particles as carrier substances.

$^{210}Pb_{ex}$ is a product of the ^{238}U decay series originating from the decay of gaseous ^{222}Rn, a daughter radionuclide of ^{226}Ra. ^{226}Ra is found naturally in the soil and generates ^{210}Pb which is usually in equilibrium with its parent ^{226}Ra. Diffusion of small amounts of ^{222}Rn introduces ^{210}Pb into the atmosphere, and the fallout of this quantity of ^{210}Pb on the soil surface is termed the excess ^{210}Pb (hence $^{210}Pb_{ex}$) that can be used to investigate its passive distribution through the soil profile. Dörr (1995) used $^{210}Pb_{ex}$ to quantify the movement of organic matter and clay particles in forest soils, whereas Jagercikova et al. (2014) used $^{210}Pb_{ex}$ to quantify clay migration under different farming practices.

Different to ^{210}Pb which is a naturally occurring radionuclide, ^{137}Cs stems from fallout of nuclear weapon tests (1950s to the 1970s) and accidents (e.g. Chernobyl, Ukraine, in 1986), and its accumulation in the surface soil can therefore be dated quite accurately, but its concentration in the world soils is also diminishing

following these fallout events. Because of its radioactive decay, ^{137}Cs will not be available in the near future to be used in particle migration studies (Mabit et al. 2008).

19.1.3.2 Bioturbation

Models of Bioturbation

Gabet et al. (2003) reviewed quantitative models of bioturbation and sediment transport. They proposed a general slope-dependent model to calculate the horizontal volumetric flux of sediment (q_{sx}) caused by root growth and decay:

$$q_{\mathrm{sx}} = \frac{(\mathrm{xr\tau})}{(\rho_r)} \tag{19.10}$$

where x (m) is the net horizontal displacement of soil, r (kg m^{-2}) is the root mass per unit area, τ (year^{-1}) is the root turnover rate which is a measure of the annual belowground production to the maximum belowground standing biomass and ρ_r (kg m^{-3}) is the density of the root material. This model was used to derive estimates of sediment flux by root growth and decay for temperate grasslands (2.1×10^{-4} m^{-2} year^{-1}), sclerophyll shrubs (6.8×10^{-4} m^{-2} year^{-1}) and temperate forests (8.8×10^{-4} m^{-2} year^{-1}). Gabet et al. (2003) also formalized the horizontal alteration of soil along a hillslope caused by tree throw based on uphill and downhill mound building:

$$x_n = \overline{x_d} - \overline{x_u} = \frac{2}{\pi}(W + D)\sin\theta \tag{19.11}$$

where x_n is the long-term net horizontal transport distance (Eq. 19.11), x_d the horizontal distance of displacement of the root plate centroid caused by trees that were falling directly upslope (Eq. 19.13), x_u the horizontal distance of displacement of the root plate centroid caused by trees that were falling directly uphill (Eq. 19.12), W the width of the root plate and D the depth of the excavated pit.

$$x_u = \frac{W}{2}(\cos\theta - \sin\theta) - \frac{D}{2}(\cos\theta + \sin\theta) \tag{19.12}$$

$$x_d = \frac{W}{2}(\cos\theta + \sin\theta) + \frac{D}{2}(\cos\theta - \sin\theta) \tag{19.13}$$

Following on, Gabet and Mudd (2010) proposed a numerical model on bedrock erosion by root fracture and tree throw. Tree throw and associated processes of pit excavation and mound building that are responsible for the large-scale topography at the soil surface are modelled implementing concepts from the 2003 paper

(e.g. Eq. 19.11=x_n). Other processes, such as the fill in of the soil pits and flattening of the mounds that occur on smaller scales, are represented through slope-dependent processes of soil creep which is modelled through simple linear diffusion:

$$q_{sc} = DS \tag{19.14}$$

where q_{sc} (m^2 $year^{-1}$) is the sediment flux, D (m^2 $year^{-1}$) is the diffusivity and S (m m^{-1}) is the local slope.

This coupled biogeomorphic model is driven by field data from the Pacific Northwest (USA) on conifer population dynamics, rootwad volumes, tree throw frequency and soil creep. Model outcomes show a humped soil production relationship between bedrock erosion through biota and developing soil thickness. This relates to the principle that as the soil thickens, it becomes less likely for tree roots to disrupt the bedrock through weathering and that the growing soil medium provides a more favourable habitat for trees.

In this regard, recently, Shouse and Phillips (2016) investigated the effects of parent material on the biomechanical process of deepening of soils by trees. Especially the deepening of shallow soils is affected by this process of root penetration of parent material rocks. Two kinds of tree habitable bedrock were assessed in this study in comparison to adjacent non-tree sites, dipped and contorted rock with plenty of joints and bedding planes accessible for tree roots and flat level-bedded sedimentary rock. This study found that soils beneath tree stumps were significantly deeper, and the authors concluded that soil deepening effects through trees are an important mechanism under both easy and not so easily root accessible lithologies.

Quantifying the Rate of Bioturbation

Radionuclides and also the technique of optically stimulated luminescence (OSL) have been used for 'particle tracking' within the soil profile and thus for generating rates of soil mixing. OSL is a dating technique that measures the time since soil particles (usually sand-sized grains) have been last exposed to sunlight before burial in the soil (Aitken 1998). A burial age in years for individual grains can be calculated using the dose (D_e in Gy, 1 Gy = 1 J kg^{-1}) the grains accumulated since burial, together with the annual dose rate (D_r in Gy $year^{-1}$) the site studied receives:

$$\text{Burial age} = \frac{D_e}{D_r} \tag{19.15}$$

For example, Wilkinson and Humphreys (2005) and Stockmann et al. (2013) used OSL to investigate rates of soil mixing of forest soils, whereas Kaste et al. (2007) employed the short-lived radionuclides ^{7}Be originating from cosmic radiation and the fallout radionuclide $^{210}Pb_{ex}$ to explore rates of soil bioturbation.

19.2 Soil Profile Models

Pedologists have long been studying soils in many different environments to understand their distribution in the landscape (Dijkerman 1974). The basis for these studies is an approach at the pedon scale. These studies have allowed a good understanding of soil genesis processes, conducting to conceptual models of soil formation. These conceptual models that were discussed in Chap. 18 have in particular been of great use for soil mapping. But the major challenge for pedologists is to develop quantitative techniques, to be able to communicate their qualitative understanding of soil evolution to other disciplines (Hoosbeek 1994). Indeed, whereas other compartments of the ecosystem are quite well addressed, the soil seen as a whole entity was and still is often seen as a black box by non-pedologists (Wagenet et al. 1994).

On the other hand, many studies in the last decades have allowed a reasonably good understanding of individual soil processes, whether physical (e.g. heat transport, water movement, physical breakdown), chemical (e.g. ion exchange) or mineralogical (e.g. mineral dissolution). These studies are of great importance to understand the fate of solid and dissolved matter in the soils in specific situations, but do not directly address our approach. The PROFILE biogeochemical model (Sverdrup and Warfvinge 1993) is one of the attempts to integrate sub-models into a single operational model. The focus has been put exclusively on biogeochemical changes and transfer of dissolved matter at the profile scale, generally for short timescales. This model is of major importance to quantify the evolution of soils in particular in response to present human activities (e.g. critical loads), but reflects only partially the evolution of soils, as it does not take into account, for example, the evolution of particle size.

There is therefore a great need to develop an interdisciplinary approach (Brantley et al. 2011) to allow a connection between these processes to integrate them in a single global model, with a pedologic perspective (Levine and Knox 1994). This modelling at the profile scale would moreover be a possibility to test our understanding of soil formation from the incipient stages (Oreskes et al. 1994; Heuvelink and Webster 2001).

The mechanistic modelling of soil formation encounters several difficulties:

– *The variation of the factors of soil formation over the timescale of soil formation, affecting the rate of the processes, but as well the processes themselves of soil formation*:

  - Variation of climate conditions in the past and the difficulty to reconstruct those
  - Changes in soil genesis processes occurring over time, e.g. changes in land use at the Holocene, recent mechanization (Sommer et al. 2008), changes in vegetation due to climate changes or changes in soil properties, etc.

- The integration of processes into a single model when the level of temporal resolution is not the same (e.g. organic matter input and decay vs. particle size evolution)

– *A good knowledge of the modelling of some processes (in particular biogeochemical processes), but a limited knowledge for others* (e.g. *bioturbation, clay migration, physical fractionation)*.

– *A difficulty to validate these models due to the large temporal scales involved*. One of the solutions proposed is to work specifically on chronosequences which we discussed in the previous chapter.

Only a few studies have emerged to address this purpose since the founding theoretical work developed by Kirkby (1977). Two approaches can be envisaged: using existing sub-models and merging them in a single soil profile model or modelling soil profile evolution using simple equations that limit the number of input parameters with the aim to remain sufficiently generic and simple. An illustration of the first approach is the work performed by Finke and Hutson (2008) and Finke (2012). They model the evolution of soils from loose parent material by integrating two existing sub-models (focusing on carbon dynamics and biogeochemistry, respectively) and adding new soil formation processes. The second approach can be illustrated by the work performed by Salvador-Blanes et al. (2007), where soils developed from hard bedrock and are modelled using as a basis the work performed by Minasny and McBratney (1999, 2001).

19.2.1 The Founding Work of M. J. Kirkby

Kirkby (1977, 1985) developed the first comprehensive mathematical model of soil profile evolution. The basis for the soil profile modelling is here to consider the 'proportion p of substance remaining' at any depth, the value of p approaching 1 asymptotically at depth. In this model, there is no assumption about the exact limit between soil and unweathered parent material. Processes such as change in the bulk density of the soil, physical translocation of clays and particle size evolution are discarded.

Three sub-models with different timescales are considered: organic matter, nutrient cycling and the weathering profile. In situ processes linked with these sub-models comprise nutrient uptake, organic matter input and decay through leaf-fall, nutrient cycling, mixing of the topsoil, solute transfer through leaching and ionic diffusion. A mechanical denudation rate, introducing geomorphic processes, is as well considered. The soil chemistry is simplified to integrate all processes at once and in a geomorphic perspective.

19.2.1.1 Percolation

The percolation process is modelled as an annual flow proportional to the pore-space available, the latter being considered as equal to the accumulated deficit of weathered material. Annual evapotranspiration and total rainfall are specified as input parameters. Evapotranspiration is a function of the distribution of roots that follows an exponential decay with soil depth. This distribution is considered as constant in the model. The amount of water percolating at any depth is a function of rainfall less evapotranspiration.

19.2.1.2 Solubility

A key issue for modelling the solubility of the mineral constituents is that thermodynamic equilibrium between solutes and soil composition is assumed: the water residence times are considered long enough to approach equilibrium. The mineral constituents of soils are simplified and considered as a mixture of their constituent oxides. These dissolve independently to create ions, using Gibbs free energy values that are empirically adjusted by comparing the values for the minerals and the sum of those values considering their constituent oxides. Weathering at a given soil depth is a function of the quantity of water passing through, combined with values of partial pressure of CO_2 linked with the vegetation cover. Simulations show that soils formed from various parent materials are all eventually enriched in sesquioxides after a variable duration. Both the vegetation and the climate component are strongly affecting the degree of weathering through the partial pressure of CO_2 and the accumulated flow, respectively.

19.2.1.3 Leaching

The leaching process is only considered for the inorganic profile. As for the organic profile, nutrients are only released through decomposition. The total solute concentration in the inorganic profile for a given proportion of substance remaining is considered as equivalent to the product of the difference in concentration of inorganic materials relative to the weathering profile and their solubility.

19.2.1.4 Ionic Diffusion and Organic Mixing

Ionic diffusion allows a redistribution of solutes in the soil profile, proportionally to the concentration gradient. It is a key process in areas where the flow of water is very low to allow weathering, especially close to the unweathered bedrock. The ionic diffusion coefficient is here defined as a function of the porosity of the soil.

The organic mixing process is modelled as a diffusion process of soil material through the burrowing activity of the soil fauna. It decreases exponentially with depth. Therefore, while organic mixing is a dominant process close to the surface, ionic diffusion is dominant at depth.

19.2.1.5 Nutrient Uptake, Leaf-Fall and Organic Matter Decomposition

Inorganic nutrients are supplied to plants through evapotranspiration, without considering seasonal variations. The total uptake of a nutrient is approached as the product of the ion concentration and of evapotranspiration rate. These nutrients are considered to be taken only from the inorganic part of the soil. The model considers that there is a constant proportion of plant biomass that falls annually. Only aboveground biomass input is considered. An equilibrium rate balancing uptake is reached over 25 years in the model. Organic matter decomposition rates are considered as constant for a given climate, whatever the organic matter composition. This rate is set at 0.2 a^{-1} for a climate with a mean annual temperature of 10 °C.

The equations of these processes are combined whenever relevant in the three sub-models: organic matter, nutrient cycling and the weathering profile. The simulations performed allow producing more or less thick soils, organic-enriched topsoil, base-depleted topsoils in aridic conditions and more or less base-depleted intermediate horizons (root uptake activity) with a base-rich topsoil in humid conditions.

This pioneering work is the first attempt for the development of a comprehensive model of soil profile evolution. However, no further work derived from this initial attempt.

19.2.2 The Pedogen Model

Salvador-Blanes et al. (2007, 2011) designed the Pedogen model of soil formation at the profile scale. The aim here was an attempt to translate the pedologist's approach – corresponding to a general phenomenological model – to a quantitative model of the key soil genesis processes. The idea is here to quantitatively model the transformation of a hard rock into soil material at the profile scale through various pedological processes that result in mineralogical transformations, organic matter input and decay and the translocation of solid or dissolved matter. In that respect, the model focuses on physical and chemical weathering, bioturbation, organic matter input and decay (Figs. 19.9 and 19.10) and on the retroactions between these processes with time steps of several decades to centuries that require simplifications in the processes modelled.

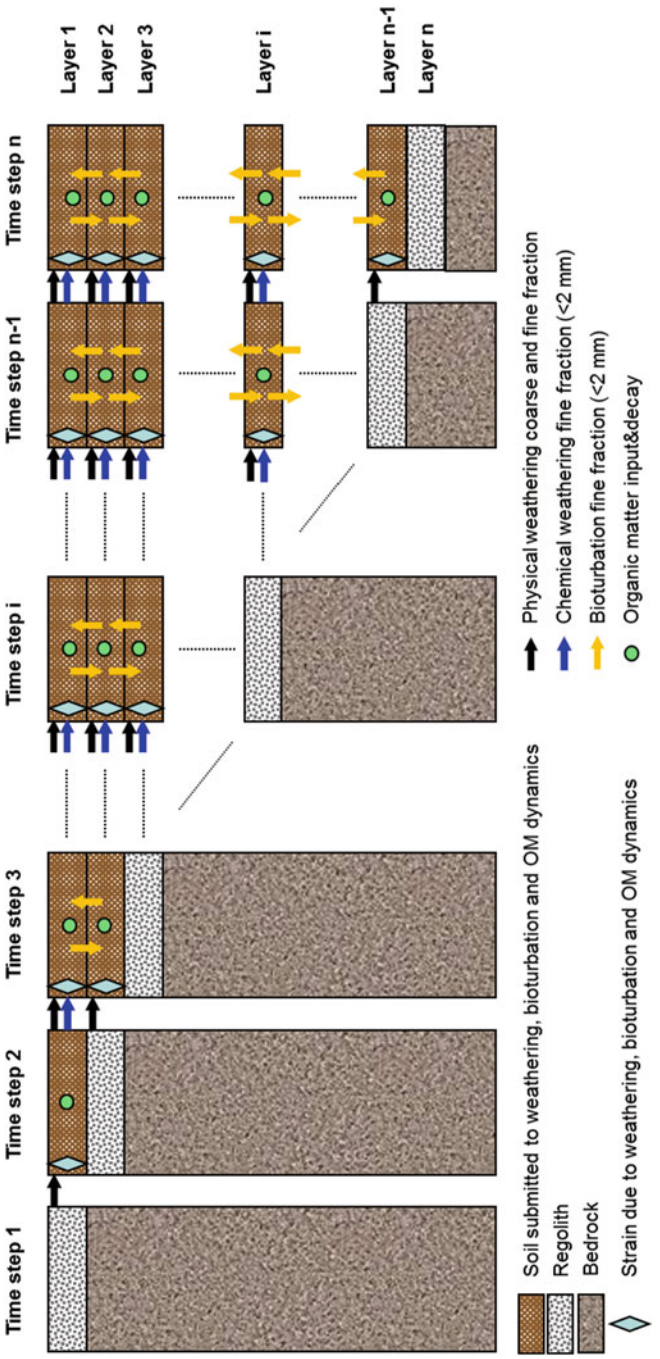


Fig. 19.9 Structure of the Pedogen model (After Salvador-Blanes et al. 2007, used with permission)

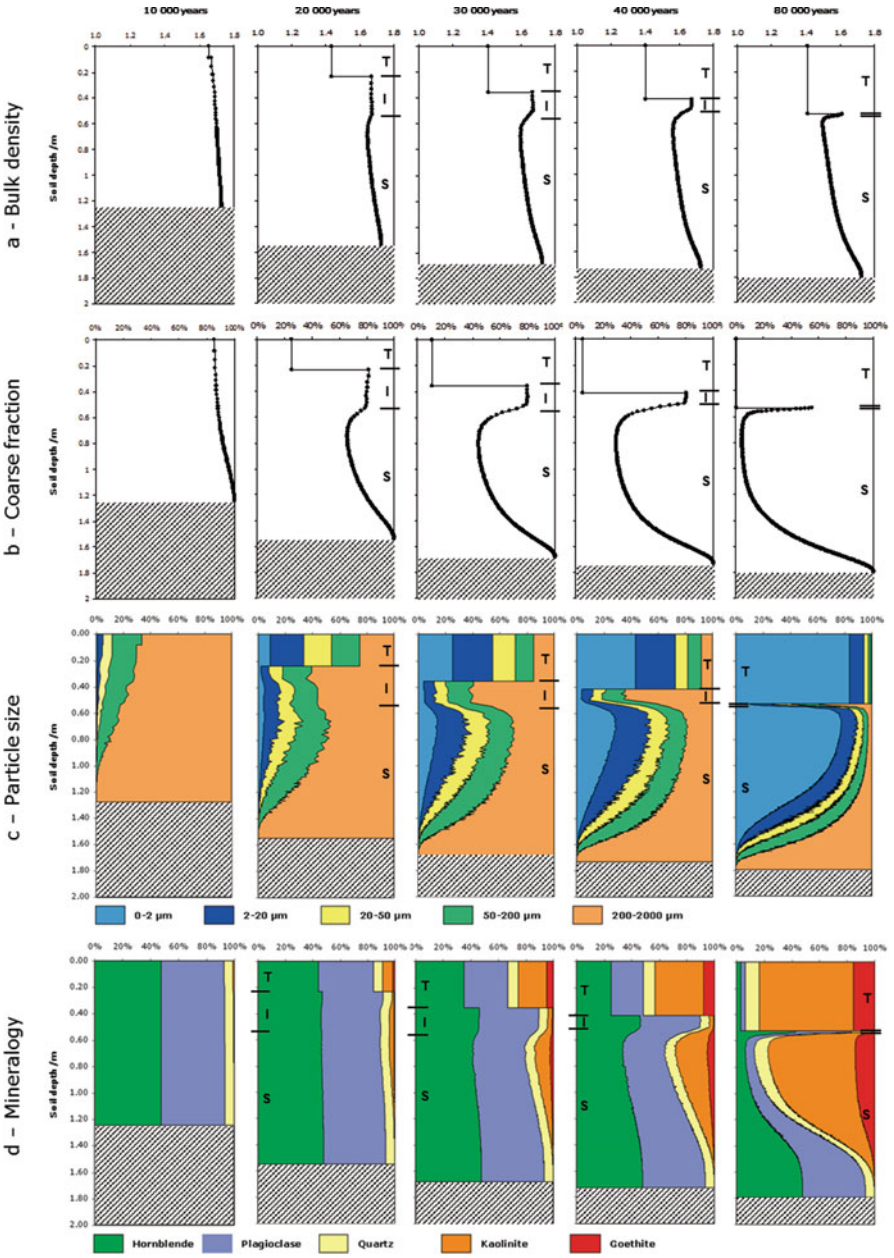


Fig. 19.10 Example of the Pedogen model outputs: bulk density, coarse fraction, particle size and mineralogy in the soil profile after 10,000, 20,000, 30,000, 40,000 and 80,000 years (*T* topsoil, *I* intermediate, *S* subsoil horizon) (After Salvador-Blanes et al. 2007, used with permission)

19.2.2.1 Release of Regolith

The model is based on lowering rate of the bedrock surface that follows a negative exponential decline with increasing soil thickness (Minasny and McBratney 1999). Therefore, at each time step, a given quantity of regolith is released by the bedrock and constitutes a so-called layer with a given thickness. This layer is submitted at each subsequent time step to weathering, bioturbation and organic matter input and decay. The aim is here to follow the evolution of each layer; the sum of these layers constitutes the soil profile.

19.2.2.2 Physical Weathering of Coarse Fragments

The regolith released by the bedrock is considered to be made of spherical coarse fragments of a given size fraction. These rock grains, which are an assemblage of minerals, are considered to be submitted to physical weathering only, as the surface in contact with weathering agents is supposed to be negligible. The probability that fragments will break decreases with decreasing size, due to the fact that smaller particles contain fewer defects (Sharmeen and Willgoose 2006). Therefore, they break down to smaller fragments of various sizes according to first-order kinetic reactions. Once these fragments reach the size of 2 mm, they pertain to the fine fraction of the soil and are submitted to both physical and chemical weathering.

19.2.2.3 Physical and Chemical Weathering of the Fine Fraction

The intensity of physical and chemical weathering of the fine fraction is a function of the mineral type. For each of the defined layers, the fine fraction is divided into 1000 classes from 1000 to 1 μm that correspond to the radius of the particles as they are considered spherical. This approach is similar to the one developed by Legros and Pedro (1985). All the individual mineral particles resulting from the physical fractionation of the coarse fraction are first considered to have an initial 1000 μm radius.

The physical weathering of the fine fraction consists of the potential microdivision of a mineral into smaller particles. This microdivision is programmed as a conditional test based on the resistance of minerals to fractionation with a stochastic component. This resistance varies as well according to the size of the particle and its depth in the soil profile. The chemical weathering is assumed to consist in a congruent dissolution: the number of moles of a given mineral that are weathered is the product of its weathering rate constant and its surface area (White et al. 1996). The total surface area of the minerals corresponds here to the sum of the surface area of the spheres of a given mineral that compose a soil layer. Roughness and an internal porosity factor can be implemented to account for the nonspherical shape of the minerals, resulting in an underestimation of their surface area (White et al. 1996). To summarize, the quantity of the primary mineral that is weathered,

if relevant, the quantity of secondary minerals formed and finally the new radius of the weathered primary mineral and its redistribution in the relevant size class are calculated for each class, of each layer and at each time step.

19.2.2.4 Bioturbation

The horizonation of in situ soil profiles results from translocation processes of soil material within the profile. Bioturbation is one of the processes that allows translocation of soil particles between horizons due to animals and plants (Hole 1981) and is addressed in this model. This process can be summarized as the homogenization of the topsoil and material transport between the subsoil and the topsoil (Müller-Lemans and van Dorp 1996). The amount of soil material translocated within the soil varies greatly according to several parameters. The model accounts for the maximum mass of soil translocated up and downwards, using data on surface casting (e.g. 1–5 kg m^{-2} $year^{-1}$, Paton et al. 1995), the coarse fragment content as a limiting factor to bioturbation and the position of the layer in the soil profile, as bioturbation rates decrease exponentially with depth in the soil.

19.2.2.5 Organic Matter Dynamics

Further developments of the initial model have been made considering the organic matter dynamics (Salvador-Blanes et al. 2011). Organic matter dynamics are implemented through the application of a simple one-compartment model (Hénin and Dupuis 1945). Although simple, the model has, for example, been used for modelling the evolution of organic carbon contents at the landscape scale (Walter et al. 2003). This one-compartment model is moreover simple to use for time steps of the order of several decades to centuries. The input parameters to be addressed relate to input of fresh organic carbon to soil, organic carbon incorporation to the soil profile and mineralisation dynamics.

Fresh organic carbon input to soils depends on plant production that itself depends on climate and edaphic parameters. The annual input can therefore be approached by considering it equivalent to net primary productivity (NPP), for which many data exist in the literature. While vegetation is a buffer to the transfer of carbon from the atmosphere to the soil compartment, the model, with time steps of several decades to a century, allows to discard this issue. The simple global Miami model (Leith 1975) that links NPP to mean yearly temperature and rainfall has been used in Pedogen. Organic carbon production being strongly linked to soil moisture and nutrient availability, NPP values have been limited using soil available water content (AWC) as a proxy when rainfall is a limiting factor (annual potential evapotranspiration > rainfall), with threshold values equivalent to the approach in the TRIFFID model (Cox 2001). The soil AWC is calculated at each time step using a PTF linking field capacity and permanent wilting point to several soil properties.

The organic carbon incorporation to soil accounts for the root/shoot ratio (r/s) and the input depth distribution in the soil profile. Constant averaged *r*/*s* values for given biomes are used (Jackson et al. 1996), according to the location of the modelled profile on the earth. The input depth distribution of organic carbon in the soil profile is correlated to the root depth distribution. This distribution varies according to the vegetation communities; it is modelled as a negative exponential decline with depth in the soil (Gale and Grigal 1987), according to the biome where the soil profile is located. Such an approach to constrain the input of OC according to soil depth in a numerical simulation of organic carbon dynamics has already been used by Elzein and Balesdent (1995). The mineralization dynamics is determined by using values of isohumic coefficients and mineralization rates given in the literature (Bayer et al. 2006).

19.2.2.6 Strain

Several processes occur in the soil that lead to strain, e.g. a collapse or a dilation of the soil (increased weathering, bioturbation, arrangement of soil particles into peds, incorporation of organic matter). This has in turn a consequence of the intensity of the processes modelled. To account for these changes, the bulk density of the layers in the soil profile is calculated at each time step according to one of the numerous bulk density PTFs available, which links bulk density to particle size properties, depth in the soil (Tranter et al. 2007).

Pedogen is a simple and 'open' model (additional processes can be incorporated), integrative of many complex pedogenetic processes, that requires few input parameters and can be adapted and implemented in a 2D/3D model as shown in the latter section of this chapter. Recent developments of this model, which incorporate the organic matter dynamics, allow an integration and interaction between processes with very different dynamics (physical and chemical weathering vs. bioturbation/organic matter), allowing to make a link with ecosystem modelling.

19.2.3 The SoilGen Model

Finke and Hutson (2008) followed by Finke (2012) devised a soil profile model called SoilGen. This model can be assimilated to a solute transport model that aims at simulating soil profile development over unconsolidated parent materials, accounting for factors of soil formation (Finke 2012). It is one of the few complete soil evolution models that simulates the changes in soil properties over millennium timescales, taking into account a wide range of processes (Opolot et al. 2015). As the processes occurring in the soil operate at very different timescales, they are modelled in SoilGen according to differing time steps (Finke and Hutson 2008; Finke 2012): milliseconds to hourly time steps (chemical and transport processes),

hourly time steps (heat flow and physical weathering), daily time steps (mineral weathering and organic matter dynamics) and yearly time steps (bioturbation, erosion/sedimentation and fertilization).

The description below is largely based on Opolot et al. (2015) that provides a complete summary of the SoilGen model, which was originally thoroughly described in Finke and Hutson (2008) and Finke (2012).

19.2.3.1 Water, Solute and Heat Transfer

Water, solute and heat transfers are based on the concepts of the LEACH-C code (Hutson 2003). The model solves the Richardson equation for the unsaturated vertical water flow, the convection/dispersion equation of the transfer of solutes and the heat flow equation for temperature distribution. Input parameters such as precipitation and evapotranspiration are corrected according to the local slope and exposition of the modelled profile.

19.2.3.2 Soil Chemical System and Chemical Equilibria

The model chemical system is divided into five phases: solution, precipitated, exchange, organic and unweathered phases. The input of ions to the solution is due to the dissolution of primary minerals, the decomposition of organic matter and external inputs through atmospheric deposition and fertilization. The removal of ions from the soil solution is due to plant uptake, leaching and precipitation. The equilibrium of the soil solution with precipitated and exchange phases is ensured by the application of several solubility laws and rate constants, with calculations at short time steps (Finke and Hutson 2008). The cations are adsorbed onto the solid phase by a Gapon exchange mechanism, and the exchange capacity is defined according to a regression equation combining organic carbon and clay contents, according to Foth and Ellis (1996), modified by a factor matching the initial CEC in the simulated pedon. However, the effect of pH on the CEC is not yet implemented.

19.2.3.3 Weathering Processes

Both physical and chemical weathering processes are described. Physical weathering processes are due to the strain caused by temperature gradients. The reduction in grain size is expressed by the probabilistic break-up of particles, as in Salvador-Blanes et al. (2007). Here, 12 size classes in the fine fraction (<2 mm) are considered. The probabilistic process is implemented as the splitting probability of a particle following a Bernoulli process according to the temperature gradient over a given time interval. However, this splitting process is restricted to fine, unconsolidated material. Chemical weathering is the major source of cations in nonagricultural soils. Here anorthite, chlorite, microcline and albite are considered

as major pools of Ca^{2+}, Mg^{2+}, K^{+} and Na^{+}, respectively. The weathering flux of each of these cations is defined as in Kros (2002) as a function of soil bulk density, the soil layer thickness, the volumic hydrogen concentration, a constant representing the effect of pH on the weathering rate and finally the content of the considered element in the primary mineral. The weathering flux of Al is determined considering a congruent weathering of the aforementioned minerals.

19.2.3.4 Vegetation, Carbon Cycling and Plant Uptake Processes

SoilGen allows the interaction between the soil and the vegetation through annual litter input, carbon cycling and ion uptake, according to four vegetation types (grass/scrub, agriculture, conifers, deciduous wood) (Finke and Hutson 2008). The carbon cycling is simulated according to the RothC-26.3 model (Jenkinson and Coleman 1994).

19.2.3.5 Soil Phase Redistribution Processes

Solid phase redistribution processes are considered through several aspects. The clay migration process is extensively described in the model, combining detachment, dispersion, transportation and deposition processes. The detachment process occurs at the soil surface through the impact of raindrops and is modelled according to Jarvis et al. (1999) modified by Finke (2012). The dispersion process is described both at the surface and within the soil profile, when the solute concentration decreases below a threshold value. Finally, the filtering process, defined as the entrapment of clay particles in small pores, is based on calculated pore water velocities. The bioturbation process is described as an incomplete mixing process, whereas tillage is considered as an extreme bioturbation process within the surface of the soil profile (Finke and Hutson 2008). Other processes, such as erosion or sedimentation, which result in the addition or removal of soil material at the surface of the soil profile and dissolution and precipitation of calcite and gypsum are described as well. However, the collapse and/or dilation through these various processes are not accounted for in the model.

19.2.3.6 Applications of the SoilGen Model

Successful applications of the SoilGen model have been performed, both at the profile and at the landscape scale. At the profile scale, the model was initially implemented to demonstrate its ability to simulate the effect of climate/vegetation/organisms on the soil formation on calcareous loess in Belgium and Hungary (Finke and Hutson 2008). Further developments of the model allowed to test its sensitivity to historic climatic fluctuations in different topographic conditions over a similar parent material (Finke 2012). A quality test of the

model was performed on soil chronosequences developed in marine sediments in Norway (Sauer et al. 2012), which showed that the model fitted well for some soil properties such as particle size distribution, but underestimated some other properties. Discrepancies were analysed, and possible improvements of the model were suggested, such as the description of soil structure formation. At the landscape scale, the model was applied over 96 soil profiles disseminated over 584 km^2 in Northern Belgium (Zwertvaegher et al. (2013). This application allowed to reconstruct soil characteristics (texture, bulk density, organic carbon, calcite content, pH) that fitted well with the current properties and allowed to reconstruct soil characteristic maps at specific points in the past. Finally, the model was applied over 108 soil profiles to test the possible effect of tree uprooting on some soil properties (Finke et al. 2014). The simulations, coupled with regression kriging, allowed to prove that tree uprooting is an important process determining horizon thicknesses.

The SoilGen model is probably the most complete integrated soil profile model to date. It has been tested both at the pedon and landscape scales and confronted to actual soil properties, with good matches for some soil properties and some discrepancies for others. Further developments include the extension to additional primary minerals and elements, to the formation of secondary minerals and to the further development of the interactions between soil and vegetation (Opolot et al. 2015).

19.3 Soil-Landscape Models

19.3.1 Soil-Landscape Models

Recent developments in process modelling focus on mechanistic simulations of soil formation in the landscape based on the principals of mass balance. Huggett's (1975) homomorphic modelling approach can perhaps be regarded as the first representation of soil-landscape evolution modelling. Huggett (1975) proposed to model the evolution of a soil system in a three-dimensional way on the catena scale over millennial timescales, considering also the formation of soil horizons over time. He explained further that concaving contours in a downslope direction should lead to convergent flow lines whereas convex contours should lead to divergent flow lines and that all flow lines should converge in hollows and diverge over spurs. All 'flowlines' should then join in one complex network, based on first- and second-order streamlines.

The simplest model of soil-landscape evolution that has been formulated implements the change in elevation as a function of material transport (Stockmann et al. 2011):

$$\frac{\partial z}{\partial t} = -\nabla q_s \tag{19.16}$$

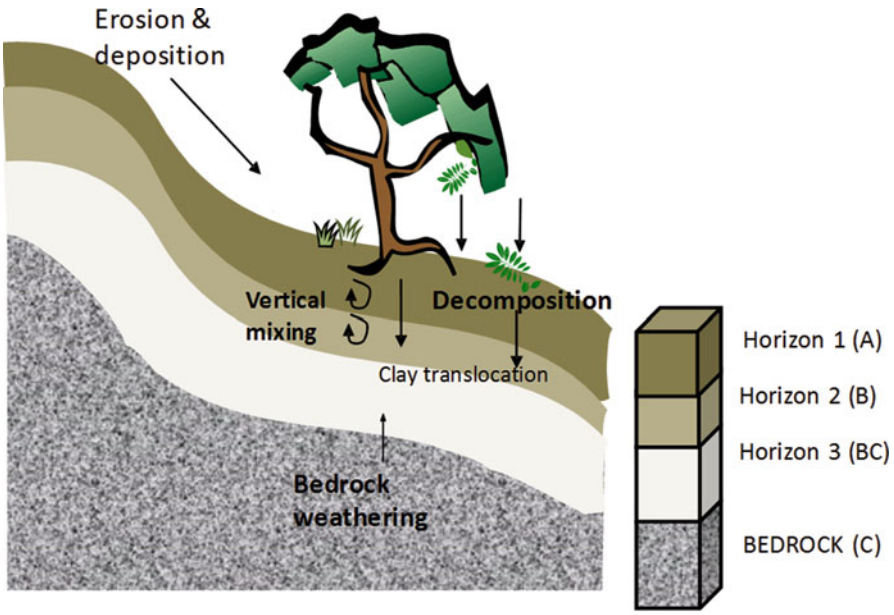


Fig. 19.11 Schematic representation of a simple soil-landscape model

where z is the elevation, t is time, q_s is the material flux and ∇ is a partial derivative vector. Dietrich et al. (1995) and later Heimsath et al. (1997) introduced soil into the continuity equation of mass transport along a hillslope (Eq. 19.16):

$$\rho_s \frac{\partial h}{\partial t} = -\rho_r \frac{\partial e}{\partial t} - \nabla q_s \tag{19.17}$$

where h is soil thickness, ρ_s and ρ_r are the bulk densities of soil and rocks, e is the elevation of the bedrock-soil interface, t is time and q_s is the material flux in the horizontal direction (Fig. 19.11).

Minasny and McBratney (1999) and following on Minasny and McBratney (2001) used some of the basic concepts described in Huggett (1975) and Heimsath et al. (1997) to introduce a two-dimensional rudimentary mechanistic pedogenetic model. Based on a digital elevation model (DEM), pedogenesis is simulated by a combination of several sub-models: (1) physical weathering starting from bedrock employing the rate of exponential decline of soil production with increasing soil thickness, (2) chemical weathering represented as a negative exponential function of both soil thickness and time and (3) movement of soil material as characterized by a diffusion transport model. The upscaling result of such an analysis is illustrated in Fig. 19.12, whereby, after 10,000 years, soil accumulation is predominant in the gullies compared with the ridges, where soil erodes.

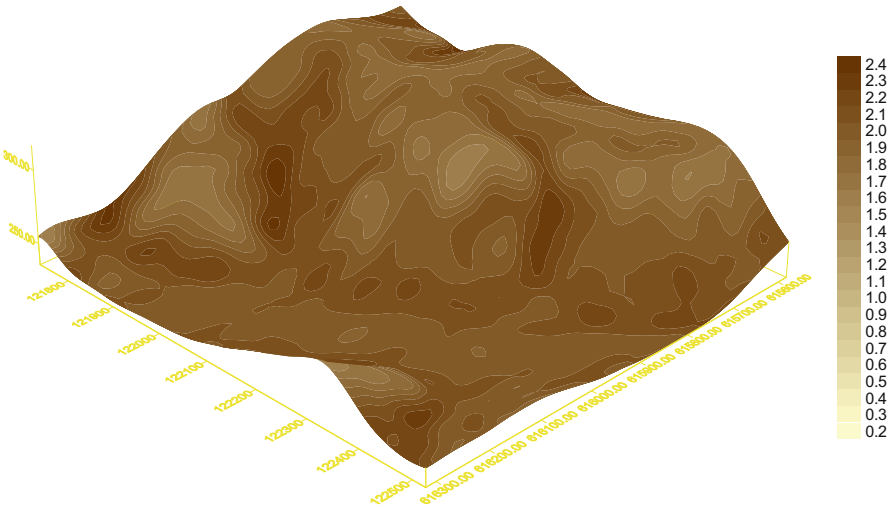


Fig. 19.12 Simulated soil formation in a landscape after 10,000 years, shading represents soil thickness (*m*) (After Minasny and McBratney 2001)

19.3.2 Linking Landscape-Scale Models and Soil Profile-Scale Models

While current soil models allow a detailed simulation at the profile level as discussed in Sect. 19.2, they lack horizontal integration, with models from the geomorphology community, where the most advanced LEMs include a wide array of erosion processes and accurate description of erosion-deposition processes, but lack attention to vertical soil formation processes.

One of the first attempts to such an integrated model was the model for integrated soil development (MILESD) by Vanwalleghem et al. (2013). MILESD is a four-layer model with five texture classes. It includes the main soil-forming processes: physical weathering, chemical weathering, clay migration and neoformation, bioturbation and carbon cycling. Landscape evolution is represented by concentrated flow erosion and creep, allowing for selective transport and deposition and with negative feedbacks from stoniness and vegetation on the erosion rates. The model was developed so that soil formation and evolution could be modelled with enough detail while at the same time reducing runtime to allow landscape-scale simulations. Potential drawbacks of this simplified approach include the fact that the soil solution and soil chemistry are not included explicitly. Therefore, all soil-forming processes need to be calibrated on a site-specific basis.

Vanwalleghem et al. (2013) applied MILESD to a 6.25 km^2 study area in Werrikimbe National Park (NSW, Australia) where it was validated against field profile data (Fig. 19.13). The results showed that trends in soil thickness were

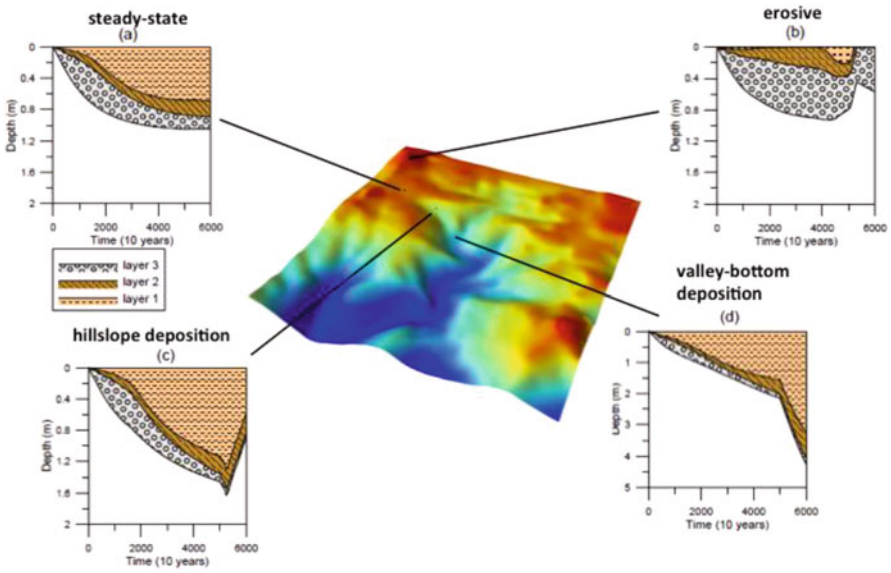


Fig. 19.13 Evolution of soil horizons in four different landscape positions, from erosive over stable (steady-state) to depositional for a constant erosion rate during the first 50,000 years followed by higher erosion period during the last 10,000 years

predicted well along a catena. Soil texture and bulk density could be predicted reasonably well, with errors in the order of 10%. Figure 19.13 clearly shows the potential of these types of integrated models. The evolution of soil horizons over time is shown for different positions in the landscape, ranging from erosive over stable to depositional. The results show a model run to 60,000 years where a constant erosion rate was followed by a higher erosion rate during the last 10,000 years. It can be seen how important differences along the catena emerge with a stable soil profile on the plateau (a), an erosive hillslope where the top two horizons disappear altogether (b), a hillslope deposition site that eventually erodes when erosion rates increase at the end of the simulation (c) and a deep depositional soil in the valley bottom (d).

Figure 19.14 shows the result of soil formation-erosion interactions on the organization and evolution of soil properties in the entire catchment. Increasing erosion rate and increasing age of the landscape both lead to an increase in semivariance, which implies a higher spatial variability over time. In this particular context, it seems that erosion and deposition are key drivers of soilscape heterogeneity as the semivariogram responds strongly to increasing erosion rates.

Temme and Vanwalleghem (2016) presented a new soil-landscape model, called LORICA, that integrated MILESD with the existing landscape evolution model LAPSUS. The coupling with a more advanced landscape evolution model allows taking into account different erosion processes, e.g. landsliding, which improves the applicability of the soilscape model to a wider range of environments. With respect

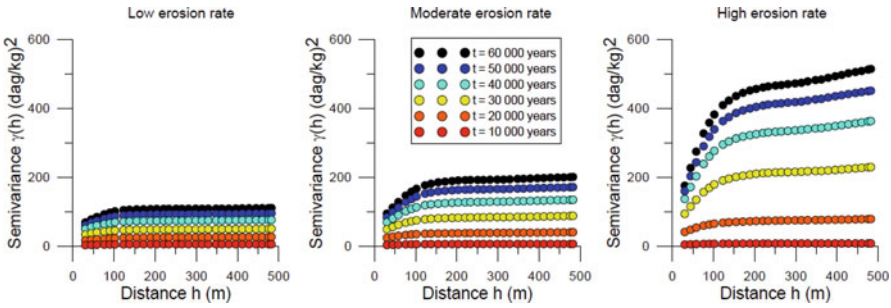


Fig. 19.14 Evolution of the spatial variability of clay content in topsoil (0–0.2 m) as a function of time and erosion rate. Erosion rate increases from left to right

to the representation of soil genesis, the main novelty is the elimination of the fixed layer limitation. LORICA uses an advanced multilayer approach, where new layers are generated or existing layers are joined on a need-be basis. van der Meij et al. (2015) applied LORICA to simulate the development of Arctic soils, which allowed to single out the importance of aeolian deposition.

Another suite of advanced, fully three-dimensional soilscape models was developed by Cohen et al. (2010, 2015) and Welivitiya et al. (2016). Both mARM3D (Cohen et al. 2010) and its extension, mARM5D (Cohen et al. 2015), simulate the evolution of soil particle size distribution via physical weathering, profile depth as a function of weathering, aeolian deposition and diffusive and fluvial sediment transport. With the model SSSPAM, the latest version in this model family, Welivitiya et al. (2016) performed an extensive sensitivity analysis, generalizing the physical dependence of the relationship between contributing area, local slope, and the surface soil particle size distribution.

It is clear from the previous discussion that the existing soilscape models will need to be developed and tested further. The issue of the ideal number of soil horizons to consider is not trivial. Although, ideally, an infinite number of horizons assure a full representation of the profile's complexity, increasing the number of horizons in any model will imply increasing computation time significantly. Moreover, soil scientists often record field data with a number of limited soil horizons. This implies that in a validation exercise, the results of a model with many horizons have to be 'converted back' into a profile with less horizons. To that extent, the four-layer approach of the simple MILESD model corresponds to the way soil profiles are described in the field. Several processes are currently represented poorly or not at all in soilscape models. Soil chemical weathering is probably the most critical process that is currently represented in an overly simplified manner. Several studies, both in the laboratory (e.g. Maher 2010) and in field conditions (e.g. Schoonejans et al. 2016), have shown the dependency of chemical weathering on soil hydrological fluxes. Pore water dynamics are currently not explicitly represented in soilscape models. This limitation could be solved by coupling more detailed point-based models, such as SoilGen, to landscape evolution

models, as proposed by Opolot et al. (2015). This would be increasingly beneficial in terms of computational resources in landscapes where landscape evolution is several orders of magnitude faster compared to soil formation processes or vice versa, and the exchange of input-output data between models is not necessary during every model time step.

So clearly, much work is required to better model the effect of climate and organisms on the soil's chemistry, mineralogy, physics and biology (Chadwick et al. 2003). It is also essential to integrate soil processes, which are usually only represented at a profile scale, with landscape processes (Viaud et al. 2010).

References

Ahnert F (1977) Some comments on the quantitative formulation of geomorphological processes in a theoretical model. Earth Surface Processes 2:191–201

Aitken MJ (1998) An introduction to optical dating. The dating of quaternary sediments by the use of photon-stimulated luminescence. Oxford University Press, Oxford

Barrett LR, Schaetzl RJ (1992) An examination of podzolization near Lake Michigan using chronofunctions. Can J Soil Sci 72:527–541

Bayer C, Lovato T, Dieckow J, Zanatta JA, Mielniczuk J (2006) A method for estimating coefficients of soil organic matter dynamics based on long-term experiments. Soil Tillage Res 91:217–226

Brantley SL, Megonigal JP, Scatena FN, Balogh-Brunstad Z, Barnes RT, Bruns MA, Van Cappellen P, Dontsova K, Hartnett HE, Hartshorn AS, Heimsath A, Herndon E, Jin L, Keller CK, Leake JR, Mcdowell WH, Meinzer FC, Mozdzer TJ, Petsch S, Pett-Ridge J, Pregitzer KS, Raymond PA, Riebe C, Shumaker S, Sutton-Grier A, Walter R, Yoo K (2011) Twelve testable hypotheses on the geobiology of weathering. Geobiology 9:140–165

Brimhall GH, Dietrich WE (1987) Constitutive mass balance relations between chemical composition, volume, density, porosity, and strain in metasomatic hydrochemical systems: results on weathering and pedogenesis. Geochim Cosmochim Acta 51:567–587

Burke BC, Heimsath AM, White AF (2007) Coupling chemical weathering with soil production across soil-mantled landscapes. Earth Surf Process Landf 32:853–873

Burke BC, Heimsath AM, Dixon JL, Chappell J, Yoo K (2009) Weathering the escarpment: chemical and physical rates and processes, south-eastern Australia. Earth Surf Process Landf 34:768–785

Chadwick OA, Gavenda RT, Kelly EF, Ziegler K, Olson CG, Elliott WC, Hendricks DM (2003) The impact of climate on the biogeochemical functioning of volcanic soils. Chem Geol 202:195–223

Chmeleff J, von Blanckenburg F, Kossert K, Jakob D (2009) Determination of the ^{10}Be half-life by multi collector ICP-mass spectrometry and liquid scintillation counting. Goldschmidt abstracts 2009 – C. Geochim Cosmochim Acta 73:A221–A221

Cohen S, Willgoose G, Hancock G (2010) The mARM3D spatially distributed soil evolution model: three-dimensional model framework and analysis of hillslope and landform responses. J Geophys Res Earth 115:F04013

Cohen S, Willgoose G, Svoray T, Hancock G, Sela S (2015) The effects of sediment transport, weathering, and aeolian mechanisms on soil evolution. J Geophys Res Earth 120:260–274

Cox P (2001) Description of the "TRIFFID" dynamic global vegetation model. Technical note 24. Hadley Centre, Met Office, London

Dietrich WE, Reiss R, Hsu M-L, Montgomery DR (1995) A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrol Process 9:383–400

Dijkerman JC (1974) Pedology as a science: the role of data, models and theories in the study of natural soil systems. Geoderma 11:73–93
Dixon JL, Heimsath AM, Amundson R (2009) The critical role of climate and saprolite weathering in landscape evolution. Earth Surf Process Landf 34:1507–1521
Dörr H (1995) Application of ^{210}Pb in soils. J Paleolimnol 13:157–168
Dosseto A, Turner SP, Chappell J (2008) The evolution of weathering profiles through time: new insights from uranium-series isotopes. Earth Planet Sci Lett 274:359–371
Dunne J, Elmore D, Muzikar P (1999) Scaling factors for the rates of production of cosmogenic nuclides for geometric shielding and attenuation at depth on sloped surfaces. Geomorphology 27:3–11
Egli M, Dahms D, Norton K (2014) Soil formation rates on silicate parent material in alpine environments: different approaches-different results? Geoderma 213:320–333
Elzein A, Balesdent J (1995) Mechanistic simulation of vertical distribution of carbon concentrations and residence times in soils. Soil Sci Soc Am J 59:1328–1335
Finke PA (2012) Modeling the genesis of Luvisols as a function of topographic position in loess parent material. Quat Int 265:3–17
Finke PA, Hutson JL (2008) Modelling soil genesis in calcareous loess. Geoderma 145:462–479
Finke PA, Vanwalleghem T, Opolot E, Poesen J, Deckers J (2014) Estimating the effect of tree uprooting on variation of soil horizon depth by confronting pedogenetic simulations to measurements in a Belgian loess area. J Geophys Res-Earth Surf 118:1–16
Foth HD, Ellis BG (1996) Soil fertility, 2nd edn. CRC Press, Lewis
Furbish DJ, Fagherazzi S (2001) Stability of creeping soil and implications for hillslope evolution. Water Resour Res 37:2607–2618
Gabet EJ, Mudd SM (2010) Bedrock erosion by root fracture and tree throw: a coupled biogeomorphic model to explore the humped soil production function and the persistence of hillslope soils. J Geophys Res Earth 115:F04005
Gabet EJ, Reichman OJ, Seabloom EW (2003) The effects of bioturbation on soil processes and sediment transport. Annu Rev Earth Planet Sci 31:249–274
Gale MR, Grigal DF (1987) Vertical root distributions of northern tree species in relation to successional status. Can J For Res 17:829–834
Gilbert GK (1877) Report on the geology of the Henry Mountains (Utah). United States Geological Survey, Washington, DC
Granger DE, Muzikar PF (2001) Dating sediment burial with in situ-produced cosmogenic nuclides: theory, techniques, and limitations. Earth Planet Sci Lett 188:269–281
Green EG, Dietrich WE, Banfield JF (2006) Quantification of chemical weathering rates across an actively eroding hillslope. Earth Planet Sci Lett 242:155–169
Hay RL (1960) Rate of clay formation and mineral alteration in a 4000-year-old volcanic ash soil on St. Vincent, B.W.I. Am J Sci 258:354–368
He Q, Walling DE (1996) Interpreting particle size effects in the adsorption of ^{137}Cs and unsupported ^{210}Pb by mineral soils and sediments. J Environ Radioact 30:117–137
Heimsath AM, Dietrich WE, Nishiizumi K, Finkel RC (1997) The soil production function and landscape equilibrium. Nature 388:358–361
Heimsath AM, Dietrich WE, Nishiizumi K, Finkel RC (1999) Cosmogenic nuclides, topography, and the spatial variation of soil depth. Geomorphology 27:151–172
Heimsath AM, Fink D, Hancock GR (2009) The 'humped' soil production function: eroding Arnhem Land, Australia. Earth Surf Process Landf 34:1674–1684
Hénin S, Dupuis M (1945) Essai de bilan de la matière organique du sol. Ann Agronomiques 11:17–29
Heuvelink GBM, Webster R (2001) Modelling soil variation: past, present and future. Geoderma 100:269–301
Hole FD (1981) Effects of animals on soil. Geoderma 25:75–112
Hoosbeek MR (1994) Towards the quantitative modeling of pedogenesis: a review – reply – pedology beyond the soil-landscape paradigm: pedodynamics and the connection to other sciences. Geoderma 63:303–307

Huggett RJ (1975) Soil landscape systems: a model of soil genesis. Geoderma 13:1–22
Huggett RJ (1998) Soil chronosequences, soil development, and soil evolution: a critical review. Catena 32:155–172
Humphreys GS, Wilkinson MT (2007) The soil production function: a brief history and its rediscovery. Geoderma 139:73–78
Hutson JL (2003) LEACHM e a process-based model of water and solute movement, transformations, plant uptake and chemical reactions in the unsaturated zone. Version 4. Research series no R03-1 (Dept. of Crop and Soil Sciences, Cornell University, Ithaca, NY). J Soil Sci 36:97–121
Jackson RB, Canadell J, Ehleringer JR, Mooney HA, Sala OE, Schulze ED (1996) A global analysis of root distributions for terrestrial biomes. Oecologia 108:389–411
Jagercikova M, Evrard O, Balesdent J, Lefèvre I, Cornu S (2014) Modeling the migration of fallout radionuclides to quantify the contemporary transfer of fine particles in Luvisol profiles under different land uses and farming practices. Soil Tillage Res 140:82–97
Jarvis NJ, Villholth KG, Ulén B (1999) Modelling particle mobilization and leaching in macroporous soil. Eur J Soil Sci 50:621–632
Jenkinson DS, Coleman K (1994) Calculating the annual input of organic matter to soil from measurements of total organic carbon and radiocarbon. Eur J Soil Sci 45:167–174
Jenny H (1941) Factors of soil formation. A system of quantitative pedology. McGraw-Hill Book Company, New York
Kaste JM, Heimsath AM, Bostick BC (2007) Short-term soil mixing quantified with fallout radionuclides. Geology 35:243–246
Kirkby MJ (1977) Soil development models as a component of slope models. Earth Surface Process 2:203–230
Kirkby MJ (1985) A basis for soil profile modelling in a geomorphic context. J Soil Sci 36:97–121
Kitayama K, Edward AGS, Drake DR, Mueller-Dombois D (1997a) Fate of a wet montane forest during soil ageing in Hawaii. J Ecol 85:669–679
Kitayama K, Schuur EAG, Drake DR, Mueller-Dombois D (1997b) Fate of a wet montane forest during soil ageing in Hawaii. J Ecol 85:669–679
Kros J (2002) Evaluation of biogeochemical models at local and regional scale. PhD thesis. Wageningen University, Wageningen, The Netherlands
Lal D (1991) Cosmic ray labeling of erosion surfaces: in situ nuclide production rates and erosion models. Earth Planet Sci Lett 104:424–439
Larsen IJ, Almond PC, Eger A, Stone JO, Montgomery DR, Malcolm B (2014) Rapid soil production and weathering in the southern alps, New Zealand. Science 343:637–640
Legros JP, Pedro G (1985) The causes of particle-size distribution in soil profiles derived from crystalline rocks, France. Geoderma 36:15–25
Leith H (1975) Modelling the primary productivity of the world. In: Leith H, Whittaker RH (eds) Primary productivity of the biosphere. Springer-Verlag, Berlin
Levine ER, Knox RG (1994) A comprehensive framework for modeling soil genesis. In: Bryant RB, Arnold RW (eds) Quantitative modeling of soil forming processes. Soil Science Society of America, Madison, pp 77–89. SSSA Special Publication No 39
Ludwig W, Probst JL (1998) River sediment discharge to the oceans: present-day controls and global budget. Am J Sci 298:265–295
Mabit L, Benmansour M, Walling DE (2008) Comparative advantages and limitations of the fallout radionuclides ^{137}Cs, $^{210}Pb_{ex}$ and ^{7}Be for assessing soil erosion and sedimentation. J Environ Radioact 99:1799–1807
Maher K (2010) The dependence of chemical weathering rates on fluid residence time. Earth Planet Sci Lett 294:101–110
Minasny B, McBratney AB (1999) A rudimentary mechanistic model for soil production and landscape development. Geoderma 90:3–21
Minasny B, McBratney AB (2001) A rudimentary mechanistic model for soil production and landscape development II. A two-dimensional model incorporating chemical weathering. Geoderma 103:161–179

Minasny B, McBratney AB (2006) Mechanistic soil-landscape modelling as an approach to developing pedogenetic classifications. Geoderma 133:138–149
Minasny B, Finke P, Stockmann U, Vanwalleghem T, McBratney AB (2015) Resolving the integral connection between pedogenesis and landscape evolution. Earth Sci Rev 150:102–120
Montgomery DR (2007) Soil erosion and agricultural sustainability. Proc Natl Acad Sci 104:13268–13272
Müller-Lemans H, van Dorp F (1996) Bioturbation as a mechanism for radionuclide transport in soil: relevance of earthworms. J Environ Radioact 31:7–20
Nishiizumi K, Kohl CP, Arnold JR, Klein J, Fink D, Middleton R (1991) Cosmic ray produced ^{10}Be and ^{26}Al in Antarctic rocks: exposure and erosion history. Earth Planet Sci Lett 104:440–454
NRC (2010) Landscapes on the edge. New horizons for research on Earth's surface. The National Academy Press, Washington, DC
Opolot E, Yu YY, Finke PA (2015) Modeling soil genesis at pedon and landscape scales: achievements and problems. Quat Int 376:34–46
Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation and confirmation of numerical models in the earth sciences. Science 263:641–646
Paton TR, Humphreys GS, Mitchell PB (1995) Soils: a new global view. University College London Press, London
Portenga EW, Bierman PR (2011) Understanding Earth's eroding surface with ^{10}Be. GSA Today 21:4–10
Riebe CS, Kirchner JW, Finkel RC (2003) Long-term rates of chemical weathering and physical erosion from cosmogenic nuclides and geochemical mass balance. Geochim Cosmochim Acta 67:4411–4427
Riebe CS, Kirchner JW, Finkel RC (2004a) Erosional and climatic effects on long-term chemical weathering rates in granitic landscapes spanning diverse climate regimes. Earth Planet Sci Lett 224:547–562
Riebe CS, Kirchner JW, Finkel RC (2004b) Sharp decrease in long-term chemical weathering rates along an altitudinal transect. Earth Planet Sci Lett 218:421–434
Salvador-Blanes S, Minasny B, McBratney AB (2007) Modelling long term in-situ soil profile evolution – application to the genesis of soil profiles containing stone layers. Eur J Soil Sci 58:1535–1548
Salvador-Blanes S, Minasny B, McBratney AB (2011) Modelling soil formation at the profile scale. European Geosciences Union General Assembly 2011. Geophys Res Abstr 13:EGU2011–EG13012
Sauer D, Schellmann G, Stahr K (2007) A soil chronosequence in the semi-arid environment of Patagonia (Argentina). Catena 71:382–393
Sauer D, Finke PA, Schülli-Maurer I, Sperstad R, Sørensen R, Høeg HI, Stahr K (2012) Testing a soil development model against southern Norway soil chronosequences. Quat Int 265:18–31
Schaetzel RJ, Barrett LR, Winkler JA (1994) Choosing models for soil chronofunctions and fitting them to data. Eur J Soil Sci 45:219–232
Schoonejans J, Vanacker V, Opfergelt S, Ameijeiras-Mariño Y, Christl M (2016) Kinetically limited weathering at low denudation rates in semiarid climatic conditions. J Geophys Res Earth 121:336–350
Sharmeen S, Willgoose GR (2006) The interaction between armouring and particle weathering for eroding landscapes. Earth Surf Process Landf 31:1195–1210
Shouse M, Phillips J (2016) Soil deepening by trees and the effects of parent material. Geomorphology 269:1–7
Sommer M, Gerke HH, Deumlich D (2008) Modelling soil landscape genesis: a "time split" approach for hummocky agricultural landscapes. Geoderma 145:480–493
Sparks DL (1995) 7 – Kinetics of soil chemical processes. Environmental soil chemistry. Academic, Boston, pp 159–185
Stevens PR, Walker TW (1970) The chronosequence concept and soil formation. Q Rev Biol 45:333–350

Stockmann U (2010) Quantifying processes of pedogenesis. A field study situated in the Werrikimbe National Park in south-eastern Australia. Faculty of Agriculture, Food and Natural Resources. The University of Sydney, p 234

Stockmann U, Minasny B, McBratney A (2011) Quantifying processes of pedogenesis. Adv Agron 113:1–74

Stockmann U, Minasny B, Pietsch TJ, McBratney AB (2013) Quantifying processes of pedogenesis using optically stimulated luminescence. Eur J Soil Sci 64:145–160

Stockmann U, Minasny B, McBratney AB (2014) How fast does soil grow? Geoderma 216:48–61

Stone JO (2000) Air pressure and cosmogenic isotope production. J Geophys Res 105:23,753–723,759

Suresh PO, Dosseto A, Hesse PP, Handley HK (2013) Soil formation rates determined from uranium-series isotope disequilibria in soil profiles from the southeastern Australian highlands. Earth Planet Sci Lett 379:26–37

Sverdrup H, Warfvinge P (1993) Calculating field weathering rates using a mechanistic geochemical model – profile. Appl Geochem 8:273–283

Syvitski JPM, Vorosmarty CJ, Kettner AJ, Green P (2005) Impact of humans on the flux of terrestrial sediment to the global coastal ocean. Science 308:376–380

Temme AJAM, Vanwalleghem T (2016) LORICA – a new model for linking landscape and soil profile evolution: development and sensitivity analysis. Comput Geosci 90(Part B):131–143

Tranter G, Minasny B, McBratney AB, Murphy B, McKenzie NJ, Grundy M (2007) Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use Manag 23:437–443

van der Meij WM, Temme AJAM, de Kleijn CMFJJ, Reimann T, Heuvelink GBM, Zwoliński Z, Rachlewicz G, Rymer K, Sommer M (2015) Arctic soil development on a series of marine terraces on Central Spitsbergen, Svalbard: a combined geochronology, fieldwork and modelling approach. SOIL Discuss 2015:1345–1391

Van Wambeke A (1972) Mathematical expression of eluviation-illuviation processes and the computation of the effects of clay migration in homogeneous soil parent materials. J Soil Sci 23:325–332

Van Wambeke A (1976) A mathematical model for the differential movement of two soil constitutents into illuvial horizons: application to clay ratios in argillic horizons. J Soil Sci 27:111–120

Vanwalleghem T, Stockmann U, Minasny B, McBratney AB (2013) A quantitative model for integrating landscape evolution and soil formation. J Geophys Res Earth 118:1–17

Viaud V, Angers DA, Walter C (2010) Towards landscape-scale modeling of soil organic matter dynamics in agroecosystems. Soil Sci Soc Am J 74:1–14

Wagenet RJ, Hutson JL, Bouma J (1994) Modeling water and chemical fluxes as driving forces of pedogenesis. In: Bryant RB, Arnold RW (eds) Quantitative modeling of soil forming processes. Soil Science Society of America, Madison, pp 17–35. SSSA Special Publication No 39

Walter C, Viscarra-Rossel RA, McBratney AB (2003) Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. Soil Sci Soc Am J 67:1477–1486

Welivitiya WDDP, Willgoose GR, Hancock GR, Cohen S (2016) Exploring the sensitivity on a soil area-slope-grading relationship to changes in process parameters using a pedogenesis model. Earth Surf Dynam Discuss 2016:1–43

Wells T, Willgoose GR, Hancock GR (2008) Modeling weathering pathways and processes of the fragmentation of salt weathered quartz-chlorite schist. J Geophys Res Earth 113:F01014

White AF, Blum AE (1995) Effects of climate on chemical weathering in watersheds. Geochim Cosmochim Acta 59:1729–1747

White AF, Blum AE, Schulz MS, Bullen TD, Harden JW, Peterson ML (1996) Chemical weathering rates of a soil chronosequence on granitic alluvium: I. Quantification of mineralogical and surface area changes and calculation of primary silicate reaction rates. Geochim Cosmochim Acta 60:2533–2550

Wilkinson MT, Humphreys GS (2005) Exploring pedogenesis via nuclide-based soil production rates and OSL-based bioturbation rates. Aust J Soil Res 43:767–779

Wilkinson BH, McElroy BJ (2007) The impact of humans on continental erosion and sedimentation. Geol Soc Am Bull 119:140–156

Yaalon DH (1975) Conceptual models in pedogenesis: can soil-forming functions be solved? Geoderma 14:189–205

Yoo K, Amundson R, Heimsath AM, Dietrich WE, Brimhall GH (2007) Integration of geochemical mass balance with sediment transport to calculate rates of soil chemical weathering and transport on hillslopes. J Geophys Res 112:1–15

Zapata F (2003) The use of environmental radionuclides as tracers in soil erosion and sedimentation investigations: recent advances and future developments. Soil Tillage Res 69:3–13

Zwertvaegher A, Finke P, De Smedt P, Gelorini V, Van Meirvenne M, Bats M, De Reu J, Antrop M, Bourgeois J, De Maeyer P, Verniers J, Crombe P (2013) Spatio- temporal modeling of soil characteristics for soilscape reconstruction. Geoderma 207-208:166–179

Part VII
Applications of Pedometrics

"Consider what each soil will bear, and what each refuses."

Virgil

Pedometrics has been considered a rather theoretical and methodological approach to soil science. This final section presents a subsample of applications beyond research. We recognise how the pedometric approach can be applied to agricultural, environmental and wider problems.

Among the agricultural issues considered is precision agriculture, where information on soil variation can be put directly to work on optimising management, particularly inputs. From an environmental point of view, we are interested in soil carbon, and soil change more generally, so we develop methods for improved mapping, auditing and monitoring.

A key application of pedometrics is any endeavour that requires sampling in space or time or both. Several chapters reflect on sampling particularly in relation to soil evaluation and change.

Chapter 20
Site-Specific Crop Management

Brett Whelan

"naturam expelles furca, tamen usque recurret. 'You may drive out nature with a pitchfork, but she will always return'".

Horace, *Epistles* i. x. 24, 65BC – 8BC

20.1 Introduction

This chapter takes a look at the impetus and main approaches for a crop management system that utilises high-resolution spatial and temporal data derived mainly from proximal and remote sensors. With the development of pedometrics and the associated tools (e.g. kriging, multivariate techniques) providing the ability to describe and understand soil variability in space and time (well explored in Chaps. 10, 11, 12, 13, 14 and 15), the concept of Precision Agriculture was born in the early 1990s (Robert 1993) as a practical crop management response to the impact of the described soil variation on crop production.

The Precision Agriculture (PA) concept has developed into a philosophy of managing variability within agricultural industries to improve profitability and/or environmental impact in the short and long term, while minimising risk (Whelan and McBratney 2000). The philosophy is now relevant to plant and animal agricultural enterprises. Site-specific crop management (SSCM) is a form of PA that relates to crop production, whereby decisions on resource application and agronomic practices are improved to better match soil and crop requirements as they vary in the field. In practice, it creates the opportunity to increase the number of (correct) decisions per hectare/per season made in the business of crop management (McBratney et al. 2005). SSCM is a logical step in the evolution of agricultural

B. Whelan (✉)
Precision Agriculture Laboratory, Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW 2006, Australia
e-mail: brett.whelan@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_20

management systems towards increased efficiency of input use relative to production quantity, minimised waste and improved product quality, with potential flow-on benefits for improving marketability through digital records of management.

In a business sense these management decisions could be made at the scale of the farm, field, within-field or even individual plants. Ultimately, the heart of SSCM is the use of (and need for) fine-scale information on spatial (in the range of metres to low tens of metres) and temporal variation in the production output/quality and critical resource/terrain/environmental attributes of a cropping system. Initial observations at a fine scale can be aggregated at various scales to inform decisions at the various scales in the farm business.

Differentially corrected Global Navigation Satellite System (GNSS) receivers and harvester-mounted crop yield and quality monitoring systems can provide the data at the required spatial scale on actual production and terrain variability. Optical and thermal remote and proximal sensing can provide data on crop production surrogates. A number of proximal soil sensing systems are commercially available that can match the survey scale of the production information (e.g. electromagnetic induction, electrical resistivity and gamma radiometrics). These sensors provide information on secondary soil properties (apparent electrical conductivity (ECa), or gamma emissions) that have proven useful in combining with the productivity and terrain data to direct physical site sampling for primary causes and explanations of the variability in production (Taylor et al. 2007). Examples of the data obtainable from the application of these systems are shown in Fig. 20.1.

20.2 Goals and General Strategy for Implementing SSCM

SSCM aims to use knowledge of the importance of changes in soil properties and other factors on crop production and yield potential to better identify and understand the reasons for observed changes in growth/yield within a field. The goal is then to improve decision-making about the use of inputs such as fertilisers, agrochemicals, soil ameliorants, water and energy to better match any spatial and temporal changes in the requirements of the soil and crop. A better match should mean that inputs are used efficiently, business profits are maximised and production risk and waste are minimised.

The way that the important soil properties (soil texture, structure, depth, pH, water, nutrients, organic matter), weed, pest and disease loads, terrain and past management interact and vary across each field on each farm is unique. The spatial and temporal scales at which data on these properties is gathered will impact on the scale at which decisions can be made. These decisions may be at the farm level (i.e. what fields are variable enough to warrant further investigation), the field level (what is the actual optimal field average application) and the within-field level (should different rates be applied within the one field).

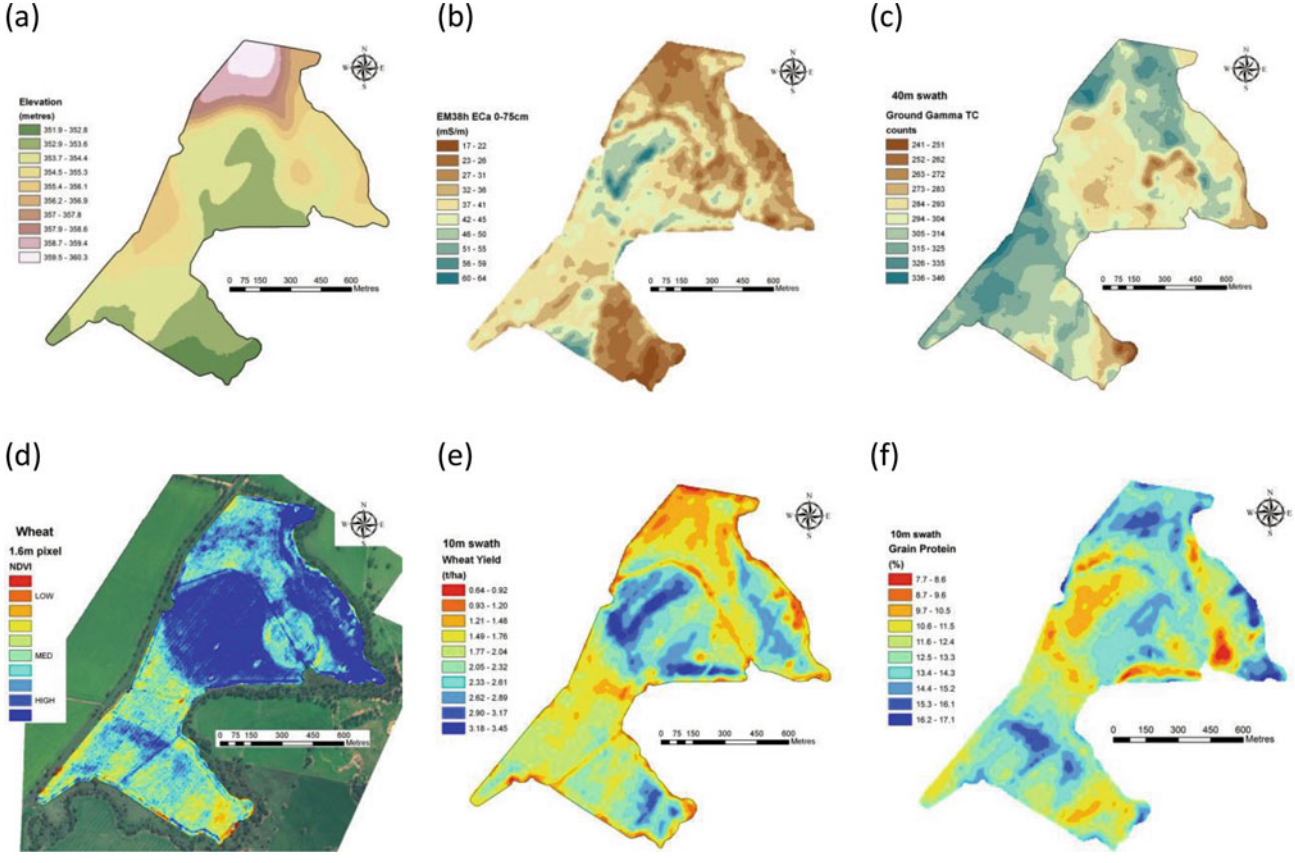


Fig. 20.1 Examples of fine-scale, sensor-derived data commercially available for use in SSCM: (**a**) elevation, (**b**) soil apparent electrical conductivity in the 0–0.75 m profile, (**c**) total gamma emission from the top 0–0.5 m profile, (**d**) normalised difference vegetation index derived from optical reflectance, (**e**) wheat grain yield from an on-harvester yield monitor, (**f**) wheat grain protein from an on-harvester sensor

20.2.1 Variable-Rate Application (VRA) of Inputs

The variable-rate application (VRA) of inputs as a practical response to variation in production potential requires that input rates be adjusted to match the identified changing local requirements. VRA can be aimed at any rate-based operations that influence crop yield and can be done with the help of variable-rate technology (VRT) or by using manual switching or multi-pass applications. Using VRT makes the job less stressful and allows more sophisticated positioning and rate adjustments. The main target operations for VRA in grain cropping are:

- Fertiliser
- Soil ameliorants (lime, gypsum)
- Agrochemicals (herbicide, insecticide, fungicide)
- Sowing
- Tillage
- Irrigation

Ideally, the use of VRA should optimise both the economic and environmental outcomes from a cropped field. In Australia, the economic considerations dominate as there is little regulatory auditing on the levels of approved chemicals in the agricultural environment. In many other countries, regulations make the environmental outcome crucial (Van Grinsven et al. 2016). However, most crop producers are well aware that maintaining a healthy environment is important for sustainability and economic success.

The potential benefits of VRA are generally higher when:

- The amount of spatial variation is larger.
- The pattern of spatial variability tends more towards coherent patches. This usually means fewer rate changes are required.
- The pattern of variability is driven by spatial rather than temporal factors, so it is likely to be relatively stable from season to season and easier to formulate VRA plans.
- The unit cost of input is high relative to the price paid for the crop.

Because of the site-specific nature of this variation, the best information for determining SSCM options for each farm/field will undoubtedly come from within its own boundaries. However, with the advent of improvements in data capture, storage and distribution, research is exploring the possibilities of widening the data usage to include local farm/farms and regional and national off-farm data. This potential advance will be discussed in Sect. 20.4.

20.2.2 A Generic Approach to SSCM

It is important to sort out what the best 'average' (or single treatment) management options are for fields and crops before considering the need/requirements for variably applying inputs. If the current single-rate treatments for fields on a farm over- or underestimate requirements to get the best output at the whole-field or farm scale, then substantial improvements in efficiencies, profit and risk management will be gained by rectifying management at this scale. This can be achieved by the following:

- Identifying any large-scale soil pH and sodicity/salinity issues. Soil samples, taken from across the farm, should be analysed to correctly identify any required treatment.
- Examining and improving, if required, the effectiveness of weed control strategies.
- Elevation mapping using GNSS will allow field boundaries and working directions to be altered to improve water management if necessary.
- Using average yield information gathered for each crop to re-examine crop yield goals on the farm and ensure average nutrient application is suitable. Yield monitors can help here, especially if historical records are missing or the concurrent use of off-farm delivery and on-farm storage makes linking tonnages to specific fields difficult.
- Employing vehicle navigation aids (guidance or autosteer) which offer immediate benefits through reducing the overlap during fertiliser, sowing and chemical operations and providing improvements in soil structure. Autosteer systems can be used to sow into the inter-row between last year's stubble to reduce the incidence of root diseases, etc.

Once any alterations are put in place to optimise the 'average' or uniform agronomy on a farm, the aim should then be to test whether SSCM can be used to further improve management decisions on the farm. This requires that within an area:

- The amount and pattern of variability in yield and soil conditions can be measured/estimated with a degree of accuracy.
- Reasons for the variability can be determined.
- The agronomic implications of the variability and its causes present a practical opportunity to vary management.

The easiest way to start this process is to investigate areas of high, average and low production within fields. The mapped data from a yield monitoring system (e.g. Fig. 20.1e) makes it easy to identify these areas, and the extent, and provides the locations to then perform:

- A complete soil profile examination and analysis.
- Crop observation for disease or pests.
- Assessment of the position in the landscape (slope, hill, hollow).

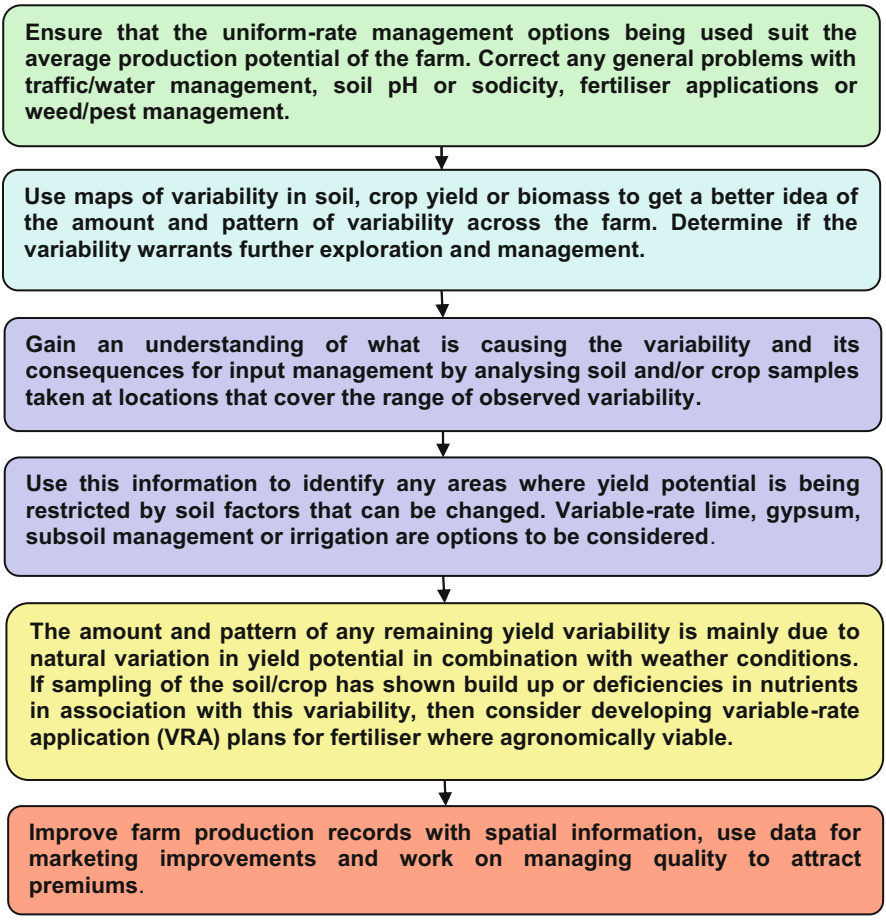


Fig. 20.2 General strategy to drive the incorporation of SSCM on a farm

The focus here isn't on providing a map of the actual soil/disease/pest variability as this would invariably involve extensive, expensive sampling and analysis. Rather, the goal is that the information gathered from the site analyses should be used with the available maps, local agronomic understanding and advice to determine whether the amount, pattern and causes of variability, along with the extent of production variability, warrant changing from a uniform application of inputs. With this understanding that local variability should be seen as the driving factor, it is possible to provide a general strategy to consider for the incorporation of SSCM on any farm (Fig. 20.2).

20.3 Developing VRA Strategies for SSCM

At present, VRA can be used to deal with spatial variability in crop production at three different scales:

- Whole-field variability
- Variability between potential management classes
- Variability within potential management classes

20.3.1 Whole-Field Variability

Input rate changes are made across a continuous range to deal with continuous variability in crop requirements. This requires that the variability in requirements can be determined accurately at a fine scale across a field. This type of operation may utilise map-based or real-time VRT. In the case of map-based applications, a prescription map for VRA of inputs may be derived from a single attribute map or the combination of a number of maps. Single maps of variability can be used directly to formulate management plans when the variability and its impact on management operations are obvious and directly linked. Options available at present include the following:

- Topography maps can be used to realign working directions, field boundaries or seedbed formation to improve water management (Webb et al. 2004).
- In broadacre irrigated fields, topography maps can be used to derive cut/fill maps for levelling purposes (SRA 2015).
- Crop yield maps can be used to calculate nutrient removal and drive a VRA replacement strategy. This is most suitable for the less variable, less mobile phosphorus (Grove and Pena-Yewtukhiw 2007).
- Soil pH maps made from currently available real-time sensing systems can be used to direct lime application (Lund et al. 2005).

As more on-the-go sensing systems are commercialised for mapping soil properties directly (e.g. soil nutrients and carbon), decision-making directly from single maps may become more commonplace. At present, VRA application at the whole-field scale using combinations of maps as a prescription is restricted to the standardisation and combination of a number of seasons of crop yield data to provide a continuous map of variation in the average nutrient removal.

Real-time VRA to treat continuous variability remains the domain of crop reflectance sensing systems that identify the presence and vigour/health of plants. Systems that calculate and prescribe nitrogen requirements, growth regulator and agrochemical applications using the data from these sensing systems have all been commercialised (Raun et al. 2005; Holland and Schepers 2013).

Ultimately, it is the technical specifications of the application equipment that will control the minimum treatment area and minimum incremental rate change in any continuous VRA operation, specifically:

- The minimum independent section width of the application equipment
- The speed at which a rate change can be relayed to the equipment
- The time it takes for the equipment to change up and down rates
- The speed of travel

20.3.2 Potential Management Classes (PMC)

In defining PMC, fields are divided into areas (classes) that have shown differences in production potential that may require different management treatments. The variability of an attribute of interest is then described by the average value in each class:

- A management class in SSCM is the total area for which a specific management treatment can be identified. Management classes are distinguished from each other based on the requirement for different management treatments.
- A management zone is an unbroken area to which a specific management class treatment is applied.
- A management class may therefore be allocated to one or more management zones within field or farm.

This approach may be regarded as a risk-averse compromise between uniform management (single-rate application) and continuous management of whole field variability. Essentially, the PMC should partition the variability within a field so that:

- Within-class variability is reduced below whole field variability.
- Average within-class variability is significantly different between management classes.
- A reduction in variability will also be expressed in important attributes that have not been used to make the management classes.

The identification of the spatial boundaries of the classes and zones may be determined using a wide range of initial information and methods. Examples of the type of initial information range from hand drawn farmer ‘mud’ maps (Fleming et al. 2000), yield data (Diker et al. 2004; Flowers et al. 2005), soil survey data (Fridgen et al. 2000; Franzen et al. 2002; Dillon et al. 2005), reflectance imagery (Stewart and McBratney 2001, Taylor et al. 2002), proximal soil sensors (Lund et al. 2002) and combinations of these (Whelan et al. 2002; Koch et al. 2004; Fleming et al. 2004; Schepers et al. 2005).

Equally, there have been many different approaches to define the shape and number of management zones using the information that range from the simplistic to

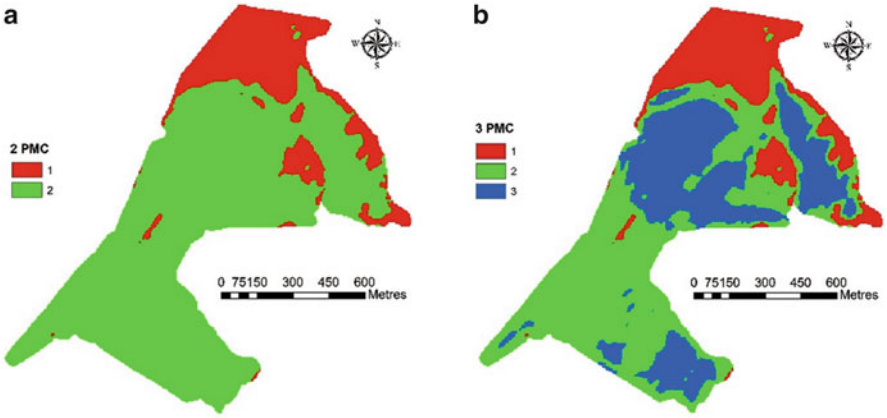


Fig. 20.3 PMC delineated using hard k-means multivariate clustering of the yield, soil and elevation layers shown in Fig. 20.1. Here (**a**) 2 PMC and (**b**) 3 PMC

statistically complex. These include hand-drawn polygons on yield maps or imagery (Fleming et al. 2000; Nehmdahl and Greve 2001), normalised yield classification (Swindell 1997), supervised or unsupervised classification of imagery (Anderson and Yang 1996; Stewart and McBratney 2001), spectral filters using fast Fourier transformation (Zhang and Taylor 2001), hard k-means cluster analysis (Taylor et al. 2002; Whelan et al. 2002), fuzzy k-means cluster analysis (Lark and Stafford 1997; Burrough and Swindell 1997) and multivariate hard k-zones (Shatar and McBratney 2001).

Determining the optimum data layers needed to delineate management zones for a particular field without any a priori information is difficult. In general emphasis should be given to sensors that record known or expected yield determining factors and sensors that measure the final crop response if available (Taylor et al. 2007).

Figure 20.3 shows the result of combining the elevation, soil and yield data layers from the field in Fig. 20.1 using a hard k-means clustering process to form potential management classes. Using previously gathered maps of variability and statistical processes takes the guesswork out of setting boundaries, but it is important to make sure any maps that are used correctly reflect the variability in the field and don't have any problems in the patterns caused by poor past crop management, weather damage or collection errors.

As can be seen from Table 20.1, when the number of management classes is increased, the differences between the mean in each class for the properties used in the classification process decrease. As these differences decrease, the potential benefits of managing the classes differently are also reduced. So any decision on the number of management classes to use should be made with consideration to both maintaining an agronomically significant difference (Taylor et al. 2007) and ensuring the pattern of variation is not overly broken up into very small unmanageable zones.

Table 20.1 The mean values in each identified PMC for the properties used in the multivariate classification process when 2 and 3 PMC are defined

Number of PMC	PMC	Soil ECa (mS/m)	Elevation (metres)	Wheat yield (t/ha)	Wheat protein (%)	Gamma TC (counts)
2	1	33.3	381.2	1.63	13.3	307.9
	2	38.8	379.7	2.49	11.7	295.6
	1	29.8	383.3	1.44	14.0	305.7
3	2	35.5	380.0	1.79	12.9	307.5
	3	39.1	379.7	2.59	11.5	294.6

There is no set number of classes that should be identified. The decision is influenced by the treatment to be applied, the range of variability in the identifiable driver for the treatment and the importance to production of matching small incremental changes in the driver with changes in the treatment. In practice, it is most commonly one to three management classes that are identified so that:

- One – uniform treatment
- Two – a division between high and low
- Three – high, medium and low treatments are required

20.3.3 Variability Within PMC

Here the base-rate treatment requirements for each management class are identified prior to the application operation, but the actual input rate is modified by a measure of variability gathered during the operation within each zone/class.

This option is a mix of dealing with variability using PMC and continuous whole-field treatment. It relies on using historical information and/or knowledge in combination with real-time VRT. Real-time systems allow the operator to set base rates, and some systems also provide the option of importing base-rate maps to be combined with the data from the real-time sensor during application.

Given the above, it is possible to partition the practical VRA options into four main categories based on the scale at which information on variability is gathered/treated and the category of VRT equipment that is used (Table 20.2). The categories are:

- Map-based management class operations
- Map-based whole field operations
- Real-time management class operations
- Real-time whole field operations

Table 20.2 The four categories of VRA operations are defined by the scale at which information on variability is gathered and the category of VRT equipment

VRT equipment	Management class treatment of variation	Whole field treatment of variation
Map based	A map of broad areas where different management is required is used to describe where rate changes will occur	A map of variability across a whole field is used to determine continuous rate changes
	Examples:	*Examples:*
	Nitrogen applied at three different rates to three management classes	Phosphorus applied based on the pattern of removal from previous yield map/maps
	Lime only applied in areas previously identified as requiring pH adjustment	Pre-emergent herbicides applied based on a map of soil organic matter content
Real-time sensors	A map of broad areas where different management is required is used to describe where base rates will change. Real-time sensors are then used to measure variability within each management class and vary application around the base rates	Variability across the whole field is sensed and used to control treatment rates during application
	Examples:	*Examples:*
	Base nitrogen rates are calculated for each of three classes. A crop reflectance sensor is then used during application to measure variability in crop condition within each class and vary application up or down from the base rate	Only green weeds, which are detected in fallow using reflectance sensors, are treated with herbicide
	Maps of areas where broad-spectrum herbicide is to be applied are combined with reflectance sensors to deliver specific action chemicals to detected weeds	Crop reflectance sensors are used to control nitrogen application rate variation around a single base rate

Deciding on the applicability and category of VRA operations for a situation can be done using a tree structure of questions that require only a positive or negative answer. An example is shown in Fig. 20.4, where the decision begins with the premise that variability in crop yield is the initial signal that VRA might be warranted. Another model might begin with soil/landscape variability or variation in crop reflectance.

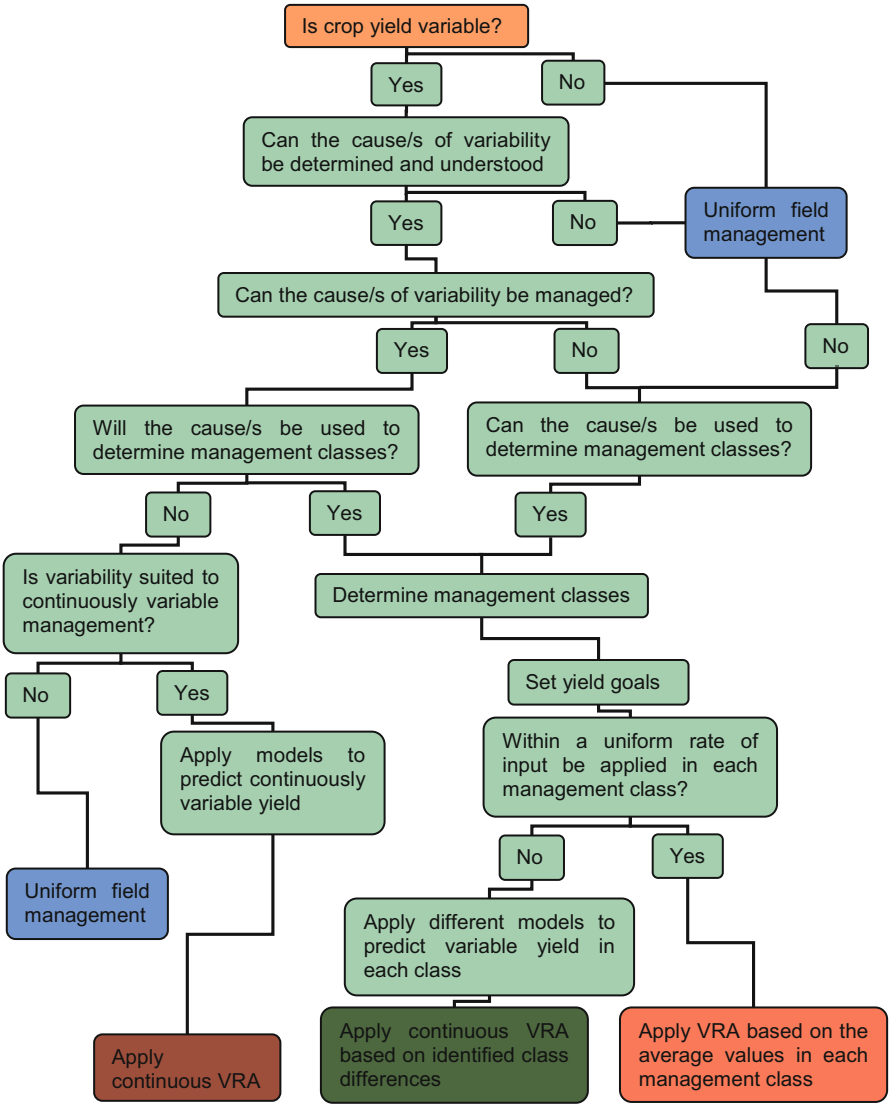


Fig. 20.4 Basic decision tree for VRA management options based on a measure of crop yield variability

20.4 Examples of Data-Driven VRA Cropping Decisions Using Maps of On-Farm Variability

20.4.1 Map-Based Whole-Field Variability

20.4.1.1 Nutrient Replacement

The amount of nutrient removed by a previous crop can be calculated using a yield map and a formula relating the amount of nutrient exported per tonne of grain. Using this process to prescribe fertiliser requirements for a following crop assumes that the optimum yield has been achieved across the field and/or soil levels of the target nutrient were at or above critical levels prior to sowing the previous crop. Here, the wheat yield map for a 110 ha field (Fig. 20.5a) has been used to calculate the amount of phosphorus removed using Eq. 20.1:

$$P \text{ removed}_{(\text{kg } P/\text{ha})} = \text{wheat yield}_{(t/\text{ha})} \times P \text{ content of wheat grain}_{(\text{kg } P/t)} \quad (20.1)$$

where the P content of wheat grain = 4 kg P/t.

The map produced (Fig. 20.5b) can be used as the basis for phosphorus replacement rates in the field. However, it does not allow any margin for error in the mapped yield nor the possibility that a base level of phosphorus may be required in the initial stages of crop growth. Equation 20.2 includes a base rate of 5 kg P/t to be applied over the whole field as part of the calculation of the total required phosphorus application rate (Fig. 20.5c):

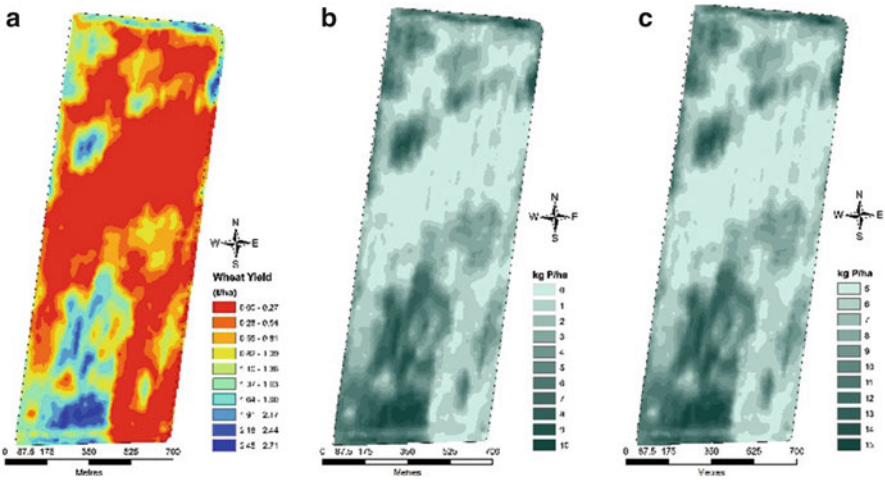


Fig. 20.5 Nutrient replacement calculation: (**a**) wheat yield map, (**b**) actual phosphorus removed, (**c**) phosphorus removed plus a base rate of 5 kg P/ha

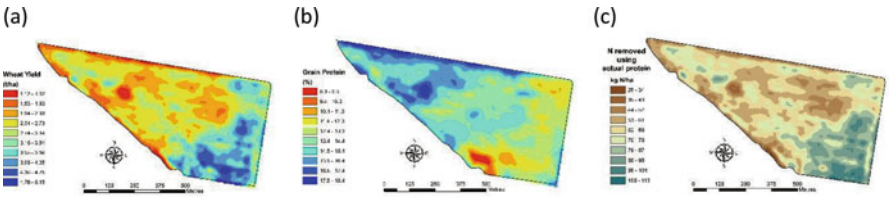


Fig. 20.6 Maps for a 42 ha field: (**a**) Wheat yield map, (**b**) grain protein map and (**c**) nitrogen removed (kg *N*/ha) calculated using the data from the two production maps

$$P \text{ removed}_{(\text{kg } P/\text{ha})} = \text{wheat yield}_{(t/\text{ha})} \times P \text{ content of wheat grain}_{(\text{kg } P/t)} + 5_{(\text{kg } P)} \tag{20.2}$$

For nitrogen in wheat, the same approach can be applied (Eq. 20.3):

$$N \text{ removed}_{(\text{kg } N/\text{ha})} = \text{wheat yield}_{(\text{kg/ha})} \times \text{Protein}_{(\%)} \times \%N \text{ in wheat protein factor} \tag{20.3}$$

where the %N in wheat = 17.5% (the factor = 0.00175).

Figure 20.6a shows the wheat yield map for a 42 ha field, and Fig. 20.6b shows a grain protein map from the same season as gathered by an on-harvester protein sensor. The result of combining this data with the yield data using Eq. 20.3 is shown in Fig. 20.6c.

The maps in Figs. 20.5b, c and 20.6c can be used to calculate continuous VRA maps for the application of phosphorus and nitrogen fertiliser products, respectively.

20.4.2 Variability Classified Into PMC

Here, a 75 ha field has been divided into three potential management classes using fine-scale soil EC_a, elevation and yield information. Soil sampling sites have been located within each of the three zones (Fig. 20.7) in an attempt to explore causes for the yield differences between the classes. Soil cores were taken at each site and the samples divided into a topsoil (0–0.3 m) and subsoil (0.3–0.9 m) layer for analysis. The results are shown in Tables 20.3 and 20.4.

Analysis of the topsoil (Table 20.3) shows that Class 1 has a higher CEC and a lower sand fraction than Classes 2 and 3, but is lower yielding (Fig. 20.7). Of note for management is that the soil nitrate is also substantially higher in Class 1. An examination of the soil below 0.3 m (Table 20.4) shows that the CEC and clay content of Class 1 are significantly lower than in the other Classes, and the soil nitrate remains double.

The interpretation of these results is that the difference in the physical properties of the subsoil, combined with the fact that the soil is on average 40% shallower

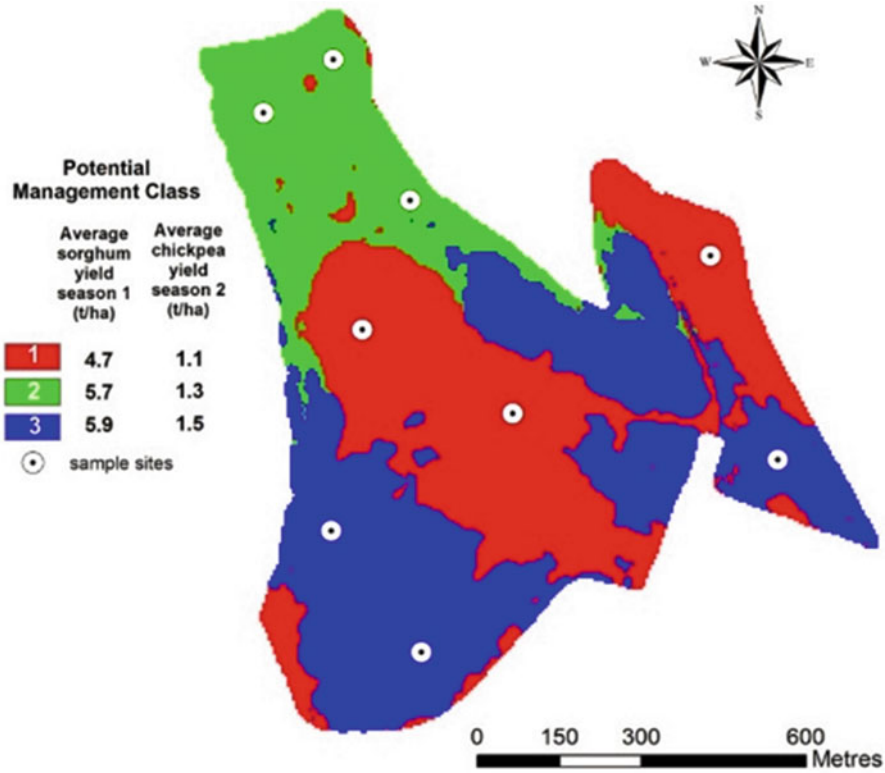


Fig. 20.7 A 75 ha field divided into three potential management classes (PMC). Average yield responses for the PMCs for sorghum and chickpea crops used in the multivariate classification process are shown in the legend

in Class 1 (Table 20.4), conspires to restrict the quantity of available moisture in the profile of Class 1 compared to Classes 2 and 3. This relative limitation in soil moisture in Class 1 would limit crop yield during most seasons and therefore reduce the nitrogen requirement. Under uniform fertiliser management, accumulation of soil nitrogen reserves in the whole profile (as evident in nitrate and total N levels in Tables 20.3 and 20.4) would then occur. While there is also some evidence for the build-up of phosphorus in the topsoil of Class 1, from a management perspective it is the manipulation of soil nitrogen that would bring the biggest benefit to the profitability of the field.

Having found the major driver/drivers for the spatial variability between the PMC in the field, there are three general options for determining input rate changes based upon either:

- Maintenance of a nutrient balance to achieve a uniform yield goal
- Modifying yield goals between classes
- Rate response experiments across classes

Table 20.3 Soil test results for the 0–0.3 m soil layer at the sample sites in Fig. 20.7

Soil attribute	Class 1 (red)	Class 2 (green)	Class 3 (blue)	Field mean
pH (CaCl)	7.6	7.2	7.8	7.5
O.C. (%C)	0.9	0.8	0.6	0.8
$N0_3^-$ (mg/kg)	30.4	19.3	10.6	20.1
P (mg/kg)	5.3	6.3	2.7	4.8
K (mmol(+)/kg)	60	90	50	70
Ca (mmol(+)/kg)	626	405	513	515
Mg (mmol(+)/kg)	132	183	221	179
Na (mmol(+)/kg)	20	50	10	60
Total N (mg/kg)	1026	1079	658	921
CEC (mmol(+)/kg)	770	600	750	700
Ca/Mg	4.8	2.2	2.3	3.0
ESP %	0.25	0.92	1.35	0.84
Sand %	10	16	12	13
Silt %	15	13	13	14
Clay %	75	71	75	74
E.C. (mS)	163	138	136	145

Table 20.4 Soil test results for the 0.3–0.9 m soil layer at the sample sites in Fig. 20.7

Soil attribute	Class 1 (red)	Class 2 (green)	Class 3 (blue)	Field mean
pH (CaCl)	7.7	7.8	8.0	7.8
O.C. (%C)	0.8	0.7	0.6	0.7
$N0_3^-$ (mg/kg)	14.7	11.9	5.6	10.7
P (mg/kg)	3.7	3.0	2.5	3.1
K (mmol(+)/kg)	4.2	6.5	4.8	5
Ca (mmol(+)/kg)	421	389	470	427
Mg (mmol(+)/kg)	95	215	249	186
Na (mmol(+)/kg)	30	21	27	17
Total N (mg/kg)	887	687	532	702
CEC (mmol(+)/kg)	523	634	748	635
Ca/Mg	5.2	1.8	1.9	3.0
ESP %	0.7	3.2	3.6	2.5
Sand %	18	15	11	15
Silt %	17	11	11	13
Clay %	65	74	78	72
E.C. (mS)	126	162	155	148
Soil depth (m)	0.68	1.17	1.24	1.03
Profile avail. H_20 at sampling (mm)	68	108	128	101

20.4.2.1 Maintenance of a Nutrient Balance to Achieve a Uniform Yield Goal

This option involves establishing a yield goal for a field or area, calculating the nutrient requirements for the goal, estimating the quantity of nutrient already available in the soil and determining the additional amount required to achieve the goal. Using this traditional agronomic technique for calculating nutrition requirements provides a simple method for including the information obtained from investigative sampling within PMC into input rate decisions. Equation 20.4 can be used to calculate nitrogen requirements for wheat crops:

$$N\ \text{required}_{(\text{kg } N/\text{ha})} = \text{yield goal}_{(\text{kg/ha})} \times \text{protein goal}_{(\%)} \times \ \%N \text{ in wheat protein factor} \times \text{efficiency factor} \tag{20.4}$$

where the %N in wheat = 17.5% (the factor = 0.00175) and the efficiency of uptake factor = 2.

According to Eq. 20.4, to achieve a yield goal of 4.5 t/ha at 13% protein would require 205 kg N/ha to be available to a wheat crop during the growing season. Applying the 4.5 t/ha yield goal uniformly across the field in Fig. 20.7, and using the soil nitrogen information in Tables 20.3 and 20.4 as sampling information prior to sowing a wheat crop, the nitrogen nutrient balance requirements can be calculated.

First, the nitrogen values from the 0–30 to 30–60 cm soil samples are used to calculate the average amount of nitrogen in kg/ha that is resident in the soil profile. Table 20.5 shows the results of these calculations for each PMC and the overall average for the field. The overall average is the estimate that would be traditionally used to calculate the uniform-rate fertiliser requirement for the field. In this instance the field average of 165 kg N/ha resident in the soil would suggest the addition of a further 40 kg N/ha is required across the field to complete the yield goal nitrogen balance (Table 20.5). In this conservative approach, no additional contribution of nitrogen from mineralisation in the soil over the growing season is included.

However, because the soil sampling has been done in classes, the class average resident nitrogen can be used to calculate the additional nitrogen required to complete the yield goal nitrogen balance in each Class. The results (Table 20.5) suggest that very different amounts of fertiliser would be required within each PMC. Applying the uniform treatment rate would over-fertilise Class 1 and 2 and under fertilise Class 3. A total of 2.7 t of urea would be wasted by the uniform application, and the substantial under-fertilisation in Class 3 would result in yield and protein losses in an average or better season. Varying the fertiliser rates to achieve the 205 kg N/ha for the 4.5 t/h yield goal in each class (Table 20.5) would be a more profitable and less environmentally risky management decision.

Table 20.5 Available nitrogen in the soil and the calculated nitrogen requirements for three PMC in Fig. 20.7

	Class 1 (29 ha)	Class 2 (13 ha)	Class 3 (33 ha)	Field average (75 ha)
Average available N in the soil (kg N/ha)	238	172	86	165
Uniform N application (kg N/ha)	40	40	40	40
Actual N required to obtain 205 kg N/ha in soil (kg N/ha)	0	33	118	
Consequence of uniform N application	2.5 t urea waste	197 kg urea waste	66% under fertilised	

Comparison between uniform application based on the average available soil nitrogen content and variable-rate treatment based on actual available soil nitrogen in each PMC

20.4.2.2 Modifying Yield Goals Between Classes

Just sampling and then modifying fertiliser rates to reach an even 'mass balance' across a field does not acknowledge that in many cases the yield potential is different between the management classes. Following directed sampling to identify the local factors impacting on yield variability, new yield goals can be assigned to the different potential management classes. This can be achieved by combining the information on the differences in the soil/landscape properties between classes with local agronomic advice on yield potential. Alternatively, the relative yield potentials can be assessed using the yield differences shown in the previous crop yield maps. Either way, local experience should be included to set a maximum yield goal for the highest performing class, and the yield goals for the other class/classes should decrease on a percentage basis.

Table 20.6 contains average sorghum yield data figures for two seasons for the whole field and each PMC in Fig. 20.7. The yield of Class 1 is between 20% and 29% (average 25%) less than Class 3, and the yield of Class 3 is between 3% and 8% (average 7%) less than Class 3. Once management is convinced that there are no amelioration questions remaining and that the yield differences are repeating over seasons, they could give consideration to reducing the yield goal and the input rates in Classes 1 and 2. There are a number of options available:

- Set a yield goal for highest yielding Class based on local advice and reduce the yield goal of the other Classes (1 and 2 here) by either the minimum/average/maximum percentage difference seen over the seasons. The most risk-averse option is to reduce the yield goal by the minimum percentage. A further option with three PMC is to set the yield goal for the middle class based on local calculations and, respectively, vary the goals up and own for the other classes.

Table 20.6 Sorghum yield differences for two seasons between potential management classes in Fig. 20.7

	Average sorghum yield (t/ha)		
Class	Season 1	Season 2	Combined
C1 (29 ha)	4.7	4.1	4.4
C2 (13 ha)	5.7	5.3	5.5
C3 (33 ha)	5.9	5.8	5.9
Whole field	5.4	5.1	5.3
% difference of lower yielding classes from class 3	C1 = 20	C1 = 29	C1 = 25
	C2 = 3	C2 = 8	C2 = 7

- Set yield goals based on the actual maximum yields obtained in each class.
- Set yield goals based on the actual averages obtained in each class.

20.4.2.3 Rate Response Experiments Across Classes

Setting up some basic experiments across a field can provide a more comprehensive understanding of the actual response to different application rates. Where PMC have been identified, field-scale experiments can be established to estimate the response in each PMC to a single input. It is best to keep these types of experiments to testing changes in a single input or practice. It is preferable to increase the number of different rates in the experiment, rather than include additional variables at this scale.

The design of the experiments should consider application equipment capability and width, incorporation of the management class pattern and a desire to minimise the area/financial impact of the experiment. These types of experiments can be run using whole field treatment strips (Lawes and Bramley 2012), but where variable-rate technology is available, the application of small strip experiments is possible (Whelan et al. 2012). An example of the physical layout for these two designs in a field is shown in Fig. 20.8. The use of small strips allows correct independent replication, more flexibility in location and greater coverage of the conditions in the field and often results in less area being subjected to potential yield-reducing treatments.

The majority of the field should be treated with the control application (i.e. traditional best practice), while the alternative treatments could include multiples of the traditional application rate (e.g. 0.25, 0.5, 1.5, 2). More than one alternative treatment rate is preferable; otherwise, the best that can be hoped for is a decision that one treatment is preferable to another. At least two alternative treatments can help provide information to decide the 'best' treatment rate.

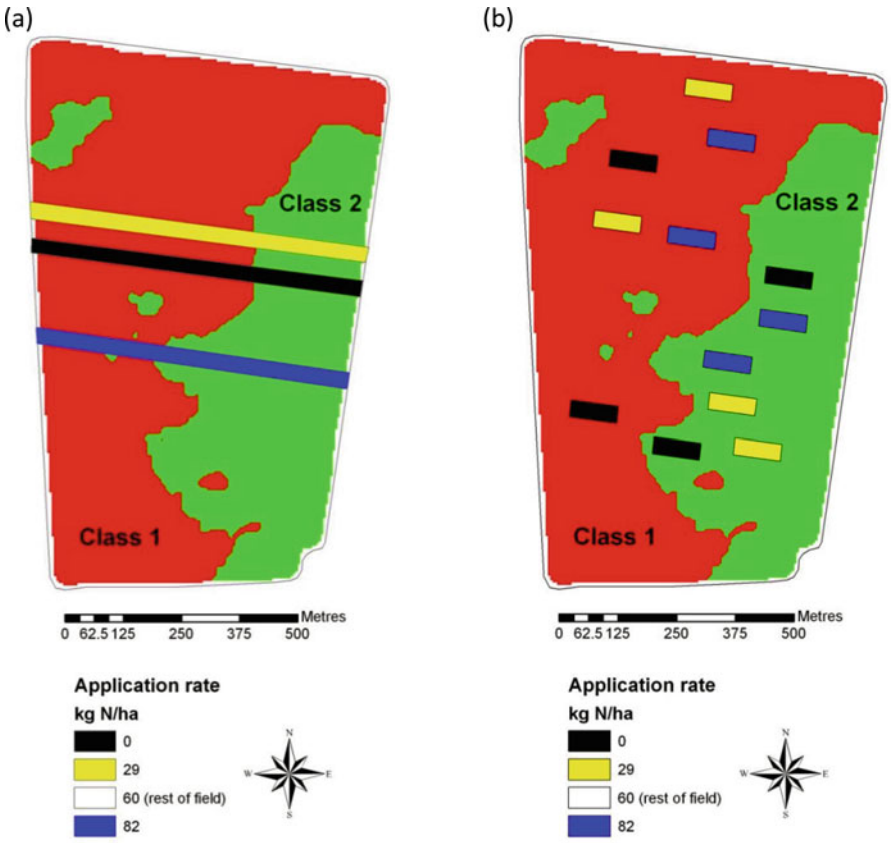


Fig. 20.8 A 79 ha field divided into two potential management classes (PMC) showing two general options for rate response experiments. (**a**) Whole-field strips and (**b**) small strips

The information from the experiment will be greatest when there is a good size difference between the treatment rates. Where possible a zero rate treatment should be included to help show the full scale of response. It is especially important where the experiment is testing whether a reduction in input rate is viable. Obviously placing zero rates in the field can cause some trepidation, but keeping the size of the treatments small can help ease such concern. With these designs, data from the whole field can be used to produce a response function for each class, and this can be interpreted using seasonally specific marginal rate analysis to determine the optimal treatment rate for each class (Fig. 20.9). Further detail on the practical application of SSCM can be found in Whelan and Taylor (2013).

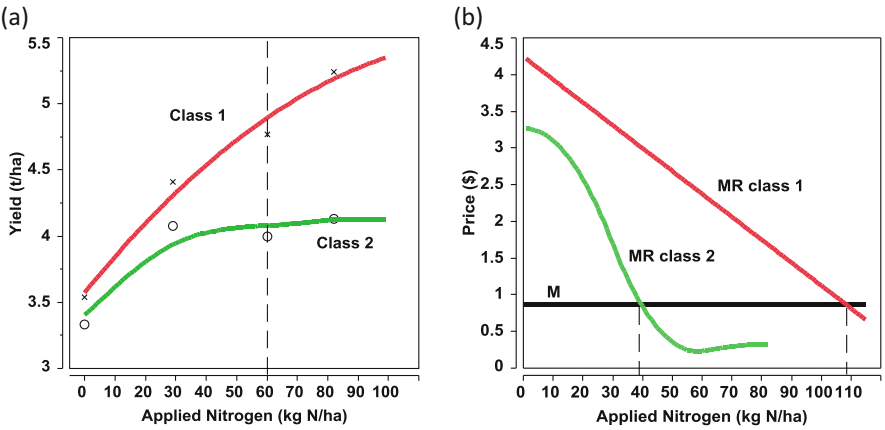


Fig. 20.9 (**a**) Actual yield response to applied nitrogen fertiliser from the application of the trial design in Fig. 20.8b where the uniform field application was 60 kg N/ha and (**b**) a marginal rate analysis (optimised where MC = MR) shows that Class 1 application optimum = 109 kg N/ha and Class 2 application optimum = 39 kg N/ha.

20.5 Data-Driven Cropping Decisions Using On-Farm and Off-Farm Data

The development and application of SSCM in cropping enterprises has been in parallel with an increase in the volume and sources of data. Long before the term 'big data' was dragged from the literature on digital data storage (Bryson et al. 1999), through the filter of business management analysts, to the present day, agriculture has been working on ideas for using 'big data' through the PA philosophy.

The data streams that have been discussed in this chapter come from diverse sources (mainly on-farm from a specific farm or field), with variable volumes, structures and scales. They have been integrated and analysed to varying degrees by PA practitioners using a range of decision support processes (see Fountas et al. 2015 for a review of Farm Management Information Systems) with the goal of improving decisions for SSCM. The current decision/action process at the farm level is generalised in Fig. 20.10.

The future should see this on-farm data, along with an increase in the sources, volumes, scales and structures of off-farm data (from other local/regional farms and non-farm domains) employed in improving SSCM. A schematic for this improved data-driven decision process is shown in Fig. 20.11, where on-farm data from a local region is aggregated and combined with off-farm data from the public domain (e.g. weather, remote sensing, predicted soil properties) for input into analysis and decision-making back on-farm. The process of data aggregation and storage may move to a cloud (or localised server)-based operation (Fig. 20.12) with research into

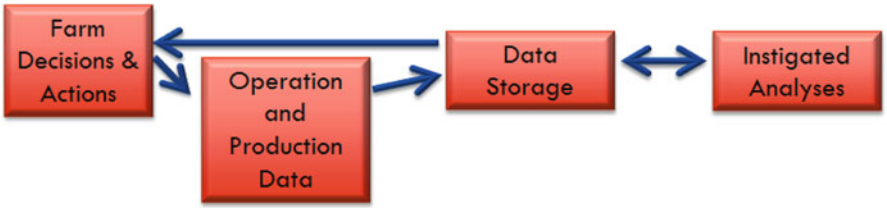


Fig. 20.10 A generalisation of the current data-driven decision/action process involving on-farm data

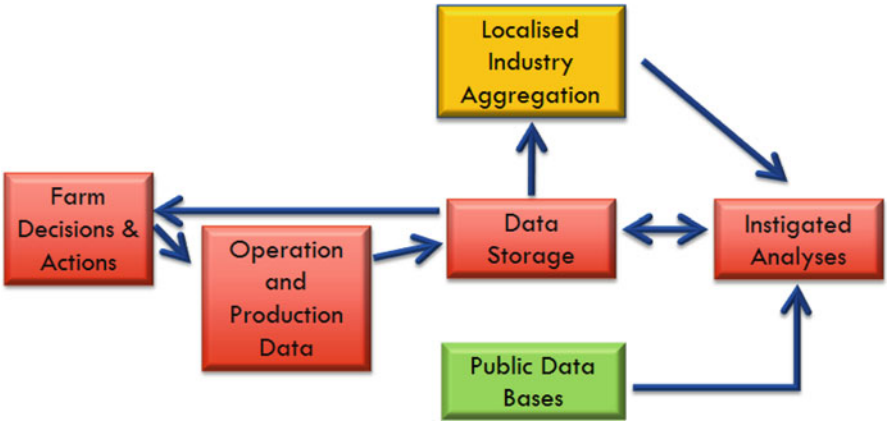


Fig. 20.11 A schematic of the data-driven decision/action process where on-farm data from a local region is aggregated and combined with off-farm data from the public domain

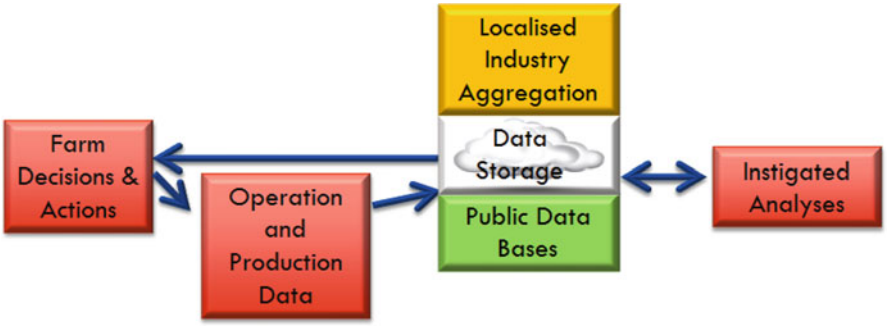


Fig. 20.12 A schematic of the data-driven decision/action process where data aggregation, storage and analysis move to a cloud (or localised server)-based operation

system design for agricultural operations underway (e.g. Kaloxylos et al. 2014). It is also feasible that the data and analyses in this process may be incorporated in a commercial sense (e.g. as a cooperative or other business entity) and made available for off-farm use (Fig. 20.13). The potential benefits from such a development

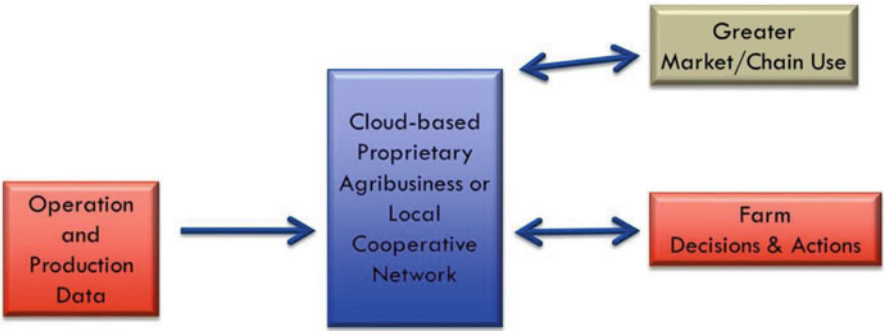


Fig. 20.13 A schematic of the data-driven decision/action process where the data storage and analysis incorporated in a commercial sense (e.g. as a cooperative or other business entity) and also made available for off-farm use

range from direct revenue from data exchange to improvements in supply chain response which would extend advantage to the whole-farm businesses involved and potentially the supply/value chain.

In practice, it will be critically important to identify and draw data from relevant sources into analytical models containing environmental, crop and farm business components that will then feed results into, or drive, key management and operational decisions.

The practical components in such a process, with respect to SSCM outcomes, may eventually include:

- **Data generation and capture (historic and real-time).** These may include information on production yield and quality, aerial/proximal in-season sensing (crop, disease, pest, soil, environment), economics, markets, distribution and consumption.
- **Data dormitories.** Cloud-based (or local subsidiary) stores of farm and off-farm data at multiple scales (production, environment, financial, markets).
- **Data-driven decisions.** Alternative options for crop management, variable-rate application and farm logistics based on the assessment of probabilistic outcomes from data-driven models of causal relationships enabled by the new combination of machine learning and mechanistic modelling approaches. These decisions may eventually be whole-business and/or whole value chain considerate.

The analytical tools would ideally contain the capability of autonomously adapting decision functions and providing farmers/agronomists with alternative management responses as the input data changes across space and/or time. However, this novel integration and analysis of relevant data from diverse domains, sources and scales to improve decision management at a fine scale should be aimed at augmenting the agronomist – not automating the agronomist – unless the decision/action process is decisive enough to warrant such an approach.

So the development of SSCM should continue with an increasing sensitivity of decisions at the within-field scale to the whole farm business outcomes, including impacts on off-farm linkages. New techniques from the field of pedometrics that enable the fine-scale description of the 3D variation in soil over space and time (especially soil moisture and nutrients) will bring the next shift in the financial and environmental optimisation of crop management.

References

Anderson GL, Yang C (1996) Multispectral videography and geographic information system for site-specific farm management. In: Robert PC, Rust RH, Larson WE (eds) Precision agriculture: proceedings of the 3rd international conference on precision agriculture. ASA, Madison, pp 681–692

Bryson S, Kenwright D, Cox M, Ellsworth D, Haimes R (1999) Visually exploring gigabyte data sets in real time. Commun ACM 42:82–90

Burrough PA, Swindell J (1997) Optimal mapping of site-specific multivariate soil properties. In: Lake JV, Bock G, Goode J (eds) Precision agriculture: spatial and temporal variability of environmental quality. Wiley, Chichester, pp 208–220

Diker K, Heerman DF, Brodahl MK (2004) Frequency analysis of yield for delineating yield response zones. Precis Agric 5:435–444

Dillon CR, Saghaian S, Salim J, Kanakasabai M (2005) Optimal water storage location and management zone delineation under variable subsurface drip irrigation. In: Stafford JV (ed), Precision agriculture '05. Proceedings of 5th ECPA, Uppsala, Sweden, June 8–11. Wageningen Academic Publishers

Fleming KL, Westfall DG, Wiens DW, Brodahl MC (2000) Evaluating farmer defined management zone maps for variable-rate fertiliser application. Precis Agric 2:201–215

Fleming KL, Heerman DF, Westfall DG (2004) Evaluating soil color with farmer input and apparent soil electrical conductivity for management zone delineation. Agron J 96:1581–1587

Flowers M, Weisz R, White JG (2005) Yield-based management zones and gridsampling strategies: describing soil tests and nutrient availability. Agron J 97:968–982

Fountas S, Carli G, Sørensen CG, Tsiropoulos Z, Cavalaris C, Vatsanidou A, Liakos B, Canavari M, Wiebensohn J, Tisserye B (2015) Farm management information systems: current situation and future perspectives. Comput Electron Agric 115:40–50

Franzen DW, Hopkins DH, Sweeney MD, Ulmer MK, Halvorson AD (2002) Evaluation of soil survey scale for zone development of site-specific nitrogen management. Agron J 94:381–389

Fridgen JJ, Kitchen NR, Sudduth KA (2000) Variability of soil and landscape attributes within sub-field management zones. In: Robert PC, Rust RH, Larson WE (eds) Precision agriculture: proceedings of the 5th international conference on precision agriculture. ASA, Madison. CD publication

Grove JH, Pena-Yewtukhiw EM (2007) Evaluating P and K fertilizer prescriptions from site-specific technologies. Soil Science News and Views. Paper 184. http://uknowledge.uky.edu/pss_views/184. Accessed 17/03/17

Holland K, Schepers J (2013) Use of a virtual reference concept to interpret active crop canopy sensor data. Precis Agric 14:71–85

Kaloxylos A, Groumas A, Sarris V, Katsikas L, Magdalinos P, Antoniou E, Politopoulou Z, Wolfert S, Brewster C, Eigenmann R (2014) A cloud-based farm management system: architecture and implementation. Comput Electron Agric 100:168–179

Koch B, Khosla R, Frasier WM, Westfall DG, Inman D (2004) Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. Agron J 96:1572–1580

Lark RM, Stafford JV (1997) Classification as a first step in the interpretation of temporal and spatial variation of crop yield. Ann Appl Biol 130:111–121

Lawes RA, Bramley RGV (2012) A simple method for the analysis of on-farm strip trials. Agron J 104:371–377

Lund ED, Wolcott MC, Hanson GP (2002) Applying nitrogen site-specifically using soil electrical conductivity and precision agriculture technology. In: Proceedings of the 6th international conference on precision agriculture, Jul 14–17, 2002, Minneapolis, Minnesota, USA

Lund ED, Adamchuck VI, Collings KL Drummond PE, Christy CD (2005) Development of soil pH and lime requirement maps using on-the-go soil sensors. In: Stafford JV (ed), Precision agriculture '05. Proceedings of 5th ECPA, Uppsala, Sweden, June 8–11. Wageningen Academic Publishers, pp 457–464

McBratney AB, Whelan BM, Ancev T, Bouma J (2005) Future directions of precision agriculture. Precis Agric 6:7–23

Nehmdahl H, Greve MH (2001) Using soil electrical conductivity measurements for delineating management zones on highly variable soils in Denmark. In: Grenier G, Blackmore S (eds) ECPA 2001, proceedings of the 3rd European conference on precision agriculture. agro Montpellier, Montpellier, pp 461–466

Raun WR, Solie JB, Stone ML, Martin KL, Freeman KW, Mullen RW, Zhang H, Schepers JS, Johnson GV (2005) Optical sensor-based algorithm for crop nitrogen fertilization. Commun Soil Sci Plant Anal 36:2759–2781

Robert PC (1993) Characterisation of soil conditions at the field level for soil specific management. Geoderma 60:57–72

Schepers AR, Shanahan JF, Liebig MA, Schepers JS, Johnson SH, Luchiaro A (2005) Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. Agron J 96:195–203

Shatar TM, McBratney AB (2001) Subdividing a field into contiguous management zones using a k-zones algorithm. In: Grenier G, Blackmore S (eds) ECPA 2001, proceedings of the 3rd European conference on precision agriculture. agro Montpellier, Montpellier, pp 115–120

Stewart CM, McBratney AB (2001) Using bare soil imagery to determine management zones for the variable-rate application of inputs for cotton. In: Grenier G, Blackmore S (eds) ECPA 2001, proceedings of the 3rd European conference on precision agriculture. agro Montpellier, Montpellier, pp 319–324

Sugar Research Australia. (2015) Precision agriculture for the sugarcane industry. Sugar Research Australia, Indooroopilly. ISBN 978-0-949678-34-8

Swindell JEG (1997) Mapping the spatial variability in the yield potential of arable land through GIS analysis of sequential yield maps. In: Stafford JV (ed) Precision Agriculture 1997. Bios, Oxford, pp 827–834

Taylor JA McBratney AB, Ciaverella R (2002) Predicting and mapping winegrape quality from multiple must properties. In: Proceedings of the 6th international conference on precision agriculture, Jul 14–17, 2002, Minneapolis, Minnesota, USA

Taylor JA, McBratney AB, Whelan BM (2007) Establishing management classes for broadacre grain production. Agron J 99:1366–1376

Van Grinsven HJM, Tiktak A, Rougoor CW (2016) Evaluation of the Dutch implementation of the nitrates directive, the water framework directive and the national emission ceilings directive. Wagening J Life Sci 78:69–84

Webb B, Blackwell P, Riethmuller G, Lemon G (2004) Tramline farming systems technical manual. Department of Agriculture and Food, Perth. Bulletin 4607

Whelan BM, McBratney AB (2000) The "null hypothesis" of precision agriculture management. Precis Agric 2:265–279

Whelan BM, Taylor JA (2013) Precision agriculture for grain production systems. CSIRO Publishing, Clayton. 199p

Whelan BM, Cupitt J, McBratney AB (2002) Practical definition and interpretation of potential management zones in Australian dryland cropping. In: Robert PC, Rust RH, Larson WE (eds)

Precision agriculture, proceedings of the 6th international conference on precision agriculture. ASA/CSSA/SSSA, Madison, Wisconsin. 15p

Whelan BM, Taylor JA, McBratney AB (2012) A 'small strip' approach to empirically determining management class yield response functions and calculating potential financial 'net wastage' associated with whole-field uniform-rate fertiliser application. Field Crop Res 139:47–56

Zhang N, Taylor RK (2001) Applications of a field-level geographic information system (FIS) in precision agriculture. Appl Eng In Agric 17:855–892

Chapter 21
Variograms of Soil Properties for Agricultural and Environmental Applications

Stacey Paterson, Alex. B. McBratney, Budiman Minasny, and Matthew J. Pringle

"Think left and think right and think low and think high. Oh, the things you can think up if only you try"!

DR Suess

21.1 Geostatistics and Precision Agriculture

The previous chapter describes how precision agriculture can be used to improve farm management to achieve economic and environmental benefits. Short-range differences in soil attributes mean that spatially differentiated management can create economic or environmental benefits. Effective precision agriculture requires accurate soil mapping at subfield scales so that management practices can be modified. Improvements in farming technology, for instance, GPS-controlled farm equipment, decrease the difficulty and cost associated with spatially differentiated management. This improves the ease of implementation and makes high-resolution soil maps more valuable.

A key challenge for geostatistics in precision agriculture is the detection of soil variability at important subfield scales and the systematic incorporation of this variability into accurate field scale soil maps.

S. Paterson (✉) • A.B. McBratney • B. Minasny
Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia
e-mail: stacey.paterson@sydney.edu.au; alex.mcbratney@sydney.edu.au; budiman.minasny@sydney.edu.au

M.J. Pringle
Department of Science, Information Technology and Innovation, EcoSciences Precinct, GPO Box 5078, Brisbane, QLD 4001, Australia
e-mail: matthew.pringle@dsiti.qld.gov.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_21

21.2 Soil Survey, the Variogram and Kriging

Soil attributes are typically difficult and expensive to observe. As a result, soil attribute maps made for the purpose of precision agriculture are typically created from point observations which represent a small proportion of the area to be mapped. Estimates of soil attributes in unknown areas are based on the observations and an expectation of the regions between them. This process of predicting attributes in unobserved areas is known as kriging (explained in more detail in Chap. 10). Assumptions about spatial variability are typically derived from the variogram, which links spatial separation distance to expectations about variability. Chapters 10 and 11 explain the variogram, its uses and the different methods of calculating the variogram in more detail.

The variogram is sometimes called the 'cornerstone of geostatistics'. Accurate estimation of the variogram is critical to the production of accurate soil maps. Because of the hidden nature of most soil attributes, we can usually only directly observe a small proportion of an area of interest. In Chap. 10, the distinction between the experimental or empirical variogram and the model variogram is described in some detail. The empirical variogram plots the average variance against separation distance for a number of distinct lags. The model variogram uses the information from the empirical variogram to estimate the expected variability at all lags. The purpose of the model variogram is to estimate the true underlying spatial variation at a level of detail that allows useful predictions.

Interpolation of results into unobserved points depends on the spatial structure estimated by the variogram. The closer the estimated variogram to the underlying spatial structure, the more accurate the subsequent interpolation. In general, a variogram computed from samples with finer spacing and more observations will estimate the underlying spatial structure more accurately than a variogram computed from samples with coarser spacing. Finer spacing allows the detection of spatial structure across more scales. The extent to which this is true will depend on the interaction of the spacing with the underlying spatial structure. For example, if there is no spatial relationship between points more than 5 m apart, then decreasing spacing from 50 m of separation distance to 10 m will not improve the variogram. We pause here to explain some key components of the variogram and how they are affected by survey design.

21.3 Key Components of the Variogram

While the variogram can take many forms (see Chap. 10, Sect. 10.1.1, for detail on some commonly used models), there are three components which are typically considered the most important indicators of spatial structure. Shown in the stylised diagram in Fig. 21.1 and subsequently described, these components are the nugget, the sill and the range. The estimation of each of these parameters depends strongly on the sampling design.

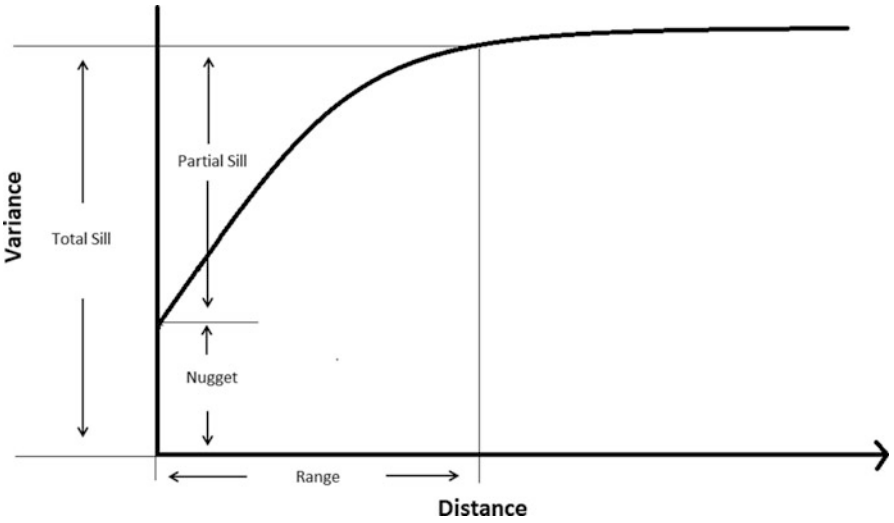


Fig. 21.1 Stylised diagram, showing the three most important indicators of spatial structure, the nugget, the sill and the range

21.3.1 Nugget

In principle the nugget effect captures nonspatial variation: measurement error; random variation. However, in practice the nugget will also capture spatial variation that occurs at scales less than the smallest sampling interval. If the sampling interval is wider than an important scale of spatial variation, then this will increase the nugget. Aliquotting (pooling of samples) will decrease the nugget. A larger support (area over which the sample is taken) will also decrease the nugget.

21.3.2 Total Sill

The total sill is defined as the maximum variability that can be expected for a particular soil property. Beyond a certain separation distance (the range, see below), the expected variability will not increase past the value of the total sill (or in some cases will increase only very slowly and slightly past the sill). Only bounded variogram models have a sill. The total sill is more likely to be affected by the number of samples and the wider separation distance than the minimum sampling spacing. When modelling spatial variability at a field scale, it is unlikely that the maximum variability of the soil will be reached. However, unbounded models are rarely fit. In the context of precision agriculture, we can think of the total sill as being the maximum variability within this particular context or a local maximum. If we extended the sampling to a regional level, it is likely we would reach another magnitude of variability.

21.3.3 Partial Sill

The distance between the nugget and the sill is known as the partial sill. The partial sill is the component of variation that can be spatially attributed. In Fig. 21.1, the semivariance increases linearly with distance at short separation distances. As the separation distance increases, the rate of increase in the semivariance decreases, before eventually reaching a plateau as the semivariance reaches the total sill. This pattern is commonly found in variogram models (see Chap. 10 for more detail on the functional forms). Accurately determining the change in variability between the nugget and the sill requires sufficiently dense sampling at appropriate scales of variability.

21.3.4 Range

Like the lag and the sill, the range estimated by the theoretical variogram will be strongly affected by the sampling design. As the lag increases, the confidence intervals widen (Oliver and Webster 2014) which can make it difficult to accurately fit a variogram at lags approaching the full extent. Precision agriculture surveys are typically conducted over areas from a few hectares to a few hundred hectares. This precludes them from capturing landscape or continental scale variability. Despite this, the majority of variogram fits from precision agriculture studies are bounded. The modelled range and sill can be thought of as the 'local sill' associated with the 'local range' associated with the particular extent and spacing of that survey. It is very likely that if the extent of the survey was increased, another degree of variability would be found. It is critical to consider the potential effect of both the extent and the spacing on the range when considering precision agriculture studies and how their findings may inform your own work.

21.4 Soil Survey Design: Capturing Spatial Variability

The task of ensuring that a soil survey captures the necessary scales of variability is not a trivial one. The expense of a soil survey, and the commonly destructive nature of soil sampling, creates a pressure to reduce the number of sampling points as much as possible. However, if sampling is insufficient or too sparse (relative to the underlying spatial structure), it will compromise the variogram and thus the accuracy of the maps which are calculated from it. It is noted in Chap. 11 that one of the chief difficulties associated with the design of a soil survey for estimating a

variogram is that the underlying spatial structure is unknown at the time the survey is designed. If the underlying spatial structure of a soil attribute is known, we can design a soil survey sufficiently to create a map at a particular level of detail.

Where budget allows it, the best practice is to undertake a preliminary soil survey to estimate important scales of variation before a more comprehensive soil survey is designed. This survey should be nested in its design to increase the chances of capturing important scales of variability (Pettitt and McBratney 1993; Webster and Oliver 2001). However, a preliminary survey will rarely be economically feasible. If it is not possible to conduct a preliminary survey, a soil survey design is likely to benefit from the consultation of alternative sources of information about soil variability, such as existing literature or covariates.

Budget or practical pressures may be sufficient to impede the collection of sufficient data to reliably calculate an empirical variogram. A stable variogram calculated from classical geostatistical methods (as described in Chap. 10) requires around 100 observations. More modern methods (as described in Chap. 11) typically require around 50. It is not possible to estimate spatial variability at distances less than the minimum separation distance. In cases where there are few or widely spread soil observations, alternative sources of information may be useful for the calculation of a variogram for kriging.

The expense of gathering soil observations creates a need for cheaper sources of information about soil variability, either to assist in the planning of a soil survey or to use in the process of kriging itself. Many authors have identified sources and strategies for the production of this information.

21.5 Variograms from the Precision Agriculture Literature

Variograms calculated for the same soil attribute may be a useful source of information. However, variograms calculated for the same property can vary significantly for a number of reasons that should be carefully considered. As outlined in Chap. 12, parent material, soil type, land use and climate will all have an effect on soil spatial variability. Consideration of these factors will be important in the selection of a variogram. Unfortunately, knowledge about soil spatial variability does not extend to the quantification of which of these factors are most important for determining spatial variability of different soil attributes.

The underlying variability of soil attributes might be different for the reasons mentioned above. In addition, the methods used to detect soil variability might create differences in the shape of the variogram. Different projects may focus on detecting variability at different scales for management or budget reasons. Even if the soil type is similar between two studies, the spatial variability might not be measured in a way that provides useful information.

It is also very important that the statistical methods used are assessed critically, before results are used or duplicated. There are a number of variograms included in Tables 21.1, 21.2, 21.3, 21.4, 21.5, 21.6, 21.7 and 21.8 which appear to use insufficient sample numbers for variogram estimation. There are several variograms which appear to assign spatial structure at a magnitude that appears to be meaningless relative to the units (i.e. total sill of less than 1% for the soil texture fraction).[1] These results have been included for completeness, but we wish to draw the readers' attention to the fact that some of the variograms included in this collection may have issues associated with them.

We describe below some key trends we have noted in the compilation of soil variograms and include a graphical summary of the variograms we have compiled. Variograms from McBratney and Pringle (1999) are included as well as those we have collected from the intervening period. Summary details and references for each variogram are given in Tables 21.1, 21.2, 21.3, 21.4, 21.5, 21.6, 21.7 and 21.8. We encourage the reader to consult each source directly for more detail about the sampling design and process.

Variograms for the same soil attribute from existing literature can be a useful source of information about field scale soil. It is important to exercise caution when consulting this literature, as variograms have been created from different soil types, for different purposes and possibly with important methodological limitations.

21.5.1 Field Scale Soil Variograms: Key Trends

21.5.1.1 Variogram Forms

Across all soil properties, a number of functional forms were fit to variograms. For each soil property, at least one study found no spatial structure (i.e. pure nugget) to be the best fit. This suggests that either the spatial structure occurs at scales finer than those surveyed or that the variability in the property in question is less than the experimental error. Spherical and exponential variograms were commonly used. Some papers described the functional form as 'experimental'. These models were fit with an exponential model.

[1] In some cases the magnitude of the nugget and partial sill is extremely small compared to the magnitude of the standard deviation. This may suggest that the data has been transformed in some way before the variogram has been fit. We have reported the results as in the original article. These results should be interpreted with particular caution.

Table 21.1 Compilation of key properties for clay variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Lopez-Granados et al. (2002) (site two)	Caracol, Southern Spain		Vertic Xerochep	Crop	Before sowing	Bouyoucos densimeter	6	84	26.3	12	0–0.10	Exponential	61	84	29.9
b	Adderley et al. Adderley et al. (1997)	Nigeria			Trees			30	480	25.8	9.1	0–0.10	Experimental	30	110	250
c	Lopez-Granados et al. (2002) (site one)	Monclova, Southern Spain		Alfisol	Crop	Before sowing	Bouyoucos densimeter	11.2	80	19.6	10.8	0–0.10	Nugget	113	0	0
d	Oliver and Webster (1987), (10–15)	Wyre Forest, England		Acid brown loam	Forest			0.025	100			0–0.15	Exponential	24.9	50.7	51.6
e	Oliver and Webster (1987), (0–5)	Wyre Forest, England		Acid brown loam	Forest			0.025	100	24.4	10.8	0–5	Exponential	14.34	51.8	65.7
f	Kerry and Oliver (2007)							15	118	47.5	6.2		Spherical	0.72	36.72	71.48
g	Nolin et al. (1996)	Quebec		Aquepts	Crop			10	130	44.3	6	0–15	Spherical	10.8	20.9	76
h	Shatar (1996)	Moree, Australia		Vertisol	Crop			15	114	49	5.5	15–30	Spherical	17	13.5	310
i	Kerry and Oliver (2007)							44	294	17.9	7.3		Spherical	4.19	22.9	152
j	Farooque et al. (2012) (site two)	Nova Scotia, Canada		Sandy loam, orthic humo-ferric podzols	Orchard		Bouyoucos densimeter	1.6	86	9.6	4	0–15	Spherical	8.08	17.91	65.5
k	Miller et al. (1988)	Sacramento, California		Xererts, Xerochrepts	Crop			8	99	37.1	4.6	Surface	Spherical	7	14	75
l	Shouse et al. (1990)	Texas, USA		Vertisol				0.3	203	40	3.9	0–30	Exponential	11.7	4.8	8.47

(continued)

Table 21.1 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
m	Kerry and Oliver (2007)	United Kingdom						10.5	205	30.2	5.8	0–15	Spherical	3.12	11.99	99.86
n	Liu et al. (2008)	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Clay loam to loam	Tobacco plantation	Before sowing	Bouyoucos densimeter	87	81	41.7	3.4		Spherical	2.72	11.78	609
o	Williams (1987) (four directions average)	Oklahoma		Paleustolls	Crop			1.62	108	22	2.3	0–30	Experimental	8.38	4.88	52.5
p	Shouse et al. (1990), (field 1)	Texas, USA		Vertisol				0.3	182	51	3.1	0–30	Exponential	3.8	6.4	11.89
q	Shukla et al. (2004)	Gross-Enzersdorf, Austria	510 mm, 10C	Loam to sandy loam	Crop	May	Pipette method	6.25	60	46.8	4.6	0–15	Spherical	5.24	4.27	347
r	Shouse et al. (1990)	Texas, USA		Vertisol				0.3	205	46.1	2.6	0–30	Spherical	3.5	3.1	36.58
s	Panagopoulos and Antunes (2008)	Algarve region, south Portugal						400	81	32.7	7	0–20	Gaussian	4.83	5.77	1,997.8
t	Kristensen et al. (1995) Riso	Riso, Denmark		Sandy loam/sandy clay loam	Crop			10.9	270			0–25	Exponential	0	6	37
u	Kristensen et al. (1995) (Vindum)	Vindum, Denmark		Sandy loam	Crop			10	302			0–25	Exponential	0	4.8	450
v	Kilic et al. (2012), (5 years of cultivation)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent Clay loam to sandy clay loam	Crop/ horticulture		Bouyoucos hydrometer	0.8	46	21.7	6.3		Spherical	0.65	1.41	310.9
w	Tabor et al. (1985)	Arizona		Haplargid/mollisol	Crop			13	49	32.1	5.8	0–20	Linear	1.68	60.32	500

x	Kilic et al. (2012)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent Clay loam to sandy clay loam	Marshy plants		Bouyoucos hydrometer	0.8	46	29.8	7.4	0–20	Spherical	0.69	0.81	207.8
y	Ayoubi et al. (2007)	Golestan Province, Iran		Fine, mixed, thermic, fluventic haploxerepts	Crop	Before plant-ing	Hydrometer method (day, 1965);	1.8	101	56.3	1.1		Spherical	0.67	0.66	25.83
z	Kilic et al. (2012) cultivation	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent Clay loam to sandy clay loam	Crop/ horticul-ture		Bouyoucos hydrometer	0.8	46	20.4	5.2		Linear	0.99	0	85.66
aa	De Souza (2010)-site two	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Cropping	Fall of 2004		22	90	34	0.2	0–20	Spherical	0	0.68	340
ab	de Souza et al. (2010) – site one	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Cropping	Fall of 2003		20	80	26.1	0.5	0–20	Spherical	0.04	0.42	393
ac	Kerry and Oliver (2007)							6.9	109	4.1	1		Spherical	0.06	0.39	96.17
ad	Molin et al. (2013)- field one	Sao Palo, Southeast Brazil		Typic dystrudoxes				19	42	21.5	5.1	0–20	Spherical	0	0.15	153
ae	Molin et al. (2013) – field two	Sao Palo, Southeast Brazil		Typic hapludoxes				22	92	23.1	3.7	0–20	Spherical	0.01	0.02	307.3
af	Farooque et al. (2012): site one	Central Nova Scotia, Canada	Sandy loam, ferric podzols					1	56	41.9	4.4	0–15	Linear	0.01	19.16	23.7

[a]Unless otherwise specified mm refers to annual average rainfall, and C refers to average annual temperature

Table 21.2 Compilation of key properties for sand variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Adderly et al. (1997)	Nigeria			Trees			30	157	56.8	10	0–10	Experimental	10	190	250
b	Lopez-Granados et al. (2002) site two	Caracol, Southern Spain		Vertic Xerochep	Crop		Bouyoucos densimeter	6	84	33.4	13.4	0–10	Spherical	85	99	84.8
c	Farooque et al. (2012) – site two	Central Nova Scotia, Canada		Sandy loam, orthic humo-ferric podzols	Orchard			1.6	86	48.7	12.5	0–15	Spherical	68.8	106.4	67.9
d	Lopez-Granados et al. (2002) site one	Monclova, Southern Spain		Alfisol	Crop		Bouyoucos densimeter	11.2	80	57.4	8.5	0–10	Nugget	69	0	0
e	Liu et al. (2008)	Henan Province, Central China	Monsoon 14.6C, 680 mm	Clay loam to loam,	Tobacco plantation	Before sowing	Bouyoucos densimeter	87	81	43.8	4.8	0–20	Spherical	7.7	32.43	657
f	Shatar (1996)	Moree, Australia		Vertisol	Crop			15	114			15–30	Spherical	20	17	300
g	Miller et al. (1988)	Sacramento, California		Xererts, Xerochrepts	Crop			8	99	20.2	5.9	Surface	Spherical	0.6	33.4	75
h	Farooque et al. (2012): site one	Central Nova Scotia, Canada		Sandy loam, orthic humo-ferric podzols	Orchard			1	56	49.5	4.5	0–15	Linear	18.75	18.75	85.86
i	Wiliams et al. (1987), four directions averaged	Oklahoma, USA		Paleustolls	Crop			1.62	108	27	4.8	0–30	Experimental	8	9.25	62.5

j	Tabor et al. (1985)	Arizona, USA		Haplargid/mollisol	Crop			13	49	41.7	8.4	0–20	Linear	15.4	101.4	300
k	Shouse et al. (1990)	Texas, USA		Vertisol				0.3	205	21.1	3.3	0–30	Spherical	6.2	5.2	22.86
l	Campbell (1977) Pawnee	Kansas, USA						1.6	160	8.5	2.9		Experimental	2.2	0.8	40
m	Kilic et al. (2012)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent clay loam to sandy clay loam	Crop/ horticulture		Bouyoucos hydrometer	0.8	46	25.5	9.7	0–20	Exponential	0.75	1.13	200.4
n	Kilic et al. (2012) – 5 years of cultivation	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent Clay loam to sandy clay loam	Crop/ horticulture		Bouyoucos hydrometer	0.8	46	33.5	10.3		Exponential	0.87	0.87	306
o	Kilic et al. (2012) – cultivation	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent Clay loam to sandy clay loam	Crop/ horticulture		Bouyoucos hydrometer	0.8	46	50.8	7.9		Linear	1.02	0	85.66
p	Campbell (1977) Ladysmith	Kansas, USA						1.6	160	1.2	0.3		Experimental	0.04	0.04	30
q	Ayoubi et al. (2007)	Golestan Province, Iran		Fine, mixed, thermic, fluventic haploxerepts	Crop	Before planting	Hydrometer method	1.8	101	2.2	0.1	0–30	Gaussian	0.01	0	91.41

Table 21.3 Compilation of key properties for pH variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Kilic et al. (2012) (unmodified)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent	Crop		Water 1:2.5	0.8	46	7.9	0.1	0–20	Exponential	0.8	0.82	209.6
b	Sidorova et al. (2012)	Northwestern Russia			Crop	Before sowing	KCl		72	5.6	0.2	23	Spherical	0.08	1	93.1
c	Kilic et al. (2012) (20 year)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent.	Crop		Water 1:2.5	0.8	46	8.1	0.2	0–20	Linear	1.03	0	85.66
d	Kilic et al. (2012) (5 year)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent	Crop		Water 1:2.5	0.8	46	7.8	0.2	0–20	Linear	1.01	0	85.66
e	Thompson et al. (2004) – site one	Alabama, USA			Crop			0.4	71			11*	Spherical	0.05	0.71	186.2
f	Pierce et al. (1995) (Durand)	Durand, USA		Alfisol, fine loam	Crop			16	165	6.6	0.9	0–5	Spherical	0.13	0.47	354
g	Machado et al. (2007)	Uberlandia, Brazil	Tropical humid with dry season	Red latosol, moderate clayey texture	Crop		Water: 1:1	25	121	6	0.4	0–20	Spherical	0.07	0.51	150
h	Thompson et al. (2004) (site three)	Alabama, USA			Crop			0.4	48			11	Spherical	0.11	0.43	168.9
i	Mulla (1993)	Washington, USA		Ultic haploxeroll	Crop			8	172	6.1	0.7	0–30	Spherical	0.17	0.26	132
j	Pierce et al. (1995) (Adrian)	Adrian, USA		Alfisol, loam	Crop			10	74	6.5	0.9	0–5	Spherical	0.08	0.32	95
k	Camacho-Tamayo et al. (2007) (site one)	Puerto Lopez, Colombia	2,375 mm 27C	Typic haplustox	Crop	May 04	Water (1:1) potentiometer	1.875	42	5.1	0.4	0–10	Exponential	0.09	0.31	410
l	Webster and McBratney (1987)	Suffolk, England			Crop		Water	77	436	7.7	0.6		Spherical	0.02	0.33	185

m	Uehara et al. (1985)	Sitiung, Indonesia			Crop			0.0784	137				Experimental	0	0.35	4
n	de Souza et al. (2010) (site one)	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Crop	Fall of 2003		20	80	5.6	4	0–20	Exponential	0.07	0.23	148
o	Adderley et al. (1997)	Nigeria			Forestry			30	480	6.3	0.5	0–10	Exponential	0.15	0.1	250
p	de Souza et al. (2010) (site two)	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Crop	Fall of 2004		22	90	5.7	0.5	0–20	Spherical	0.01	0.23	97
q	Pierce et al. (1995) (Plainwell)	Plainwell, USA		Entisol, loamy sand	Crop			22	174	6.7	0.4	0–20	Spherical	0.06	0.15	190
r	Thompson et al. (2004) (site two A)	Alabama, USA			Crop			0.2	124			22	Spherical	0.09	0.1	393.2
s	Birrell et al. (1996)	Missouri, USA						28	504			0–15	Spherical	0.06	0.11	125
t	Thompson et al. (2004) (site two B)	Alabama, USA			Crop			0.4	58			22	Spherical	0.05	0.05	181.1
u	Kristensen et al. (1995) (Riso)	Riso, Denmark		Sandy loam/sandy clay loam	Crop			10.9	270			0–25	Exponential	0	0.1	17
v	Kristensen et al. (1995) (Vindum)	Vindum, Denmark		Sandy loam	Crop			10	302			0–25	Exponential	0	0.09	19
w	Liu et al. (2008)	Henan Province, Central China	Monsoonal 14.6C, 680 mm	Clay loam to loam,	Crop	Before sowing	Water (1: 2.5)	87	81	7.7	0.3	0–20	Spherical	0.01	0.06	308
x	Mondo et al. (2012)	Sao Paulo, Brazil	Not specified		Crop	After		22	33	5	0.2	0–20	Gaussian	0.02	0.05	650

(continued)

Table 21.3 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
y	Laslett et al. (1987)	Brisbane, Australia			Pasture		CaCl2	1	121	4.5	0.2		Spherical	0.02	0.03	55
z	Campbell (1977) (Pawnee)	Pawnee. Kansas, USA					CaCl2 (1:2)	1.6	160	6.5	0.3		Experimental	0.04	0	0
aa	Silva et al. (2003)	Santa Maria, Brazil		Ultisol dystrophic hapludalf	Crop	Before sowing	Water 1:1	0.3	192	4.9	0.1	0–20	Spherical	0	0.02	18.66
ab	Campbell (1977) (Ladysmith)	Ladysmith, Kansas, USA					CaCl2 (1:2)	1.6	160	6.5	0.2		Experimental	0.02	0	0
ac	Shatar (1996)	Moree, Australia		Vertisol	Crop		CaCl2 (1:5)	15	114	7.4	0.2	5:30	Spherical	0.01	0.01	310
ad	Tabor et al. (1985)	Arizona		Haplargid/Mollisol	Crop			13	49	7.3	0.2	0–20	Linear	0.02	0.06	500
ae	Lopez-Granados et al. (2002) (site one)	Monclova, Southern Spain		Alfisol	Crop	Before sowing	0.1 mol KCl	11.2	80	7.8	0.1	0–100	Spherical	0	0.01	66
af	Lopez-Granados et al. (2002) (site two)	Caracol, Southern Spain		Vertic Xerochep	Crop	Before	0.1 mol KCl	6	84	7.7	0.1	0–100	Nugget	0.01	0	0
ag	Shukla et al. (2004)	Gross-Enzersdorf, Austria	510 mm 10C	Loam to sandy loam	Crop	May		6.25	60	NA	NA	0–15	Spherical	0.01	0	158
ah	Ayoubi et al. (2007)	Golestan Province, Iran		Fluventic haploxerepts	Crop	Before sowing	0.1 mol KCl	1.8	101	7.9	0	0–30	Spherical	0.001	0.001	24.39

*Assuming 11 cm were taken from topsoil.

Table 21.4 Compilation of key properties for carbon variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Farooque et al. (2012) (site one −1)	Central Nova Scotia, Canada		Sandy loam, podzols	Orchard	May July 2009	Om (LOI)	1	56	6.6	1.5	0–15	Sph.	1.14		
b	Farooque et al. (2012) (site one −2)	Central Nova Scotia, Canada		Sandy loam, podzols	Orchard	Jun 10	Om (LOI)	1	56	6.6	1.4	0–15	Sph.	1.11	2.28	76.1
c	Panagopoulos and Antunes (2008)	Algarve region of south Portugal	Medi-terranean	Mostly lithosols	Forestry, crop		OM (WB-wet)	400	81	1.2	0.5	0–20	Exp.	1.63	2.14	70.23
d	Farooque et al. (2012) (site two −2)	Central Nova Scotia, Canada		Sandy loam, podzols	Orchard	Jun 11	Om (LOI)	1	56	24	5.7	0–16		0.25	1.95	1997.8
e	Farooque et al. (2012) (site two −1)	Central Nova Scotia, Canada		Sandy loam, podzols	Orchard	May July 2009	Om (LOI)	1.6	86	4.9	1.3	0–15	Exp.	0.22	1.86	13.7
f	Kilic et al. (2012) (5 years)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent	Crop		OM (WB-wet)	0.8	46	1	0.3	0–20	Exp.	0.09	1.82	14.6
g	Nolin et al. (1996)	Quebec, Canada		Aquepts	Crop		OC	10	130	2.8	0.7	0–15	Sph.	0.08	1.23	224.5
h	Kristensen et al. (1995), (Riso)	Riso, Denmark		Sandy loam-sandy clay loam	Crop		OC	10.9	270			0–25	Exp.	0	0.61	343
i	Kilic et al. (2012) (10 years)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent	Crop		OM (WB-wet)	0.8	46	0.7	0.3	0–20	Exp.	0	0.43	99
j	Kilic et al. (2012) (uncultivated)	Kaz Lake of Tokat, Turkey	436 mm, 12C	Typic ustifluvent	Crop		OM (WB-wet)	0.8	46	1.6	0.4	0–20	Lin.	0.33	0.35	9.6
k	Mulla (1993)	St. John, Washington, USA		Ultic haploxeroll	Crop		OC	8	172	1.2	0.5	0–30	Sph.	0.05	0	85.66

(continued)

Table 21.4 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
l	Gutierrez et al. (2010)	Municipality of Pasca, Colombia	1,800 mm, 16¬C	Entisols and others	Horticulture		OC	1.5	64	6.4	0.3	0–20	Sph.	0.17	0.19	114
m	Nanni et al. (2011)	Sao Paulo state, Brazil	Meso-thermic climate	Oxisols, entisols, alfisols, ultisols, inceptisols and mollisols	Crop		OM[b]	184	184	0.7	0.4	0–20	Sph.	0.05	0.24	10,240
n	Zanão Júnior et al. (2010)	South-eastern Brazil	1,500 mm	Oxisol hapludox (medium clay)	Crop	May to June 2003	OM	25	121	1.7	0.2	0–10	Gaus	0.05	0.15	691
o	Cahn et al. (1994) 0.25 ha	Central Illinois		Mollisol	Crop		OC	0.25	200	1.7	0.3	0–15	Exp.	0	0.15	539
p	Kristensen et al. (1995), (Vindum)	Vindum, Denmark		Sandy loam	Crop		OC	10	302			0–25	Exp.	0	0.18	50
q	Shukla et al. (2004)	Gross-Enzersdorf, Austria	Temperate and conti-nental, 510 mm 10C	Loam to sandy loam	Crop	May	OC	6.25	60	1.4	0.2	0–15	Sph.	0.04	0.12	45
r	Rowlands 1998; cited in McBratney & Pringle (1999)[c]	Wyalkatchem, WA		Duplex	Crop		OC	75	56	0.7	0.2	0–10	Sph.	0.03	0.03	163
s	de Souza et al. (2010) (site two)	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Crop	Fall of 2004	OM	22	90	15.1	9.7	0–20	Sph.	0.02	0.03	638.985
t	Mondo et al. (2012)	Sao Paolo, Brazil			Crop	After harvest	OM	22	33	0.7	0.2	0–20	Gaus	0.01	0.04	304
u	Camacho-Tamayo et al. (2008) (site 1)	Puerto Lopez, Colombia	2,375 mm: 27C	Typic haplustox	Crop	May 04	OC (mod-WB)	1.9	42	1.3	0.2	0–10	Exp.	0	0.05	498.2

v	Shukla et al. (2004)	Gross-Enzersdorf, Austria	Temperate and continental, 510 mm 10C	Loam to sandy loam	Crop	May	OC	6.25	60	1	0.2	0–15	Sph.	0.02	0.04	22.1
w	Shatar (1996)	Moree, Australia		Vertisol	Crop		OC	15	114	1.1	0.2	15–30	Sph.	0.01	0.02	184
x	Zhang et al. 2016	Jiangsu Province, China			Crop	2012	OM	7	136				Sph.	0	0.03	280
y	de Souza et al. 2010 (site one)	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Crop	Fall of 2003	OM	20	80	10.7	1.4	0–20	Sph.	0.01	0.04	38.6
z	Amirinejad et al. 2011	Uttar Pradesh, India		Inceptisol	Crop		OC (WB)	19.6	145			0–15	Gaus	0.01	0.02	324
aa	Kumhalova et al. 2011 (sampled 2005)	Prague, Ruzyne, Czech Republic	526 mm: 7.9C	Haplic luvisol.	Crop		OC	11.5	70	2.1	0.2		Sph.	0.02	0.02	380.73
ab	Camacho -Tamayo et al. 2008 (site two)	Puerto Lopez, Colombia	2,375 mm, 27 C	Typic hapludox	Crop	Aug 04	OC (mod-WB)	1.875	42	1.6	0.1	0–10	Sph.	0	0.01	240.5
ac	Liu et al. 2010	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Sandy to medium clay loam	Crop	After harvest,	OM (WB-wet)	4	111	1	0.1	0–20	Sph.	0.01	0.02	33.5
ad	Liu et al. 2008	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Clay loam to loam	Crop	Before sowing	OM[d]	87	81	0.7	0.1	0–20	Sph.	0.01	0.01	56.5
ae	Ayoubi et al. 2007	Golestan Province, Iran		Fluventic haploxerepts	Crop	Before planting	Om (WB)	1.8	101	1.5	0.1	0–30	Sph.	0.01	0.01	556
af	Silva et al. 2003	Santa Maria, Brazil		Ultisol dystrophic hapludalf	Crop	Before sowing	OM [e]	0.3	195	1.6	0.1	0–20	Gaus	0.01	0.01	29.28

(continued)

Table 21.4 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
ag	Kumhalova et al. 2011 (sampled 2004)	Prague, Ruzyne, Czech Republic	526 mm: 7.9C	Haplic luvisol	Crop		OC	11.5	70	1.7	0.2		Sph.	0.01	0.01	9.5
ah	Kumhalova et al. 2011 (sampled 2006)	Prague, Ruzyne, Czech Republic	526 mm: 7.9C	Haplic Luvisol	Crop		OC	11.5	70	1.9	0.2		Sph.	0.01	0.01	274.6
ai	Goovaerts and Chiang 1993 (average)	Belgium		Typic hapludalf			OC	0.16	73	0.7	0.1	0–20	Sph.	0.01	0.05	244.7
aj	Chung et al. 2008	Korea	12.7C, 1,560[a]	Coarse, loamy, mixed, non-acid, mesic	Crop	After harvest	OM[f]	0.3	246	1.3	0.1	0–15	Lin.	0.01	0	11.5
ak	Lopez-Granados et al. (2002) site one	Monclova, Southern Spain		Alfisol	Crop	Pre planting	OM[g]	11.2	80	0.9	0.1	0–10	Nug.	0.01	0	NA
al	Lopez-Granados et al. (2002) site two	Caracol, Southern Spain		Vertic Xerochep	Crop	Pre planting	OM[g]	6	84	0.8	0.1	0–10	Sph.	0	0	0
am	Chatterjee et al. 2015	West Bengal, India	1,443 mm, hot and humid		Crop	After harvest	OC (WB)	81	100	0.5	0.1	0–15	Lin.	0	0.01	44.8
an	Kumhalova et al. 2011-(sampled 2007)	Prague, Ruzyne, Czech Republic	526 mm: 7.9C	Haplic luvisol	Crop	Not specified	OC	11.5	70	2	0.2		Sph.	0	0.01	58
ao	Bai and Wang 2011	Shaanxi Province, China	393 mm	Silty loams	Orchard		OC[h]	0.275	125	0.3	0.1	0–10	Exp.	0	0.07	247.1

[a] Measurement method refers to the original property which was measured. All values have been converted into OC
[b] Organic matter (OM), total and effective acidity (determined by 1 M calcium acetate – Ca(CH3 COO)2 H2 O and 1 M KCl titulometric method, respectively),
[c] Unpublished data provided by Pringle for the publication of McBratney & Pringle (1999)
[d] Potassium bichromate titrimetric method
[e] Photocolorimetry
[f] Laboratory analysis was completed by the Soil Management Division – NIAST, RDA
[g] Redox-Electrode (Methrom Titroprocesador)
[h] Oil bath titration

Table 21.5 Compilation of key properties for available nitrogen variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Liu et al. (2010)	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Sandy to medium clay loam	Crop	After harvest	Alkaline hydrolysable N (AN)	4	111	75	14.3	0–20	Spherical	105.6	127.3	112.6
b	Liu et al. 2008	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Clay loam to loam	Crop	Before sowing	Alkalytic N (AN)	87	81	70.6	10.3	0–20	Linear	110.46	0	NA
c	Lopez-Granados et al. (2002) (site two)	Caracol, Southern Spain		Vertic Xerochep	Crop	Before winter sowing	Nitrate	6	84	23.2	10.2	0–10	Exponential	10	91	31.7
d	Chatterjee et al. (2015)	West Bengal, India	Hot and humid		Crop	After harvest	Available N	81		76.7	7.8	0–15	Spherical	17.38	45.26	43
e	Shire 1997; cited in McBratney & Pringle (1999)[d]	Wyalkatchem, Western Australia		Duplex	Crop		Nitrate	90	88	22.9	7.6	0–10	Spherical	22.14	36.86	281.432
f	Lopez-Granados et al. (2002) site one	Monclova, Southern Spain		Alfisol	Crop	Before winter sowing	Nitrate	11.2	80	7.1	5.3	0–10	Nugget	27.2	0	0
g	Cahn et al. (1994) 0.25 ha	Central Illinois		Mollisol	Crop		Nitrate	0.25	200	6.2	3.7	0–15	Experimental	7	3	5
h	Everett and Pierce (1996)	Michigan		Loamy sand, sandy loam	Crop		Nitrate	22.6	60	6.2	2.2	0–30	Spherical	3.05	3.7	167.05

(continued)

Table 21.5 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
i	Tabor et al. (1985)	Arizona		Haplargid/mollisol	Crop		Nitrate	13	49	13.6	4.2	0–20	Linear	6.66	11.12	200
j	van Meirvenne and Hofman (1989)	Belgium		Udifluvent	Horticulture		Nitrate	1	247	8.9	2.7	0–100	Experimental	2.25	4.25	25
k	Wade et al. (1996) (pasture)				Pasture		Nitrate		86	0.8	0.8	0–10	Experimental	0.44	0.57	120
l	Cambardella et al. (1994)	Iowa		Udic/argic mollisols	Crop		Nitrate	10	72	7.2	1.3	0–15	Spherical	0.14	0.2	201
m	Wade et al. (1996), (arable)	Warwickshire, England					Nitrate		81	0.7	0.5	0–10	Experimental	0	0.3	60

[a]Alkaline hydrolysable N was measured using the alkaline hydrolysis diffusion method
[b]Nitrate determined by colorimetry in SKALAR
[c]Available N content was determined by alkaline permanganate method as provided in Chatterjee et al. (2015)
[d]Unpublished data provided by Pringle for the publication of McBratney & Pringle (1999)

Table 21.6 Compilation of key properties for total nitrogen variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Ganawa et al. (2003)	Sawah Sempadan, Malaysia		Jawa, Teluk, Karang, Sempadan and Sedu	Crop	Before planting	[a]	2,300	120	4140	870	0–20	Spherical	100,000	700,000	460
b	Yana et al. (2000)	Takatsuki, Japan	15.8 °C, 1,240 mm	Clay loam		After transplanting	Dry combustion	0.5	91	3,100	449.5		Spherical	30,000	270,000	35.5
c	Yana et al. (2000)	Takatsuki, Japan	15.8 °C, 1,240 mm		Crop	After harvest	Dry combustion	0.5	91	3,400	452.2		Spherical	80,000	200,000	19.5
d	Zhang et al. (2016)	Jiangsu Province, China			Crop			7	136				Exponential	0	180,000	39.08
e	Shukla et al. (2004)	Gross-Enzersdorf, Austria	510 mm, 10C	Loam to sandy loam	Crop	May	Kjeldahl method	6.25		1,328.6	381	0–15	Spherical	106,576	31,746	243.6
f	Nouri et al. (2010)	Fesaran village, Esfahan		Clay loam and loam	Horticulture	Before fertigation	Nitrate	87	60			0–30	Spherical	30,000	30,000	141.8
g	Liu et al. (2008)	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Clay loam to loam	Crop	Before sowing	Kjeldahl method	87	81	730	100	0–20	Spherical	2,000	9,000	274
h	Ayoubi et al. (2007)	Golestan Province, Iran		Fine, mixed, thermic, fluventic haploxerepts	Crop	Before sowing	Kjeldahl method	1.8	101	1,300	110	0–30	Gaussian	600	10,000	23.99
i	Chung et al. (2008)	Korea	12.7C, 1,560 mm[b]	Coarse loamy	Crop	After harvest	[b]	0.3	246	1,500	100		Exponential	0	0	4.5

[a]Sulfuric-salicylic acid digestion method

[b]Laboratory analysis was completed by the Soil Management Division – NIAST, RDA

Table 21.7 Compilation of key properties for phosphorus variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	McBratney et al. (1985)	Narrabri, NSW, Australia		Vertisol	Crop		P (NaHCO3 soluble)	1	386	129.6	58.8	0–7.5	Gaussian	152.85	7,489.55	109.64
b	Chatterjee et al. (2015)	West Bengal, India	Hot and humid 1443 mm		Crop/ horticulture	After harvest	Available P2O5	81		172.4	82.4	0–15	Spherical	3,700.68	3,925.85	283
c	Mondo et al. (2012)	Sao Paolo, Brazil			Crop	After harvest	P resin	22	33	239.6	59.6	0–20	Spherical	1,484.69	2,030.1	89.9
d	Cambardella et al. (1994) (pothole field)	Iowa, USA		Udic/aquic mollisols	Crop		P (bray no. 1)	6.25	241	126	55.6	0–15	Spherical	596.9	2,839	71
e	Delcourt et al. (1996)	Leefdaal, Belgium		Silty	Crop		P (colorimetric)					0–25	Gaussian	310	850	74.5
f	Cahn et al. (1994), 0.25 ha	Central Illinois, USA		Mollisol	Crop		P (bray no. 1)	0.25	200	74	26.8	0–15	Experimental	390	760	50
g	Pierce et al. (1995), (Plainwell)	Plainwell, USA		Entisol, loamy sand	Crop		P (not specified)	22	174	124	32	0–20	Spherical	233	844	172
h	Camacho -Tamayo et al. (2008) (site two)	Puerto Lopez, Colombia	2,375 mm 27 C	Typic hapludox	Crop	Aug 04	P (bray II)	1.875	42	28.8	25.8	0–10	Spherical	2	660.5	38.4
i	Ganawa et al. (2003)	Sawah Sempadan, Malaysia		Jawa, Teluk, Karang, Sempadan and Sedu	Crop	Before planting	Available P NH4	2,300	120	50.2	22.5	0–20	Spherical	69	422	597
j	Nolin et al. (1996)	Quebec, Canada		Aquepts	Crop		P (Melich III)	10	130	52	18.5	0–15	Exponential	35.1	255.5	13

k	Chung et al. (2008)	Korea			Crop	After harvest	[c]	0.3	246	119	15.5		Spherical	17.8	210.3	4
l	Pierce et al. (1995) Durand	Durand, USA		Alfisol, fine loam	Crop		P (not specified)	16	165	35	21	5:20	Linear	223	163	294
m	Camacho -Tamayo et al. (2008) (site one)	Puerto Lopez, Colombia	2,375 mm 27 C	Typic haplustox	Crop	May-04	P (bray II)	1.875	42	11.1	14.2	0–10	Linear	154.4	262.9	373.4
n	Frogbrook et al. (2002) (sampled 1998)					Before fertili-sation		16.5					Exponential	17.95	131.5	17.07
o	Frogbrook et al. (2002) (sampled 1999)							16.5					Spherical	54.58	66.61	54.9
p	Frogbrook et al. (2002) (sampled 1997)							16.5					Spherical	56	59.72	49.95
q	Kristensen et al. (1995) (Riso)	Riso, Denmark		Sandy loam/sandy clay loam	Crop		Available P	10.9	270			0–25	Exponential	0	110	60
r	Shatar (1996)	Moree, Australia		Vertisol			Available P	15	114	18	1	15–30	Spherical	55	50	290
s	Sidorova et al. (2012)	Northwestern Russia			Crop		Available P[d]		72	484.9	65.6		Spherical	11	93	63.7
t	Pierce et al. (1995) (Adrian)	Adrian, USA		Alfisol, loam	Crop		P (not specified)	10	74	23	9	0–5	Exponential	23.6	60.49	68
u	Machado et al. (2007)	Uberlandia, Brazil	Tropical humid with dry season,	Red latosol, moderate clayey texture	Crop			25		12.4	9.1	0–20	Linear	82.33	0	NA
v	Mulla (1993)	St. John, Washington, USA		Silt loam (ultic haploxeroll)	Crop		P (acetate extractable)	8	172	15.2	7.7	0–30	Spherical	27.58	35.8	145

(continued)

Table 21.7 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
w	Rowlands 1998; cited in McBratney & Pringle (1999)[e]	Wyalkatchem, WA, Australia		Duplex	Crop		P (not specified)	75	56	31.9	7	0–10	Spherical	22.32	28.72	414.604
x	Kristensen et al. (1995) (Vindum)	Vindum, Denmark		Sandy loam	Crop		Available P	10	302			0–25	Exponential	0	45	148
y	Liu et al. (2008)	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Clay loam to loam,	Crop	Before sowing,	Available P[f]	87	81	16.1	5.9	0–20	Spherical	13.4	22.59	337
z	Lopez-Granados et al. (2002) (site two)	Southern Spain (Caracol)		Vertic Xerochep	Crop	Before crop	Available P[g]	6	84	11.3	5.6	0–10	Exponential	0.7	33.3	28.3
aa	Lopez-Granados et al. (2002) (site one)	Southern Spain (Monclova)		Alfisol	Crop	Before r crop	Bray extractable P[g]	11.2	80	15.5	4.4	0–10	Exponential	0.6	18.6	27.4
ab	Ayoubi et al. (2007)	Golestan Province, Iran		Fine, mixed, thermic, fluventic haploxerepts	Crop	Before crop	Available P[h]	1.8	101	27.2	1.3	0–30	Spherical	1.08	0.57	35.58

[a]Available P2O5 Olsen method by way of extracting 2.5 g of soil with 50 ml of 0.5MNaHCO3 (pH 8.5) for 30 min and determining the phosphorus in the extract by the L-ascorbic acid method
[b]Available P NH4F-HCl extraction method
[c]Laboratory analysis was completed by the Soil Management Division – NIAST, RDA
[d]Available P (the TsI-NAO modification of the Kirsanov method)
[e]Unpublished data provided by Pringle for the publication of McBratney & Pringle (1999)
[f]Available P Olsen extraction method using alkaline sodium bicarbonate as the extractant in a 20:1 ratio
[g]Bray extractable P measured by colorimetry using ascorbic acid-ammonium molybdate reagents
[h]Available P measured by colorimetry using ascorbic acid-ammonium molybdate reagents

Table 21.8 Compilation of key properties for potassium variograms

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
a	Lopez-Granados et al. (2002) (site two)	Caracol, Southern Spain		Vertic Xerochep	Crop	Before planting	Exchangeable K	6	84	741	156	0–0.10	Spherical	0	15,210	54.6
b	Lopez-Granados et al. (2002) (site one)	Monclova, Southern Spain		Alfisol	Crop	Before planting	Exchangeable K	11.2	80	468	117	0–0.10	Exponential	0	15,210	34.5
c	Cahn et al. (1994) (0.25 ha)	Cemtral Illinois, USA		Mollisol	Crop		K (not specified)	0.25	200	268.2	114.9	0–15	Experimental	5000	7500	50
d	Pierce et al. (1995) (Adrian)	Adrian, USA		Alfisol, loam	Crop		K (not specified)	10	74	210	71	0–5	Exponential	1850	4168	93
e	Frogbrook et al. (2002) (sampled 1997)			Clay loam		Before fertilisation		16.5	110				Spherical	994.5	1831	55.09
f	Chatterjee et al. (2015)	West Bengal, India	Hot and humid, 1443 mm		Crop/ horticulture	After harvest	Available K2O	81		134.8	51.9	0–15	Exponential	1864.51	923.4	55
g	Chung et al. (2008)	Korea	12.7C, 1560[b]	Coarse, loamy, mixed, non-acid, mesic	Crop	After harvest		0.3	246	269.1	27.3		Gaussian	304.2	2281.5	32.9
h	Frogbrook et al. (2002) (sampled 1998)					Before fertilisation	K[c]	16.5					Circular	1121	1266	57.71

(continued)

Table 21.8 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
i	Frogbrook et al. (2002) (sampled 1999)					Before fertilisation		16.5					Pentaspherical	1050	1034	79.03
j	Nanni et al. (2011)	Sao Paulo state, Brazil	Mesothermic climate	Oxisols, entisols, alfisols, ultisols, inceptisols and mollisols	Crop			184	184	61.3	44.6		Spherical	706.18	1373.56	353
k	Pierce et al., (1995) (Plainwell)	Plainwell, USA		Entisol, loamy sand	Crop		K (not specified)	22	174	121	35.9	0–20	Spherical	887	391	157
l	Ferraz et al. (2012) (sampled 2008)	Tres Pontas, Brazil	Mild tropical altitude	Red-yellow latosol	Crop		K Mehlich 1	22	48	7.1	6.5	0–20	Spherical	0	1268.37	437
m	Liu et al. (2010)	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Sandy to medium clay loam	Crop	After harvest	Available K[d]	4	111	161.4	30.8	0–20	Spherical	435	744.1	312.4
n	Silva et al. (2003)	Santa Maria, Brazil		Ultisol dystrophic hapludalf	Crop	Before sowing	Available K (flame photometry)	0.3	194	111	31.9	0–20	Gaussian	527	619	16.6
o	Pierce et al. (1995), (Durand)	Durand, USA		Alfisol, fine loam	Crop		K (not specified)	16	165	97	31	5–20	Spherical	302	833	174
p	Delcourt et al. (1996)	Leefdaal, Belgium		Silty	Crop		K (AES)			NA	NA	0–25	Spherical	425	565	149.7
q	Rowlands 1998; cited in McBratney & Pringle (1999)[b]	Wyalkatchem, WA		Duplex	Crop		K (not specified)	75	56	59.1	24.7	0–10	Spherical	185.17	581.23	800
r	Camacho-Tamayo et al. (2008) (site two)	Puerto Lopez, Colombia	2375 mm 27C	Typic hapludox	Crop	Aug 04	K[e]	1.875	42	74.1	15.6	0–10	Exponential	182.52	349.83	366.6

s	Machado et al. (2007)	Uberlandia, Brazil	Tropical humid with dry season	Red latosol, moderate clayey texture	Crop		K[f]	25	121	79	21.5	0–20	Linear	466.4	0	NA
t	Kristensen et al. (1995) (Riso)	Riso, Denmark		Sandy loam/sandy clay loam	Crop		Available K	10.9	270			0–25	Exponential	0	389	129
u	Kristensen et al. (1995) (Vindum)	Vindum, Denmark		Sandy loam	Crop		Available K	10	302			0–25	Exponential	0	377	25
v	Ferraz et al. (2012) (sampled 2007)	Tres Pontas, Brazil	Mild tropical altitude	Red-yellow latosol	Crop		K Mehlich 1	22	54	11.5	12.5	0–20	Exponential	0	365.82	165
w	Liu et al. (2008)	Henan Province, Central China	Monsoon climate 14.6C, 680 mm	Clay loam to loam, and the soil is slightly alkaline (pH, 7.7)	Crop	Before sowing	Available K (AK)[d]	87	81	105.9	14.3	0–20	Spherical	66.9	149.2	345
x	de Souza et al. (2010) (site one)	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Crop	Fall of 2003		20	80	3.2	1.3	0–20	Spherical	26.1	103.9	185
y	Sidorova et al. (2012)	Northwestern Russia			Crop		Exchangeable K[g]		72	202.5	36.2		Exponential	7	106	72.8
z	Ayoubi et al. (2007)	Golestan Province, Iran		Fine, mixed, thermic, fluventic haploxerepts	Crop	Before sowing	Available K[h]	1.8	101	334.6	8.1	0–30	Spherical	17.16	62.4	93.92
aa	Camacho-Tamayo et al. (2008) (site 1)	Puerto Lopez, Colombia	2375 mm 27C	Typic haplustox	Crop	May 04	K[e]	1.875	42	70.2	7.8	0–10	Nugget	60.84	0	0

(continued)

Table 21.8 (continued)

	Reference	Location	Climate	Soil type	Land use	Timing	Measurement method	Study area (ha)	Sample size	Mean	Standard deviation	Depth (cm)	Model	Nugget	Partial sill	Range
ab	Thompson et al. (2004) (site two A)	Alabama, USA			Crop	January		0.2	124			22	Exponential	0.1	55.1	228.3
ac	Thompson et al. (2004) (site two B)	Alabama, USA			Crop	January		0.4	58			22	Spherical	23.16	27.52	299
ad	de Souza et al. (2010) – (site two)	Araras, Southeast Brazil	Wet summer, dry winter	Oxisol (typic haplustox)	Crop	Fall of 2004		22	90	3.7	1.9	0–20	Exponential	0.9	3.2	121
ae	Thompson et al. (2004) (site one)	Alabama, USA			Crop	September		0.4	71			11	Spherical	0.01	0.06	295.2
af	Thompson et al. (2004) (site three)	Alabama, USA			Crop	September		0.4	48			11	Spherical	0	0.03	160.4

[a] Exchangeable K atomic absorption spectrophotometry (AAS)
[b] Unpublished data provided by Pringle for the publication of McBratney & Pringle (1999)
[c] Laboratory analysis was completed by the Soil Management Division – NIAST, RDA
[d] Available K (AK) neutral ammonium acetate extraction method
[e] K extraction with ammonium acetate pH 7.0
[f] HCl 0,05 mol L-1 + H2SO4 0,025 mol L-1
[g] Exchangeable K TsI-NAO modification of the Kirsanov method
[h] Available K extraction with ammonium acetate (1 N)

21.5.1.2 Total Sill

Between studies the total sill (nugget plus partial sill) changes by several orders of magnitude for each property. As expected, this variability is the least pronounced for bounded properties (pH, OM %, sand % and clay %) and much more pronounced for micronutrients, which vary by around three to five orders of magnitude.

Variability in total sill is similar to that observed by McBratney and Pringle (1999). Inspection of the summary tables (Tables 21.1, 21.2, 21.3, 21.4, 21.5, 21.6, 21.7 and 21.8) indicates that for the majority of soil properties, the range in the total sill is similar for the variograms collected by McBratney and Pringle (1999) and the more recently collected properties. The maximum variability reached within 1 km has remained within an order of magnitude for all properties. For soil pH, organic carbon and potassium, the maximum variability found in the more recent literature search is two to three times greater than the maximum variability found in the literature reviewed by McBratney and Pringle (1999). The other properties have a very similar maximum.

21.5.1.3 Nugget

Like McBratney and Pringle (1999), we find wide variability (several orders of magnitude) in the nugget parameter. McBratney and Pringle (1999) suggest that this variability is largely due to the strong effect of sampling design on the nugget. A variogram can only model the spatial structure that is detectable by the sampling design. In general, the wider the spacing, the more spatial variability will be attributed to the nugget component of the model. If the survey spacing is wider than the spatial structure, the variogram will appear as a pure nugget model. A wider support and the use of aliquotting will reduce the 'noise' in the data and decrease the nugget. It has been often proposed, and is quite likely, that the majority of soil properties would have more than one layer of soil structure. The differences in estimated nugget likely reflect both the underlying differences in spatial structure at the field scale and the capacity of different survey designs to capture this variability.

21.5.1.4 Range

Like the lag and the sill, the range estimated by the theoretical variogram will be strongly affected by the sampling design. As the lag increases, the confidence interval around the variance increases (Oliver and Webster 2014). Precision agriculture surveys are typically conducted over areas from a few hectares to a few hundred hectares which limits the extent of spatial variability they can capture. Despite this, the majority of variograms fit from precision agriculture studies are bounded. The modelled range and sill can be thought of as the 'local sill' and 'local range' associated with the particular extent and spacing of that survey. It is very likely that if the extent of the survey was increased, another degree of variability would be

found. It is critical to consider the potential effect of both the extent and the spacing on the range when considering precision agriculture studies and how their findings may inform your own work.

21.5.2 Field Scale Soil Variograms: Methodological Differences

21.5.2.1 Survey Design

Another substantial difference between the survey designs was whether or not aliquott ing was used. This is particularly significant when comparing these spatial studies because some studies model the range at distances that other studies were combining soil at. This is typically done for samples taken within 1 m of each other. This practice is likely to reduce the nugget (or white noise) and also to reduce any short-term spatial trends which might be occurring.

There is significant variation in the survey design which may influence the mapping of spatial variability. Nested designs are better placed to capture spatial trends across a variety of scales than designs with even spacing; however, because of the additional costs associated with these, they are less common.

21.5.2.2 Model Fitting Process

Oliver and Webster (2014) wrote an explanatory piece of work, describing the best methods for soil scientists to model variograms for kriging. They also described common mistakes made by soil scientists when calculating variograms. The majority of papers we assess do not follow all of Oliver and Webster's recommendations for reporting methods. This makes it difficult to assess how well a fitted variogram captures underlying spatial variability. Few papers report summary statistics for a variety of models, and few papers present variogram clouds to illustrate the utility of the fit. This does not necessarily mean that the fitted models are not accurate, but it does make it difficult to assess the model.

Another point worth considering is the possibility that trends are being overfitted. Perhaps some of the models presented in this chapter would have been better represented by a nugget. These issues around model quality are not new, but a degree of caution is required when interpreting the results.

The more recent literature has included studies which have found much lower values for the total sill for several soil attributes. For clay the lowest value found for the total sill is two orders of magnitude lower than the lowest value reported by Petit and McBratney. Sand is one order of magnitude lower. Some modelled variograms occur over a very tiny range of variability relative to the magnitude of the property they are measuring. Whether it is necessary or feasible to model a spatial structure of less than 1% for soil texture properties is questionable.

21.5.2.3 Measurement Methods

The properties we have included here are commonly measured soil properties with known agronomic implications. However, measurement of these properties is rarely simple or consistent. Differences in measurement methods and differences in which component of the property is being measured will influence both the shape and magnitude of the variogram.

pH is an extremely commonly measured property. However, within the studies we have assessed, there are differences in the solution, the dilution rate and the equipment used to measure the pH. This problem becomes more complex when considering more difficult-to-measure properties such as potassium and phosphorus. Different studies have used different extractants to target different fractions for these nutrients.

The variogram is affected by the distribution of the property it is being calculated for. The variograms calculated for potassium and phosphorus show the greatest differences in total sill. We expect that this is because the target of the measurements varies as well as the measurement method used.

Some articles were found which estimated total carbon or inorganic carbon. However, there were relatively few such studies, so we have not included them here. We have included studies which measured organic matter as a proxy for organic carbon. We converted these using the van Bemmelen factor (1.724). Pribyl (2010) illustrates that an accurate conversion factor for different soils can vary from 1.4 to 2.5. Error in the conversion will be small relative to the overall spread of the variograms.

21.5.3 Field Scale Soil Variograms: A Compilation

Figures 21.2, 21.3, 21.4, 21.5, 21.6, 21.7, 21.8. and 21.9. provide a visual compilation of field scale variograms for each of the soil properties initially examined by McBratney and Pringle (1999). We include both the original variograms used by McBratney and Pringle and variograms published since then. We only include variograms which were calculated from untransformed data and which were based on physical observations (i.e. not from remotely observed data). The black bold lines represent the average variogram (Sect. 21.7). Tables 21.1, 21.2, 21.3, 21.4, 21.5, 21.6, 21.7 and 21.8 correspond to each figure and include reference details and key parameters for each variogram. Because of the wide range of values of the source variograms, the scales used in the figures cannot include all of them, and some low sill variograms have not been included. We include the details in the tables for completeness, but suspect that they are unlikely to provide useful information. The figures affected and the number of source variograms not included are pH 1, carbon 3, total nitrogen 1 and potassium 2.

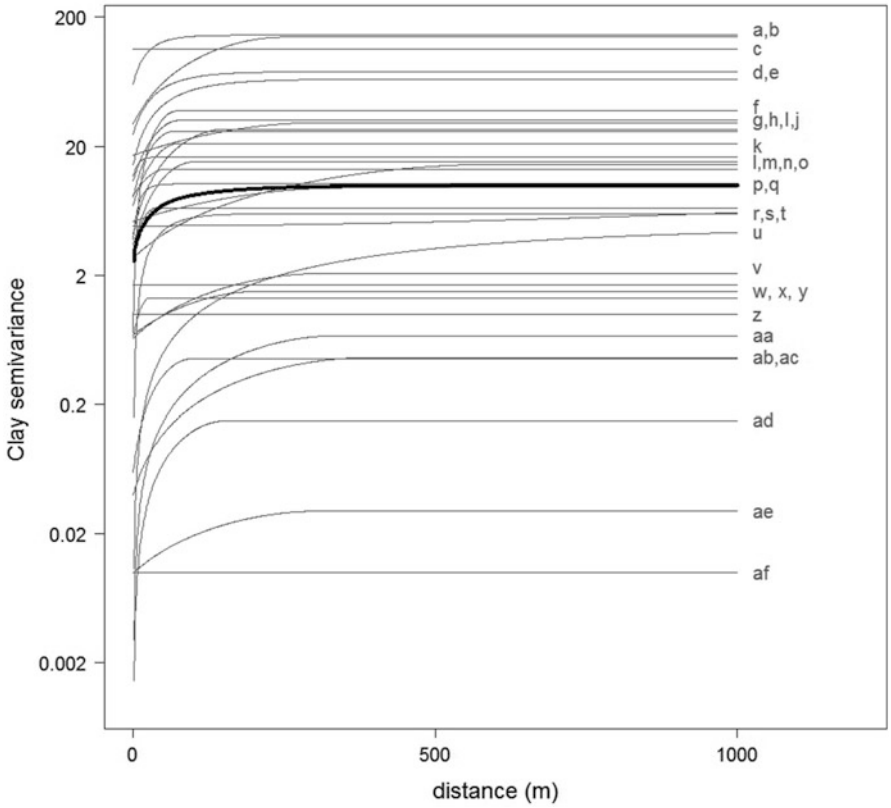


Fig. 21.2 Compilation of field scale variograms for clay. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–ag*) are given in Table 21.1

21.6 Estimation of the Variogram from Proportional Variograms

McBratney and Pringle (1999) observe a strong relationship between mean squared and variance for several soil properties and develop a method for estimating a 'proportional variogram'. This method has the advantage of capturing the much higher levels of variability that tend to occur when the mean values of the property are extreme. Similar to McBratney and Pringle (1999), we find that some soil properties (phosphorus, nitrogen, potassium and carbon) appear to have a strong linear relationship between the mean and the standard deviation. This could imply that the calculation of a proportional variogram would be useful for these properties. However, closer interrogation reveals that this relationship is largely driven by high leverage points. It is not possible to fit a robust curve to link the mean and standard deviation. Proportional variograms are based on the relationship between the mean

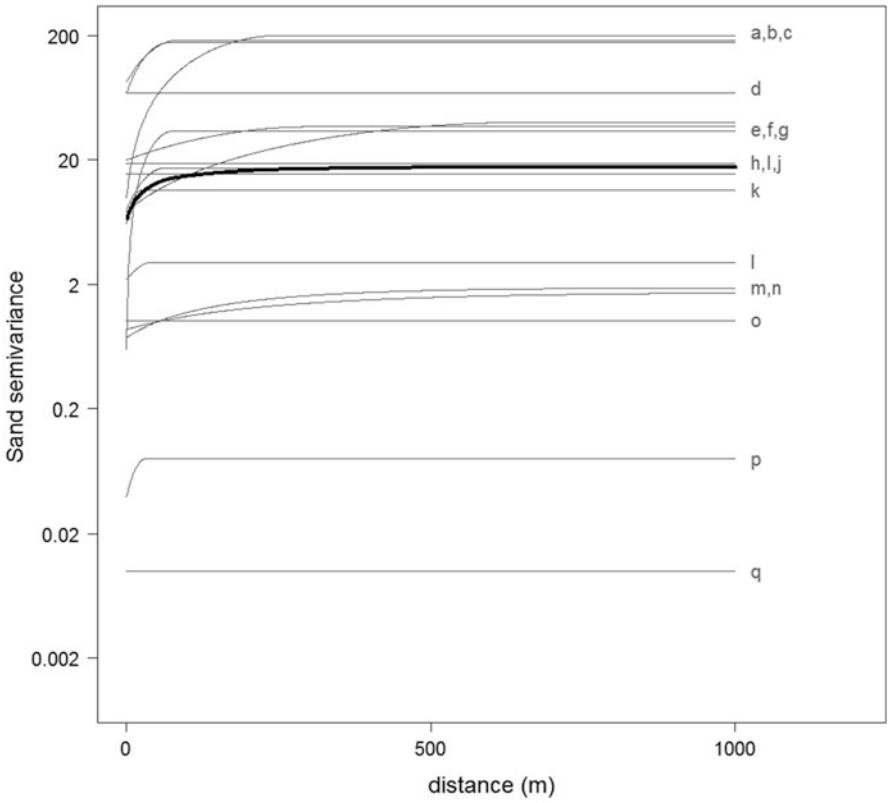


Fig. 21.3 Compilation of field scale variograms for sand. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a*–*r*) are given in Table 21.2

and the variance. Because we cannot be confident about this relationship, it is not prudent to calculate proportional variograms.

We advise against the use of proportional variograms as an estimate for variability and as such do not update McBratney and Pringle's (1999) **estimates of proportional variograms.**

21.7 Estimation of the Variogram from Average Variograms

McBratney and Pringle (1999) calculate average variograms for seven soil properties. These average variograms are calculated from the sample of variograms they compile. The fourth root transform of each approximated by the spherical model is taken and then backtransformed. Exponential or spherical models are fitted.

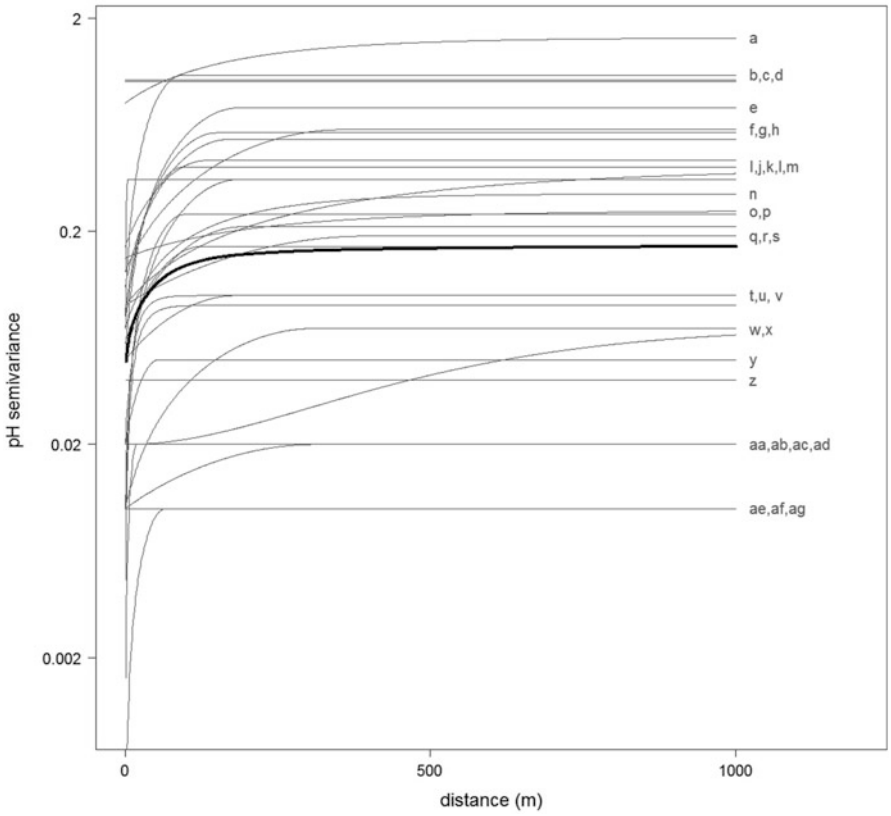


Fig. 21.4 Compilation of field scale variograms for pH. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–ag*) are given in Table 21.3

McBratney and Pringle (1999) note the broad spread of variability in the variograms they collected and suggest that this might make the 'average' variograms less useful. Despite this, they suggest that the average variogram is a useful starting point where no other information is available. Kerry and Oliver (2003) find evidence that average variograms can be useful for prediction when parent material and soil forming factors are similar, but emphasise that they do not expect a global average variogram to provide much useful information.

We do not believe a global average provides useful information for predictive purposes. However, as McBratney and Pringle (1999) suggested, the average variogram does provide a useful reference for those interested in soil variability. We produce average variograms for illustrative purposes (by averaging the fourth root transform of each variogram at finely spaced intervals and then plotting the backtransformed values), but we do not fit these with a functional form.

As suggested by McBratney and Pringle (1999) and illustrated by Kerry and Oliver (2004), the concept of the average variogram has the most use for prediction

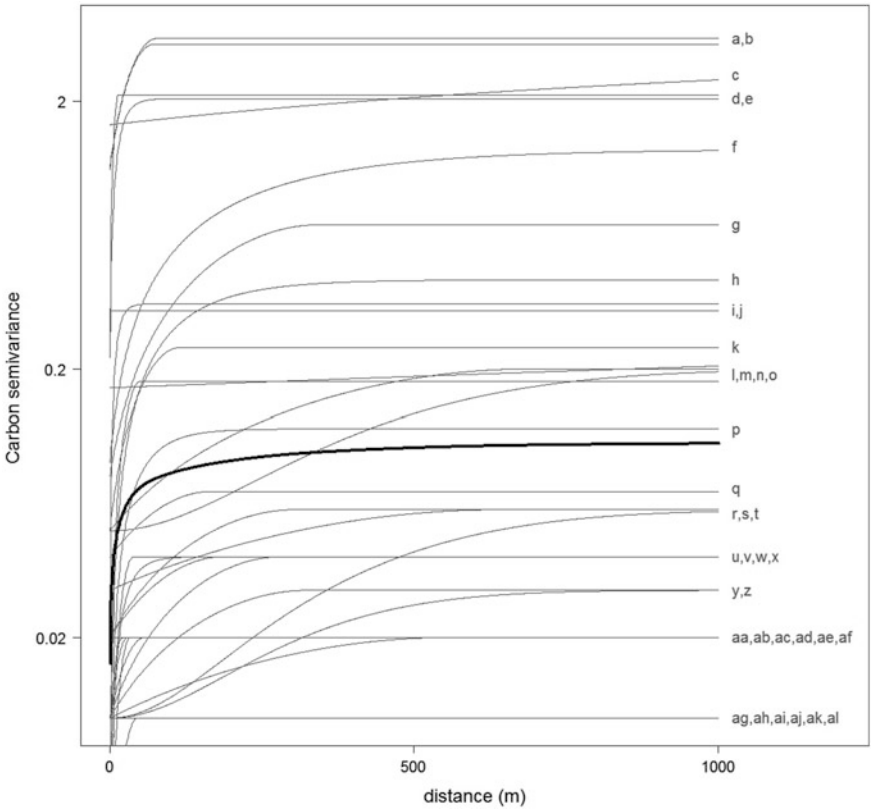


Fig. 21.5 Compilation of field scale variograms for carbon The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–al*) are given in Table 21.4

when it is calculated from a select subset of existing variograms. The diversity of climate regimes, soil types and land use for which field scale variograms have been calculated means that discretion is essential in the selection process. Differences in sampling regime, soil measurement protocols and geostatistical methods add another layer of complexity that needs to be navigated in appropriate selection. We discuss these issues further in the next section.

We advise against the use of the 'global average' variogram as an estimate of local soil variability. Instead, where suitable variograms are available, an average of variograms with similar conditions is taken. Discretion and expert knowledge will need to be used in this selection process. The process outlined by McBratney and Pringle (1999) **for calculation of an average variogram can be followed.**

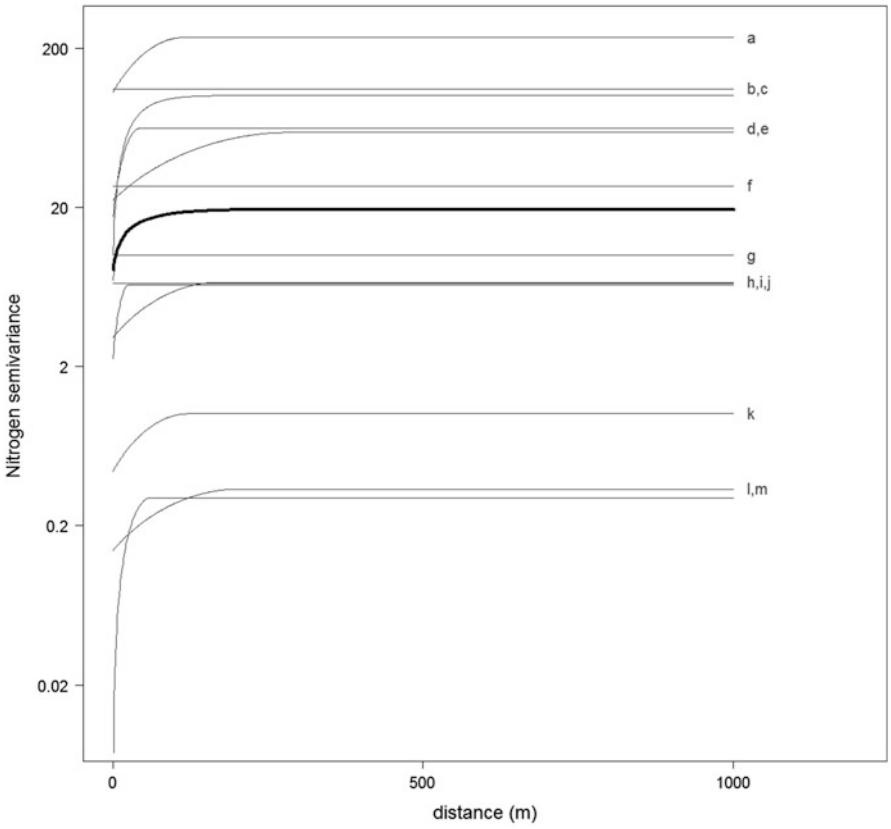


Fig. 21.6 Compilation of field scale variograms for available nitrogen. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–m*) are given in Table 21.5

21.8 Ancillary Information

Kerry and Oliver have written several papers investigating the potential of using cheaper more densely available ancillary information to supplement expensive and sometimes sparse soil survey data. In 2004, they compared the spatial structure of a number of ancillary data sources to the spatial structure of a number of fixed soil properties. They found that variograms calculated from aerial colour photographs of 3.4 m 'sampling density' can estimate range with sufficient accuracy to helpfully guide sampling density of the soil survey. Kerry and Oliver (2008) paper extends the use of ancillary information to kriging. They suggest that the primary requirement for using data is that the ancillary variogram has a similar sill-to-nugget ratio to the property being studied.

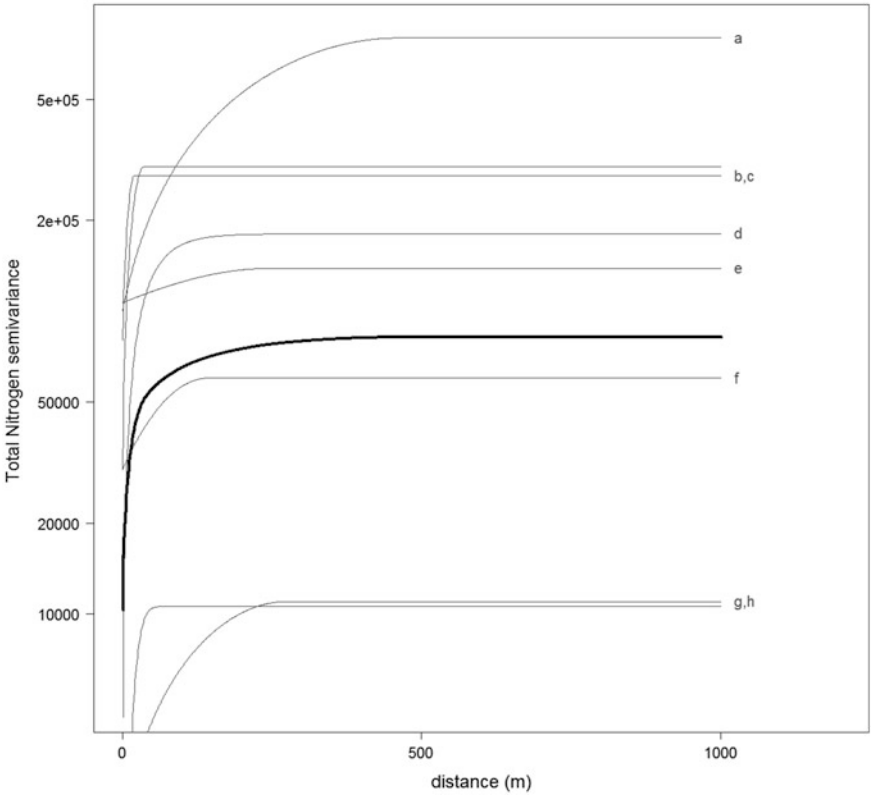


Fig. 21.7 Compilation of field scale variograms for total nitrogen. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–h*) are given in Table 21.6

Where it is available, ancillary information can be a useful source of information. Care must be taken in the selection of appropriate ancillary information. It may be useful to consult a range of ancillary variables.

21.9 Expert Knowledge

Truong et al. (2013) propose the use of expert knowledge as a means to estimate the variogram when there are not enough observations to calculate a reliable variogram using geostatistics. They point to an increasing realisation from other disciplines that experts' knowledge provides a useful source of information that can be incorporated into statistical models. Truong et al. (2013) also suggest that expert knowledge may be useful when there is no data available or even when the available data for some reason are unreliable or unsuitable.

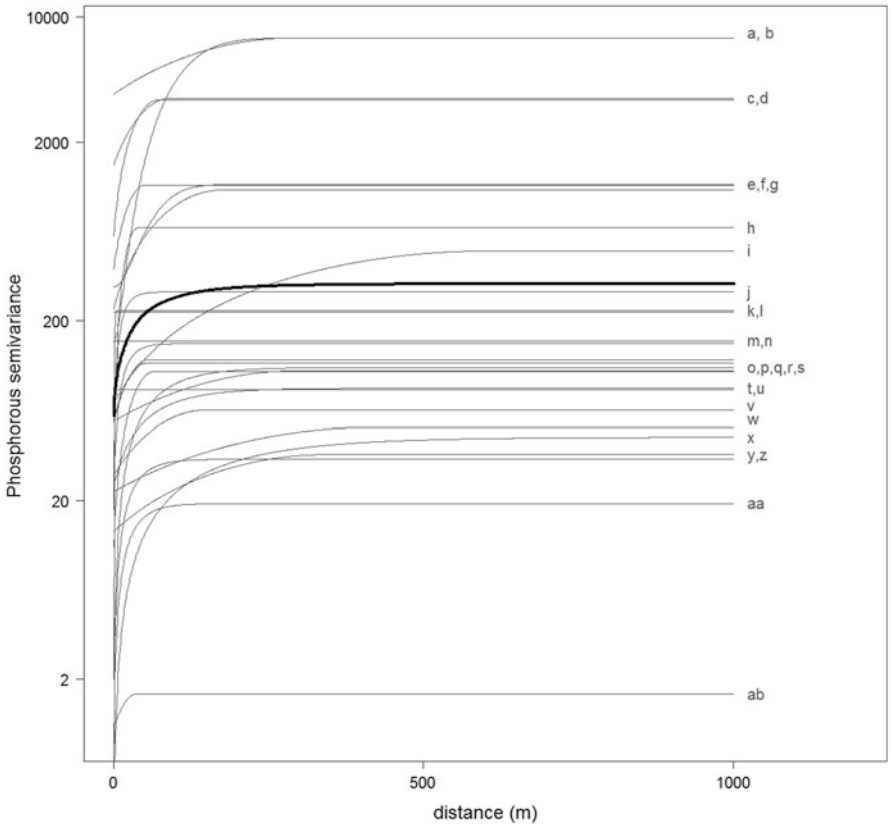


Fig. 21.8 Compilation of field scale variograms for phosphorus. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–ab*) are given in Table 21.7

Truong et al. (2013) propose a strict methodology for eliciting knowledge from experts in order to construct a variogram. Their methods are designed to avoid bias. This process is still in the prototype stage. Currently, those seeking to supplement data with expert knowledge will not be able to avoid some bias. However, in many cases, subjective expert knowledge may be the best available option.

Even when there are data available, a degree of subjectivity will be required to assess the usefulness and representativeness of these data. Where possible, it will obviously be preferred that these subjective decisions are informed by those with expertise in the area of interest (geographical or topical).

We strongly encourage the use of expert knowledge in the selection of appropriate datasets for the modelling of spatial variability. Where datasets are unavailable or deemed inappropriate, it may be necessary to rely entirely on expert knowledge to estimate the variogram. Eventually, it may be possible

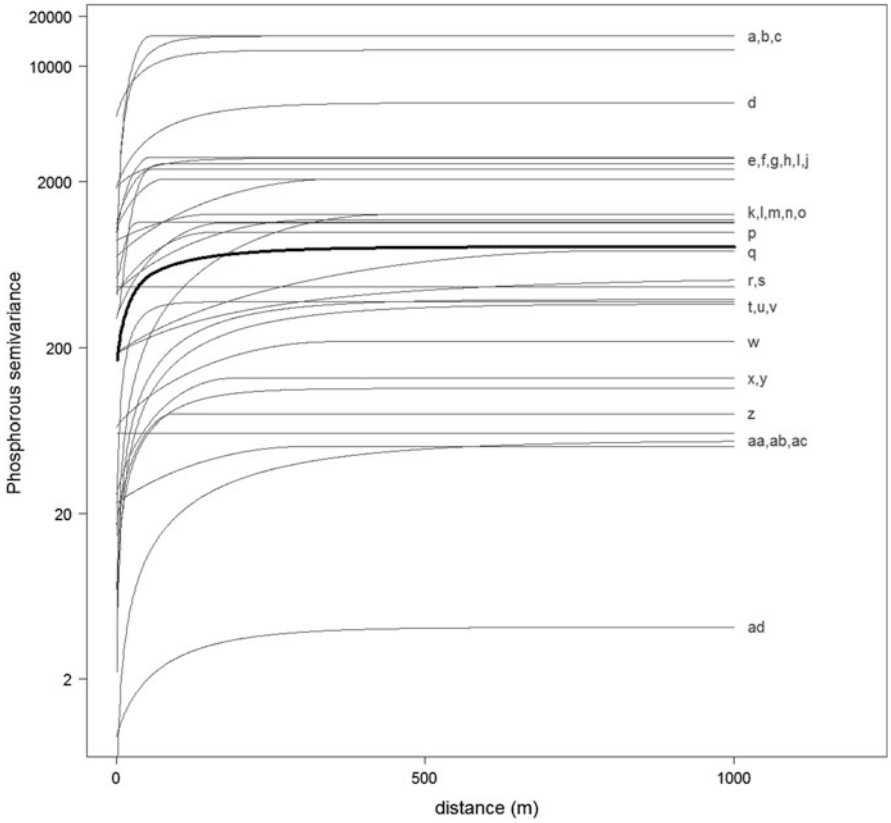


Fig. 21.9 Compilation of field scale variograms for potassium. The *bold black line* represents the average variogram. Summary details and references for each variogram (*a–ad*) are given in Table 21.8

to elicit expert knowledge using a formal process such as that described by Truong et al. (2013).

21.10 Quick Variograms

There may be situations when ancillary information, variograms from literature or even expert knowledge are unavailable or unreliable. In these cases, we would like to propose the following sampling approach that can be used to estimate a rough variogram at very low cost. The method proposed will necessarily be imprecise, but is a better alternative than not having any information. We anticipate that this method would be particularly useful in cases where alternative sources of information are available but unreliable.

Fig. 21.10 Sampling approach for estimating a 'rough' variogram. Here, it is shown that sampling is recommended in eight locations (four widely spaced points, each with a closely spaced pair, i.e. *A-B*, *C-D*, *E-F* and *G-H*)

We suggest sampling in eight locations (four widely spaced points, each with a closely spaced pair as per the diagram shown in Fig. 21.10). Obviously, the sampling design will be affected by the shape of the field. We advise that sampling occurs as close to the boundaries as possible while avoiding the edge effect.

In Fig. 21.10, one can calculate four bin sizes.

Close spacing (proxy for nugget): four pairs A-B, C-D, E-F, G-H
Maximum spacing (proxy for sill): eight pairs (A-G, A-H, B-G, B-H, C-E, C-F, D-E, D-F)
Intermediate spacing 1: eight pairs A-C, A-D, B-C, B-D, E-G, E-H, F-G, F-H
Intermediate spacing 2: eight pairs A-E, A-F, B-E, B-F, C-G, C-H, D-G, D-H

The close spacing (the close pairs) can be used as a proxy for the nugget, and the maximum spacing (the diagonals) can be used as a proxy for the total sill.

If the nugget and the sill are similar, we can assume that the appropriate model is the nugget model.

If the nugget and the sill are different, we will need to estimate the range and select a model for the variability.

There is no obvious proxy for the range that can be calculated from a small number of data points.

The two intermediate spacings may be useful to indicate where the range should occur. If they are similar to the total sill, then the range should be less than the intermediate spacings. If they are smaller than the total sill, then the range should be greater than the intermediate spacings.

We suggest that the intermediate spacings be used to determine the limits of where the range could occur. The range should then be taken as the halfway point between the limits.

For example, if the smallest intermediate bin has a variance similar to the sill, then we should set the range to be equal to the halfway point between the nugget and the smaller intermediate bin. If both of the intermediate bins have variances smaller than the sill, we should set the range to be equal to halfway between the total sill and the intermediate bin.

Modelling a range larger than the maximum separation distance is always unlikely in variograms, because of the tendency of models to break down beyond half of the field extent. We do not think that unbounded models (i.e. ranges greater than the maximum separation distance) are necessary to consider here.

Quick variograms are not able to provide a precise estimation of soil variability. Their primary and important advantage is that they are calculated from data for the property of interest in the location of interest. Quick variograms provide a useful source of information when either (i) there are no other sources of information available or (ii) for checking alternative sources of information against a reference point.

21.11 Recommendations

- Where it is economically and practically feasible, the best practice for estimating variograms is to conduct a preliminary survey and then a comprehensive field survey at the spatial scales of interest.
- Where resources for field survey are limited, it is desirable to find a variogram for the same property which has been calculated for a similar soil type. This variogram may be useful as a source of information for survey design. It may also be possible to use this variogram for kriging. Due to the large variability in soil variograms calculated for each soil property, it is critical to use discretion in this selection process.
- If more than one existing variogram from a similar soil type is identified, the information from both should be used. An averaging process may be a useful way to combine this information. Alternatively, they could be used separately to provide a range of predictions.
- Ancillary information (such as that from aerial photographs) should be considered to estimate variograms for soil properties. It can be used for survey planning and for kriging. Care must be taken to ensure that these variograms have a similar nugget-to-sill ratio to the property of interest. If available, it is preferred to use information from soil survey over ancillary information.
- If it is not possible to find a variogram for a similar area, other options are available. For example, expert knowledge could be consulted. Consultation of expert knowledge may occur in a formal process-based manner or in a more informal way.
- Quick field surveys, with limited sampling, may provide a cost-effective way to estimate rough variograms where other information is unavailable. These

methods may also be useful when it is desirable to supplement information with observations directly from the field.
- Variograms are useful for designing detailed sampling surveys for agricultural and environmental purposes.

References

Adderley WP, Jenkins DA, Sinclair EL, Stevens PA, Verinumber I (1997) The influence of soil variability on tree establishment at. Soil Use Manag 13:1–8

Amirinejad AA, Kamble K, Aggarwal P, Chakraborty D, Pradhan S, Mittal RB (2011) Assessment and mapping of spatial variation of soil physical health in a farm. Geoderma 160(3–4):292–303. doi:10.1016/j.geoderma.2010.09.021

Ayoubi S, Zamani SM, Khormali F (2007) Spatial variability of some soil properties for site specific farming in northern Iran. Int J Plant Prod 1(2):225–236

Bai YR, Wang YK (2011) Spatial variability of soil chemical properties in a jujube slope on the loess plateau of China. Soil Sci 176(10):550–558. doi:10.1097/SS.0b013e3182285cfd

Birrell SJ, Sudduth KA, Kitchen NR (1996) Nutrient mapping implications of short-range variability. In: Precision agriculture, proceedings of the third international conference, p 207

Cahn MD, Hummel JW, Brouer BH (1994) Spatial analysis of soil fertility for site-specific crop management. Soil Sci Soc Am J 3:1240–1248

Camacho-Tamayo JH, Luengas CA, Leiva FR (2008) Effect of agricultural intervention on the spatial variability of some soils chemical properties in the Eastern Plains of Colombia. Chil J Agric Res 68(1):42–55. Retrieved from <Go to ISI>://000256940500005

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Turco RF, Konopka A (1994) Field-scale variability of soil properties in central Iowa soils. Soil Sci Soc Am J 58:1501–1511

Campbell JB (1977) Spatial variation of sand content and pH within single contiguous delineations of two soil mapping units. Soil Sci Soc Am J 42(3):460–464. doi:10.2136/sssaj1978.03615995004200030017x

Chatterjee S, Santra P, Majumdar K, Ghosh D, Das I, Sanyal SK (2015) Geostatistical approach for management of soil nutrients with special emphasis on different forms of potassium considering their spatial variation in intensive cropping system of West Bengal, India. Environ Monit Assess 187(4):1–17. doi:10.1007/s10661-015-4414-9

Chung SO, Kung IK, Sung KH, Sudduth KA, Drummond ST (2008) Analysis of spatial variability in a Korean Paddy field using median polish detrending. J Biosyst Eng 33(5):362–369. doi:10.5307/JBE.2008.33.5.362

de Souza ZM, Cerri DGP, Magalhães PSG, Siqueira DS (2010) Variabilidade espacial de atributos do solo e produtividade da cultura de cana-de-açúcar em relação a localização topográfica. Rev Bras Engenharia Agric Ambient 14(12):1250–1256. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-78650798321&partnerID=tZOtx3y1

Delcourt H, Darius PL, DeBaerdemaeker J (1996) The spatial variability of some aspects of topsoil fertility in two Belgian fields. Comput Electron Agric 14(2–3):179–196. doi:10.1016/0168-1699(95)00047-X

Everett MW, Pierce FJ (1996). Variability of com yield and soil profile nitrates in relation to site-specific n management. In: Precision agriculture, proceedings of the 3rd international conference, p 43

Farooque A a, Zaman QU, Schumann AW, Madani A, Percival DC (2012) Response of wild blueberry yield to spatial variability of soil properties. Soil Sci 177(1):56–68. doi:10.1097/SS.0b013e3182376ed6

Ferraz G, Da Silva F, Carvalho LCC, Alves MDC, Franco BC (2012) Spatial and temporal variability of phosphorous, potassium and of the yield of a coffee field. Eng Agríc Jaboticabal 32:140–150. doi:10.1017/CBO9781107415324.004

Frogbrook ZL, Oliver MA, Salahi M, Ellis RH (2002) Exploring the spatial relations between cereal yield and soil chemical properties and the implications for sampling. Soil Use Manag 18:1–9. doi:10.1079/SUM200186

Ganawa S, Soom EM, Amin MM, Musa MH, Rashid A, Shariff M, Wayayok A (2003) Spatial variability of total nitrogen, and available phosphorus of large rice field in. Sci Asia 29(2003):7–12

Goovaerts P, Chiang CN (1993) Temporal persistence of spatial patterns for mineralisable nitrogen and selected soil properties. Soil Sci Soc Am J 57:372–381

Gutierrez CAG, Cortes CA, Camacho-Tamayo JH (2010) Spatial variability of some chemical. Rev UDCA Actualidad Divulg Cient 13:87–95

Kerry R, Oliver MA (2003) Variograms of ancillary data to aid sampling for soil surveys. Precis Agric 4(3):261–278. doi:10.1023/A:1024952406744

Kerry R, Oliver MA (2004) Average variograms to guide soil sampling. Int J Appl Earth Obs Geoinf 5(4):307–325. doi:10.1016/j.jag.2004.07.005

Kerry R, Oliver MA (2007) Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. Geoderma 140(4):383–396. doi:10.1016/j.geoderma.2007.04.019

Kerry R, Oliver MA (2008) Determining nugget: sill ratios of standardized variograms from aerial photographs to krige sparse soil data. Precis Agric 9(1–2):33–56. doi:10.1007/s11119-008-9058-0

Kilic K, Kilic S, Kocyigit R (2012) Assessment of spatial variability of soil properties in areas under different land use, Bulgarian Journal of Agricultural Science 18(5):722–732

Kristensen K, Simmelsgaard SE, Djurhuus JD, Olesen SE (1995) Spatial variability of soil physical and chemical parameters. In: Proceedings of the seminar on site specific farming, pp 39–55

Kumhalova J, Kumhala F, Kroulik M, Matejkova S (2011) The impact of topography on soil properties and yield and the effects of weather conditions. Precis Agric 12(6):813–830. doi:10.1007/s11119-011-9221-x

Laslett GM, McBratney AB, Pahl PJ, Hutchinson MF (1987) Comparison of several spatial prediction methods for soil pH. Eur J Soil Sci 38(2):325–341. Retrieved from http://dx.doi.org/10.1111/j.1365-2389.1987.tb02148.x

Liu G-S, Wang X-Z, Zhang Z-Y, Zhang C-H (2008) Spatial variability of soil properties in a tobacco field of Central China. Soil Sci 173(9):659–667. doi:10.1097/SS.0b013e3181847ea0

Liu G-S, Jiang H-L, Liu S-D, Wang X-Z, Shi H-Z, Yang Y-F, … Gu J-G (2010) Comparison of kriging interpolation precision with different soil sampling intervals for precision agriculture. Soil Sci, 175(8), 405–415. doi:10.1097/SS.0b013e3181ee2915

Lopez-Granados F, Jurado-Exposito M, Atenciano S, Garcia-Ferrer A, Sanchez de la Orden M, Garcia-Torres L (2002) Spatial variability of agricultural soil parameters in southern Spain. Plant Soil 1:319–326. doi:10.1023/A:1021568415380

Machado L d O, Lana ÂMQ, Lana RMQ, Guimarães EC, Ferreira CV (2007) Variabilidade espacial de atributos químicos do solo em áreas sob sistema plantio convencional. Rev Bras Ciênc Solo 31(3):591–599. doi:10.1590/S0100-06832007000300019

McBratney AB, Pringle MJ (1999) Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precis Agric 1:125–152. doi:10.1023/a:1009995404447

McBratney A, Clarke SF, Thomson HM (1985) pH, electrical conductivity and bicarbonate-extractable phosphorus values for soil specimens from a one hectare area at Myall Vale Research Station, Narrabri. CSIRO Div Soils Tech Memo 25:1985

Miller MP, Singer MJ, Nielsen DR (1988) Spatial variability of wheat yield and soil properties on complex hills. Soil Sci Soc Am J 52(4):1133. doi:10.2136/sssaj1988.03615995005200040045x

Molin JP, Di G, Faulin C (2013) Spatial and temporal variability of soil electrical conductivity related to soil moisture. Sci Agric 70:1–5

Mondo VHV, Gomes Junior FG, Pinto TLF, de Marchi JL, Montomiya AV d A, Molin JP, Cicero SM (2012) Spatial variability of soil fertility and its relationship with seed physiological potential in a soybean production area. Rev Bras Sementes 34(2):193–201. doi:10.1590/S0101-31222012000200002

Mulla DJ (1993). Mapping and managing spatial patterns in soil fertility and crop yield. Soil Specif Crop Manag 15 pp 15–26

Nanni MR, Povh FP, Demattê JAM, De Oliveira RB, Chicati ML, Cezar E (2011) Optimum size in grid soil sampling for variable rate application in site-specific management. Sci Agric 68(3):386–392. doi:10.1590/S0103-90162011000300017

Nolin MC, Guertin SP, Wang C (1996) Within field spatial variability of soil nutrients and corn yield in a montreal lowlands clay soil. In: Precision agriculture, proceedings of the 3rd international conference, pp 257–270

Nouri H, Amin MSM, Razavi SJ, Anuar AR, Aimrun W, Borujeni SC (2010) Sugar beet performance affected by uniformity of N fertigation. Am J Appl Sci 7(3):366–370

Oliver MA, Webser R (1987) The elucidation of soil pattern in the Wyre Forest of the west midlands, England I multivariate distribution. J Soil Sci 38(2):279–291. doi:10.1111/j.1365-2389.1987.tb02145.x

Oliver MA, Webster R (2014) A tutorial guide to geostatistics: computing and modelling variograms and kriging. Catena 113:56–69. doi:10.1016/j.catena.2013.09.006

Panagopoulos T, Antunes MDC (2008) Integrating geostatistics and GIS for assessment of erosion risk on low density Quercus suber woodlands of South Portugal. Arid Land Res Manag 22(2):159–177. doi:10.1080/15324980801958000

Pettitt AN, McBratney AB (1993) Spatial sampling designs for estimating variance components. J R Stat Soc Ser C (Appl Stat) 42(1):185–209

Pierce FJ, Warnke DD, Everett MW (1995). No Title. Site Specif Manag Agric Syst 133

Pribyl DW (2010) A critical review of the conventional SOC to SOM conversion factor. Geoderma 156(3–4):75–83. doi:10.1016/j.geoderma.2010.02.003

Shatar TM (1996) Site specific crop management – relationships between edaphic factors and sorghum yield. The University of Sydney, Sydney

Shouse PJ, Gerik TJ, Russell WB, Cassell DK (1990) Spatial distribution of soil particle size and aggregate stability index in a clay soil. Soil Sci. doi:10.1097/00010694-199006000-00006

Shukla MK, Slater BK, Lal R, Cepuder P (2004) Spatial variability of soil properties and potential management classification of a Chernozemic field in lower Austria. Soil Sci 169(Cv):852–860. doi:10.1097/00010694-200412000-00004

Sidorova V a, Zhukovskii EE, Lekomtsev PV, Yakushev VV (2012) Geostatistical analysis of the soil and crop parameters in a field experiment on precision agriculture. Eurasian Soil Sci 45(8):783–792. doi:10.1134/S1064229312080066

Silva VR, Reichert JM, Storck L, Feijo S (2003) Seção Iv – Fertilidade Do Solo E Nutrição De Plantas Ao Potássio Em Um Argissolo Sob. Rev Bras Ciênc Solo 5:565–571

Tabor JA, Warrick AW, Myers DE, Pennington DA (1985) Spatial variability of nitrate in irrigated cotton: II. Soil nitrate and correlated Variables1. Soil Sci Soc Am J. doi:10.2136/sssaj1985.03615995004900020024x

Thompson AN, Shaw JN, Mask PL, Touchton JT, Rickman D (2004) Soil sampling techniques for Alabama, USA grain fields. Precis Agric 5(4):345–358. doi:10.1023/B:PRAG.0000040805.01967.b0

Truong PN, Heuvelink GBM, Gosling JP (2013) Web-based tool for expert elicitation of the variogram. Comput Geosci 51:390–399. doi:10.1016/j.cageo.2012.08.010

Uehara G, Trangmar BB, Yost RS (1985) In: Nielsen DR, Bouma J (eds) Soil spatial variability. PUDOC, Wageningen

van Meirvenne M, Hofman G (1989) Spatial variability of soil nitrate nitrogen after potatoes and its change during winter. Plant Soil 120:103

Wade SD, Foster IDL, Baban SMJ (1996) The spatial variability of soil nitrates in arable and pasture landscapes: implications for the development of geographical information system models of nitrate leaching. Soil Use Manag 12 pp 95–101

Webster R, McBratney AB (1987) Mapping soil fertility at Broom's barn by simple kriging. J Sci Food Agric 38(2):97–115. doi:10.1002/jsfa.2740380203

Webster R, Oliver M (2001) Geostatistics for environmental scientists. Wiley, West Sussex

Williams RD, Ahuja LR, Naney JW, Ross JD, Barnes BB (1987) Spatial trends and variability of soil properties and crop yield in a small watershed. Trans ASAE 30(6):1653–1660

Yana J, Lee CK, Umeda M, Kosaki T (2000) Spatial variability of soil chemical properties in a paddy field spatial variability of soil chemical properties in a. Soil Sci Plant Nutr 0768(January). doi:10.1080/00380768.2000.10408800

Zanão Júnior LA, Lana RMQ, Carvalho-Zanão MP, Guimarães EC (2010) Variabilidade espacial de atributos químicos em diferentes profundidades em um Latossolo em sistema de plantio direto. Rev Ceres 57(3):429–438. doi:10.1590/S0034-737X2010000300021

Zhang X, Jiang L, Qiu X, Qiu J, Wang J, Zhu Y (2016) An improved method of delineating rectangular management zones using a semivariogram-based technique. Comput Electron Agric 121:74–83. doi:10.1016/j.compag.2015.11.016

Chapter 22
Broad-Scale Soil Monitoring Schemes

Dominique Arrouays, Ben P. Marchant, Nicolas P. A. Saby, Jeroen Meersmans, Thomas G. Orton, Manuel P. Martin, Pat H. Bellamy, R. M. Lark, and Mark Kibblewhite

"Land, then, is not merely soil; it is a fountain of energy flowing through a circuit of soils, plants and animals".

Aldo Leopold, A Sand County Almanac, 1949

22.1 Introduction

Soil resources provide many important ecosystem goods and services. However, they are at risk from a variety of threats operating over a broad range of scales. Political awareness that soil is threatened by increasing pressures has been rising for several years (European Commission 2006). Indeed, the demand for soil information is increasing continuously (Richer de Forges and Arrouays 2010). Although rates of soil degradation are often slow and only detectable over long timescales,

This chapter is a reprint from an original paper published in the journal *Pedosphere*. Changes from the original article include its formatting, the addition of cross-references to other chapters of this book and of one recent reference (Louis et al. 2014) and some minor textual changes.
Citation of the original paper: Arrouays, D., Marchant B. P., Saby, N. P. A., Meersmans, J., Orton, T. G., Martin, M. P., Bellamy, P. H., Lark, R. M. and Kibblewhite, M. 2012. Generic issues on broad-scale soil monitoring schemes: A review. *Pedosphere*. 22(4): 456–469.

D. Arrouays (✉) • N.P.A. Saby • M.P. Martin
INRA, US 1106, InfoSol Unit, CS 40001 Ardon, F-45075 Orléans Cedex 2, France
e-mail: Dominique.arrouays@inra.fr

B.P. Marchant • R.M. Lark
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK

J. Meersmans • P.H. Bellamy • M. Kibblewhite
Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

T.G. Orton
University of Queensland, Brisbane, Australia

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_22

they are often irreversible. Therefore, monitoring soil quality and condition is essential in order to detect adverse changes in their status at an early stage.

Soil monitoring is the systematic determination of soil properties so as to record their temporal and spatial variations (FAO/ECE 1994). In a recent review, Morvan et al. (2008) defined a soil monitoring network (SMN) as a set of sites/areas where changes in soil characteristics are documented through periodic assessment of an extended set of soil properties. According to this definition, national frameworks for soil monitoring exist in numerous countries and in most member states of the European Union. However, while some countries have achieved uniformity in methodology and coverage, this is far from common even among national systems (Arrouays et al. 2008a; Morvan et al. 2008; van Wesemael et al. 2011). In addition to achieving harmonisation, there are many generic issues that must be addressed by scientists when establishing and operating SMNs, including the requirement for these to be effective for different soil systems. Of particular importance is the requirement for SMNs to detect change in soil over relevant spatial and temporal scales with adequate precision and statistical power (Arrouays et al. 2008b; Desaules et al. 2010; van Wesemael et al. 2011).

In this chapter, we present some of these generic issues including the design and implementation of soil sampling in space and time, the development of statistical techniques that are general enough to describe the complicated patterns of spatial and temporal variations of soil properties and harmonisation issues.

22.2 Soil Monitoring Objectives

In a review of European SMNs, Arrouays et al. (1998) stressed that their establishment may have several objectives:

1. Determination of the current characteristics and properties of soils as well as their environmental stresses, which can be considered as an initial assessment of the soil status, often called “baseline” values, although the term “baseline” may be reserved for some assessment of soil state without the impact of human activities, inferred, perhaps, from nearby soils under climax vegetation
2. Long-term and/or early determination of changes in soils as a consequence of location-, stress- and use-specific factors, through periodic investigations
3. Assessment of the sensitivity of soils to changes and prediction of their future development;
4. Development and validation of models for the simulation of ecosystem responses and the use of these to estimate responses to actual or predicted changes and stresses and to make regional assessments in concert with survey data
5. Establishment of reference sites for calibration of environmental measurements
6. Generation of information about soil trends, to inform future national policies to protect soils from degradation and pollution, including the identification of new threats to soil quality/condition and tests of the effectiveness of existing policies

de Gruijter et al. (2006) grouped the objectives of SMNs into three broad categories that have implications when developing the options for the design of a SMN:

1. Status/ambient monitoring to characterise or quantify the status of soil and follow how its properties change over time, such as topsoil carbon content under different land uses
2. Trend/effect monitoring to assess the possible effects of pressures or drivers on soils to determine not only status but also whether a change was caused by a specific event or process
3. Regulatory/compliance monitoring to determine whether soils are failing to meet set standards or targets

22.3 General Considerations About SMN Design and Construction

The choice of design for a SMN is crucial, especially when assessing large areas and several properties that are driven by numerous controlling factors of various origins and scales.

22.3.1 Establishing a SMN

Several reviews have highlighted large differences between existing networks (Arrouays et al. 1998; Morvan et al. 2008; Saby et al. 2008b; van Wesemael et al. 2011). The geographical coverage of SMNs is very diverse between and within countries. Three broad approaches to the establishment of SMNs can be distinguished, including:

1. The design and construction of purpose-built SMNs
2. Resampling of the soil at sites where measurements have previously been made for other purposes
3. Compilation and analysis of soil data that have previously been collected in other soil analysis exercises or experiments

Purpose-built SMNs have been adopted by many countries (e.g. France, UK, Denmark, Austria, Switzerland, Germany) although in most cases the sites have, as yet, been sampled only once and hence remain inventories until sampling is repeated. The sampling design is critical when establishing new SMNs. There are continuing and extensive discussion about the choice between probability sampling, which permits design-based analyses free of any statistical model, and model-based sampling schemes, commonly regular grids with some supplementary points, which are analysed by model-based statistical methods (Brus and de Gruijter 1993, 1997;

de Gruijter et al. 2006). The probability designs include a random component in the selection of sampling locations, whereas the purposive designs select sampling locations such that a specified objective is best served (see Chap. 11). Often the purposive design consists of a regular grid since this ensures that the study region is evenly sampled. This design choice controls the types of statistical analyses that can be performed. Probability designs are generally used to answer questions about behaviour across the whole study area or within a restricted number of subareas using design-based analyses. This means that inferences from the data are based upon the probability that particular locations are included in the design. Designs other than probability designs require model-based analyses (see Chap. 11) where statistical models of the variation of the property are estimated. If these models truly reflect the variation of the property, it is possible to make localised predictions and maps and to quantify the uncertainty associated with these maps.

The decision about the scale over which results should be reported presents issues in itself. To some extent it should be controlled by the scale at which policymakers require information (see Chaps. 23 and 17). However, some effects may only be observable at particular spatial scales. For example, Wang et al. (2010) demonstrated that effects of climate on soil organic carbon (SOC), which were evident at the provincial scale, were less evident at smaller spatial scales.

Discussions are still ongoing in Europe about the effectiveness of stratified random sampling compared to purposive sampling on a grid. Previous simulations have shown that a 16×16 km grid is representative of most soil-type/land cover combinations at European and national scales (Arrouays et al. 2001; Van-Camp et al. 2004; Morvan et al. 2008). In a report about the design and implementation of a future SMN for the UK, Black et al. (2008) provided an extensive review of the advantages, limitations and relative performances of these sampling options. This study compared two purposive designs (grid and optimised grid) and two probability designs (stratified random and stratified cluster random sampling). The stratified random scheme was found to be the most suitable option for some of the specific questions being addressed, particularly in terms of the assessment of status and changes in SOC. In a review of ten national SMNs focused on SOC changes, van Wesemael et al. (2011) showed that most of these SMNs (seven out of ten) are based on stratified random sampling. Indeed, several studies dedicated to sampling schemes for SOC monitoring have pointed out that a stratified sampling design would be more efficient (Walter et al. 2003; Goidts et al. 2009b; Viaud et al. 2010; Meersmans et al. 2011). In view of these studies, there appears to be a consensus that stratified designs should be selected if the aim of the SMN is to determine the average status and change of soil properties over large regions and if the spatial patterns of factors which control the variation of all of the soil properties are known. Major soil groups and land use categories are often suitable factors for the stratification of the design.

However, grid-based surveys have the advantage of achieving good spatial coverage, with proportional representation of the regions of interest. Overall, the grid-based sampling scheme should be more flexible for incorporating unknown future requirements such as the impact of new pressures and monitoring of new

soil quality indicators and indicators for which spatial patterns are not yet known. Also, grid-based designs will in general be more appropriate if a key objective is to produce maps of status or change.

A further consideration is how the design of different phases of a SMN should relate to each other. With reference to probability sampling, de Gruijter et al. (2006) classified designs according to whether they are static and all sampling takes place at a fixed set of locations or whether the set of locations changes for each phase of the survey. Rotational designs are a compromise where only a proportion of the locations from the previous phase are resampled and new locations are selected for the remainder of the observations. de Gruijter et al. (2006) defined synchronous designs as those where multiple observations are made at the same time. There are trade-offs between these different classes of design. If locations are resampled, then the temporal variation at these sites will be well understood, but the spatial resolution of estimates can be improved if the locations change and more sites are visited. If the measurement approach is destructive or alters the soil properties at the site, then it might not be possible to revisit a particular location. Also static designs mean that any bias in the initial sample design persists throughout the life of the SMN. Static designs might be required if it is expensive to move and reinstall monitoring devices such as the lysimeters used by Brus et al. (2010). This SMN used a nonsynchronous design because the aim was to estimate the space-time means of the measured indicators. Other surveys favour synchronous designs because estimates of the indicators are required on different dates or because they lead to simple estimators (Brus and Knotters 2008). The aim of model-based surveys is often to produce a series of maps of soil indicators on different dates, and these are most easily predicted from synchronous designs (Marchant et al. 2009). However, the number of samples and the time taken to travel between them might mean it is not practical to use truly synchronous designs for national-scale SMNs; it may take more than 1 year to complete the sampling as in the National Soil Inventory of England and Wales (Bellamy et al. 2005).

Regardless of the choice between probability-based and purposive approaches, it is important to estimate, prior to implementation of the scheme, how many measurements will be sufficient to predict status and change of key soil properties with the precision required by policymakers (e.g. Black et al. 2008). The expected errors from a particular purposive sample design can only be determined if the variogram of the status and/or change of each indicator is known. The variogram (Webster and Oliver 2007) is a function which describes the variance and spatial correlation of a property (see Chaps. 10 and 21). It is the model in much model-based analysis of soil data. It is unlikely that the variograms are known exactly prior to monitoring, but approximate variograms can be estimated from previous surveys of similar indicators in similar circumstances. This approach has been used to design both probabilistic (Brus and Noij 2008) and purposive (Marchant et al. 2009) sample schemes. Often the required precision of a SMN is unclear because neither the rate of change of an indicator nor the implications of changes are known prior to sampling. In a recent study, Lark (2009) emphasised that the current status of a particular indicator and the rate of change of that indicator are different variables,

and so their variability may differ. Some plausible statistical models of change in the soil were examined, and their implications for sampling to estimate mean change in large regions were considered. These results show that taking account of knowledge of soil processes may improve the design of the SMN. Some authors recommend adapting (or calculating) the sampling time interval to make sure that the observed changes will be significantly higher than the differences that might be due to sampling and other methodological issues (Smith 2004; Bellamy et al. 2005; Saby et al. 2008b), while others (e.g. Desaules et al. 2010) argue that given these uncertainties, reducing the time interval increases the power of the scheme to observe short-term and potentially important changes in the observed trends.

Finally, it should be stressed that resources for SMN establishment and operation are always limited to some extent, and this affects the actual choice of sampling strategy and places a premium on identifying an optimal scheme taking account of the monitoring objectives and a requirement for resource efficiency. Considering this limitation, Black et al. (2008) choose to test the design of a UK SMN on SOC status and changes, as these properties are involved in processes controlling a large number of threats to soil (e.g. decline in soil organic matter, erosion, soil biodiversity, compaction, fate of contaminants). Similarly, Yu et al. (2011) assessed the sampling required to detect a change of 1.52 g kg^{-1} in SOC under various types of land management in South China. Chapter 23 gives a relevant example of designing a cost-effective monitoring scheme for farm-scale soil carbon auditing.

Resampling inventory sites from past soil mapping surveys allows immediate estimates of change and reduces the opportunity cost of establishing a SMN, as the baseline sampling exercise is already completed. This strategy has been used extensively in Belgium for monitoring SOC (Arrouays et al. 1996; Sleutel et al. 2007; Goidts et al. 2009a, b; Meersmans et al. 2009, 2011). It supports a focus on the change in SOC stock at the point scale. Although Goidts et al. (2009a) resampled within a radius of 11 m of the original site of the Belgian National Soil Survey (1947–1974), the source of error related to imprecise resampling of each location was quite large (i.e. relative RMSE ranging between 12% and 31%) due to large variability in SOC concentration, bulk density, stone content and sampling depth at very fine spatial scales (i.e. variability within the same field). Consequently, given the response time of SOC to changes in management or land use (i.e. in the order of decades), most soil inventories are probably not old enough, and/or the rates of SOC changes at individual sites are too small to be detected by resampling. Nevertheless, the latter study shows that uncertainty because of positioning error was considerably lower when studying SOC stock changes for homogeneous landscape units (characterised by same land use, agricultural region and soil type) due to the fact that multiple locations (9–47) were sampled at this aggregated level (i.e. relative RMSE ranging between 1% and 11%). Indeed, other studies have been able to detect significant temporal changes when conducting SOC stock comparisons for areas rather than individual monitoring sites. For example, Meersmans et al. (2009) studied changes in the vertical heterogeneity of SOC by resampling soil profile pits from the National Soil Survey and comparing modelled depth distribution of SOC from both time periods within homogenous land use-soil-

type combinations in North Belgium. Moreover, Sleutel et al. (2007) related average SOC stock evolution between 1990 and 2000 by municipality (from 190,000 SOC measurements) to agricultural variables (e.g. manure application, crop rotation, land use change) in order to derive the main factors explaining the overall SOC change over time in Flanders (North Belgium). Recently, Meersmans et al. (2011) identified an overall countrywide significant increase under grassland and a decrease under cropland after modelling the spatial distribution of SOC in all agricultural soils over all of Belgium in 1960 using the initial sites from the National Soil Survey and data for 629 locations resampled in 2006.

Analysing the results from existing soil measurement exercises, such as operational soil testing by farmers or fertiliser suppliers, is one potential option for detecting large temporal trends in soil characteristics. Pre-existing data, such as historic soil testing results, have often been used to assess temporal changes at national and regional levels, e.g. for phosphorus by Skinner and Todd (1998) in England and Wales, Cahoon and Ensign (2004) in eastern North Carolina (USA), Wheeler et al. (2004) in New Zealand, Lemercier et al. (2008) in France and Reijneveld et al. (2010) in the Netherlands and for carbon, Saby et al. (2008a) in France and Reijneveld et al. (2009) in the Netherlands. These studies assessed the change in soil test results with respect to land uses, cropping regimes and soil types. A spatial analysis of a soil test database performed by Baxter et al. (2006) in England and Wales contributed to designing future sampling approaches for monitoring soil properties at the national scale.

The conclusions drawn using these kinds of data may be subject to several sources of bias that are inherent in a noncontrolled sampling strategy. The farmers' agronomic concerns for soil testing may have induced skews accentuating the proportion of extreme values, especially for trace element contents. Indeed, farmers are likely to require trace element soil testing when they suspect a crop or animal deficiency or toxicity. Moreover, possible biases may arise from changes in sampling resolution in space and time.

In deciding upon the monitoring approach to be used in SMN, managers must weigh the efficiency of purpose-built designs against the reduced costs and immediacy of change estimates from the other types of designs. The benefits of the purpose-built design might be strongly felt if the SMN has a long lifetime and is to be resampled on several occasions. If soil monitoring is required to quantify an immediate threat in the short term, then the use of existing soil observations becomes more important. If the resampling of an existing inventory is being proposed, then the suitability of the inventory design for soil monitoring must be assessed. If the initial inventory was a non-probability survey, then it will not be possible to apply design-based analyses to the SMN. Model-based analyses will require that the design of the inventory is adequate to estimate a model of the spatial variation in the change of key properties and to predict this change across the study region. The data from existing soil measurement exercises should only be used if they are considered to be representative of the underlying variability of the soil.

22.3.2 Within-Site Sampling

Site area and number of subsamples. When planning sampling in a site where a soil indicator is expected to change, it is necessary to know how many samples should be taken to demonstrate a given change and after how long this change will be detectable. At the site level, numerous studies have addressed these issues (Hungate et al. 1996; Garten and Wullscheleger 1999; Conen et al. 2003, 2004; Saby and Arrouays 2004; Smith 2004). Arrouays et al. (2008a) reviewed within-site variability using data from the literature. One hundred and twenty references were collected, providing information about the short-range variability of soil indicators, for sites having areas ranging from 1 m^2 to 20 ha. The data were used to derive quantitative estimates of the mean variances, standard deviations and coefficients of variation for all available parameters. They examined the possible relationships between within-site variability and site area and/or mean values, and they found a strong relationship between the within-site variability of some parameters and the size of the site area. A marked relative increase in variability was observed for sites having areas >1 ha. This was particularly the case for some trace elements which are known to exhibit large spatial variations over quite short distances (Pb, Cd, Zn and Cu). In view of the increase in variability with site area and its implications for the number of samples that should be collected, they recommended using site areas not exceeding 1 ha to keep the number of subsamples practically feasible. If the aim of the SMN is to report the mean of an indicator over large scales such as soil-landscape units, then within-site variability is less important provided that the effect of this variability on the overall error of the mean is controlled, perhaps by forming a soil sample for analysis by aggregation of aliquots from across the site.

Due to resource constraints, most of the national monitoring sites are sampled using composite sampling, i.e. taking subsamples and bulking them. However, as has been stressed, studies of the subsampling error of monitoring sites are crucial for the interpretation of results and changes. In a study of results from the Swiss soil monitoring network (NABO), Desaules et al. (2010) showed that no certified trends can be stated after three measurement campaigns over a period of 10 years. Moreover, these authors stressed that the only way to detect reliable signals and trends earlier is to improve the overall measurement quality (precision and bias) and to shorten the sampling time interval.

Sampling depth. In their review of European SMNs, Arrouays et al. (2008a) and Morvan et al. (2008) showed that fixed-depth increments are predominantly used for core sampling (in more than 70% of the SMNs). This sampling method ensures standardisation between sites. It is also the most relevant for some anthropogenic characteristics (e.g. anthropogenic heavy metals, radionuclides, organo-chemicals) and for properties showing a strong gradient near the soil surface where the soil is often sampled over smaller increments.

Pedogenic horizons are often sampled in soil pits, outside the monitoring site, but close to it. This method of sampling is relevant for some parameters (e.g. particle-size distribution, water retention properties, mineralogy). It is also the most relevant

unit to link SMN observations to geographical soil information systems derived from soil mapping activities.

For nearly all the SMNs, the organic layers at the soil surface are sampled separately from the underlying organo-mineral soil, and this is our recommendation.

For organo-mineral layers, we recommend adoption of systematic depths in order to avoid subjectivity in sampling, harmonise sampling protocols and facilitate comparisons between SMNs.

The best practice would be to sample both by depth increments in the site and by pedogenetic horizons in soil pits, outside the monitoring area, but close to it. Arrouays et al. (2008a) examined, for each European SMN, the depths to which indicators are measured or can be calculated which were highly variable amongst SMNs.

Another way to compare vertical sampling is to calculate for each SMN the maximal depth to which sampling is realised. About 90% of the SMNs provided information down to 20 cm, whereas nearly 65% of the SMNs reached at least 30 cm.

It is very difficult to recommend sampling depths which should be adopted for all SMNs. Moreover, there may be good reasons for accepting a particular depth in a particular SMN, and changing systematic depths for a national SMN might, in some cases, make it very difficult to use the data from previous campaigns to assess changes. For example, it is not possible to compare data for indicators based on a 0–15 cm sampling depth with that for the same indicators based on a 0–30 cm resampling depth. One way to harmonise reporting at the international level could be to report the results on the basis of an equivalent mineral mass (Ellert and Bettany 1995). However, this would require the determination of bulk density at all sites and at each sampling date. General considerations about using soil depth functions or horizons and classes are given in Chap. 9.

22.3.3 Resampling the SMN

One objective when resampling should be to replicate as closely as possible the original sampling methodology and location. This requires that the original methodology is documented completely, but even when this is done, it is likely that variation in detailed procedures will occur, for example, due to differences in practice between different operators. This extends to laboratory testing as well as field sampling. While the availability of global position system (GPS), especially if this incorporates a ground station, means that the longitude and latitude for sampling locations can be precisely recorded and repeat sampling can be exactly located, this does not extend to altitude, and very often the soil surface has been altered and sometimes eroded leading to uncertainty in the equivalence of sampling exercises. Deviation from sampling and analytical protocols and location errors are likely to be confounded with those arising from actual spatial and temporal variation in the indicator being monitored. When making in situ measurements, it is possible in

principle to resample a specific location and the soil within it, but where a sample is extracted for laboratory testing, this is clearly impossible. In the latter case, it is essential to establish an adequate sampling scheme that can be applied rigorously at each sampling location, for example, by establishing a grid and removing samples from randomly determined locations at each sampling exercise.

22.4 Statistical Inference Issues

22.4.1 Design-Based or Geostatistical Methods

The variation of soil properties is very complex since soil is affected by many processes acting over different spatial and temporal scales. Local factors such as geological anomalies or anthropogenic pollution can distort and disguise the underlying relationships of interest. Therefore statistical analyses are required to test the significance of relationships and to determine the uncertainty associated with estimates and predictions. We described previously how some SMNs such as the Countryside Survey of Great Britain (Firbank et al. 2003) are based on probability sampling, whereas others such as the French National Soil Monitoring Network are based on purposive designs.

There are different statistical methods associated with these different types of designs. Design-based analyses which are reviewed by Barnett (2002) and de Gruijter et al. (2006) are associated with probability designs. These are well-established statistical techniques which can estimate summary statistics such as the mean, median or probability density function (PDF) of a soil property across the entire study region or a portion of it. They can be used to understand the underlying behaviour in the region and test hypotheses about the effect of particular factors or threats. An estimation variance is also calculated that quantifies the uncertainty associated with these estimates. Design-based methods can account for different stratifications of the data, compare different temporal phases of SMNs and determine whether a soil property has changed significantly between phases. Kravchenko and Robertson (2011) stressed the importance of performing power analyses prior to sampling to predict the sampling requirements and post sampling to determine if observed changes are significant and exactly what can be inferred from the absence of a significant change.

Soil monitoring networks based on purposive designs are generally analysed by geostatistical techniques which can be used to make local predictions and quantify uncertainties at any site of interest (see Chaps. 10, 11 and 14). Many of the geostatistical methodologies commonly used today can be directly attributed to Matheron (1965) and his analyses of the spatial variation of ore bodies. These methodologies are based on a statistical model known as the variogram which is fit to available data and describes the pattern of spatial variation of the observed variable (see Chap. 21). The fitted variogram is used in kriging (Krige 1966; Chap.

10) to predict the variable across the region and to calculate a measure of the uncertainty associated with the predictions, known as the kriging variance. These methodologies, which are introduced in an accessible manner by Webster and Oliver (2007), rely upon a number of assumptions about the statistical distribution of the soil indicator and the regularity of its variation. Brus and de Gruijter (1997) consider this a disadvantage of geostatistical methods since the inferences made from them will be invalid when these assumptions are not appropriate. In contrast, design-based methods do not fit models to the data but instead base inferences on the sample design and the probability that a point is included in the sample. The development of techniques to analyse SMNs which are sampled spatially by probability designs but temporally by non-probability designs is an active area of research (Brus et al. 2010).

A major challenge associated with geostatistical techniques, but not design-based techniques, is to ensure that the model of variation is appropriate everywhere in the study region. In the remainder of this section, we focus upon this challenge in various circumstances that might not be consistent with the standard geostatistical model.

22.4.2 Generalising the Geostatistical Model

Often the variation of a soil property is sufficiently consistent with the assumptions of standard geostatistical models for the methods of Matheron (1965) to produce adequate results. However in general the variation of soil indicators is much more complex. Therefore, since the 1960s, many methods have been proposed to generalise the geostatistical model so that, for example, the expectation (Lark et al. 2006b) or variance (Marchant et al. 2009) of the indicator can vary according to covariates such as soil type. Furthermore, in some situations, the kriging variance does not give a sufficiently complete description of the uncertainty associated with the SMN, and model-based geostatistical methods have been introduced to predict the entire PDF of the soil indicator at each site (see Chap. 11). Then the PDF can be interrogated to answer specific questions such as "What is the probability that the concentration of the soil indicator exceeds the regulatory threshold?" or "What is the probability that the concentration of the soil indicator has increased?"

The geostatistical analysis of national-scale SMNs can be particularly challenging. The vast area covered by these surveys means that the observed variation is the combined effect of processes acting over disparate spatial scales. The number of observations means that efficient computational methods are required to ensure that the statistical analysis is tractable.

22.4.3 Extreme Observations

Isolated geological anomalies or pollution can lead to outliers or extreme values amongst SMN observations. Outliers are not consistent with standard geostatistical models and can lead to the underlying uncertainty in the SMN being overestimated. Standard kriging methods can exaggerate the spatial extent of hotspots around outliers. In studies of trace elements, this can mean that excessive remediation is conducted or the areas of potential deficiencies are missed. This issue was addressed by Marchant et al. (2010) in a study of cadmium variation across France. They used robust geostatistical methods which reduced the influence of outliers when they fitted their models of variation. These models were used to identify outliers which were censored prior to kriging (Hawkins and Cressie 1984). Their methodology separated variation into geological, diffuse and anomalous components and meant that underlying relationships could be investigated. When these methods were applied to a wider group of trace elements (Saby et al. 2011), soil experts were able to identify processes contributing to variation at each scale. Figure 22.1 shows how the variation of lead across France is divided between these scales.

Robust methodologies are not appropriate for compliance monitoring where the risk of extreme values must be included and a model that can accommodate them is required. Marchant et al. (2011a) demonstrated that copula-based methods can accommodate general statistical distributions including the extreme value distribution. The PDF of the indicator of interest can be calculated at any site in the study region, and any relevant measure, such as the probability of exceeding a threshold, can be determined. They used this model to map the probability of cadmium exceeding a regulatory threshold of 0.8 mg kg^{-1} within France (Fig. 22.2).

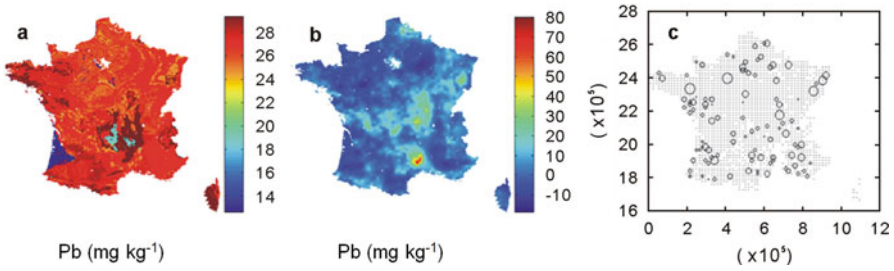


Fig. 22.1 Components of lead variation at the geological (**a**), diffuse (**b**) and anomalous localised (**c**) spatial scales estimated by robust geostatistical methods (Figure reprinted from Arrouays D, Marchant BP, Saby NPA, Meersmans J, Orton TG, Martin MP, Bellamy PH, Lark RM, Kibblewhite M (2012) Generic issues on broad-scale soil monitoring schemes: A review. Pedosphere 22(4):456–469)

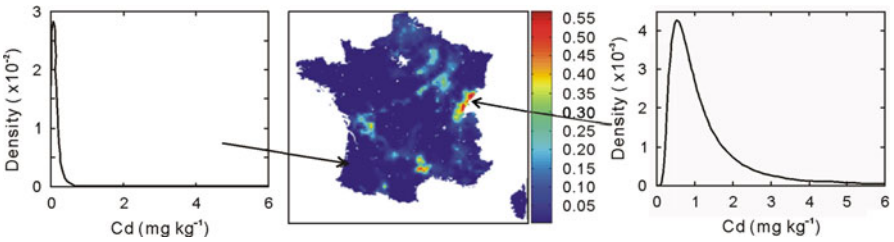


Fig. 22.2 Map of probability that regulatory cadmium threshold of 0.8 mg kg^{-1} is exceeded and (inset) PDF (probability density function) for cadmium at specified sites. Predictions are derived from a copula-based model (Figure reprinted from Arrouays D, Marchant BP, Saby NPA, Meersmans J, Orton TG, Martin MP, Bellamy PH, Lark RM, Kibblewhite M (2012) Generic issues on broad-scale soil monitoring schemes: A review. Pedosphere 22(4):456–469)

22.4.4 Different Sources of Uncertainty

Geostatistical methods can be used to quantify the uncertainty that results from the prediction of spatial maps based on observed data. However, the data obtained from a SMN includes other sources of uncertainty (see Chap. 14). In the field, there may be errors in locating observation sites. In the laboratory, there may be measurement error; for some trace elements, many observed values might be less than the detection limit, meaning that the value cannot be distinguished from zero. Our discussion focuses on continuous data such as concentrations of elements in the soil, but noncontinuous types of data such as radioactive emission counts require that the uncertainty is expressed in a different manner. Also, there can be errors in estimating spatial models. All of these components of uncertainty should be understood if we are to fully appreciate the total uncertainty of a predicted map (see Sect. 14.4.5). Rawlins et al. (2009) considered the relative influence of errors from different sources and strategies that do exist to isolate these different uncertainties. In large-scale SMNs that include many observations, the effects of these uncertainties might well be negligible. However, it is prudent to confirm that this is the case.

22.4.5 Location Uncertainty

Cressie and Kornak (2003) reviewed methods that account for location errors and suggested novel kriging equations which included such errors. Area-to-point kriging (Kyriakidis 2004) can be used to incorporate the uncertainty that results from data that are averaged over geographical units of varying sizes. The method is based on the assumption that the covariance between any two areal data units is the average of point-to-point covariances between the two units; this assumption allows a point-

to-point covariance function to be fitted to represent the variation of the areal data, which can then be used to calculate the area-to-point kriging predictions.

22.4.6 Measurement Error and Detection Limit Data

Laboratory errors can be estimated if repeated measurements are made at a small number of sites in a survey (Marchant et al. 2011b). Orton et al. (2009) used information from such repeated measurements to define a simple Gaussian measurement error model, and this was combined with the effects of the micro-scale field variation (which was assumed to be Gaussian on a log scale) using a Bayesian hierarchical modelling approach (Banerjee et al. 2004).

When laboratory measurements are reported as being less than a detection limit (DL), it is important to consider how they should be included in a spatial analysis. Commonly, measurements below the DL are incorporated in the analysis by replacing them with some function of the limit (e.g. DL/2). Although this approach is simple and allows analysis through the standard variogram estimation and kriging methods, Helsel (2006) observed that it can lead to biased estimates of the mean and variance. De Oliveira (2005) and Fridley and Dixon (2007) used data augmentation in the Bayesian framework to incorporate DL data in geostatistical prediction, in which the DL data were replaced by sampled values below the DL using a Markov chain Monte Carlo method, and their uncertainty and effect on variogram estimation and prediction were determined.

22.4.7 Other Forms of Data

Noncontinuous soil indicators, such as emission counts of radioactive material from the soil, or the presence or absence of some bacteria can be observed in SMNs. In such cases, interest will typically lie in a nonmeasurable quantity: the underlying true quantity of the radioactive contaminant in the soil or the probability of the presence of bacteria at each location. Although we can proceed with analysis as if the measured quantity were our primary focus (e.g. by indicator kriging for binary data), uncertainty in such cases can be more appropriately described by some statistical description of the data-generating mechanism: count-type data can often be described well by a Poisson distribution, and binary data might be better described by the binomial distribution. For describing uncertainty resulting from such data, the generalised linear mixed model (LMM, Diggle and Ribeiro 2007) and Bayesian hierarchical modelling approaches (Banerjee et al. 2004) provide powerful expansions of the classical kriging methods (see Sects. 11.3 and 11.4); they provide more flexible statistical representations of the data than the classical approaches, so that the processes that gave rise to the measurements can be more

accurately modelled. These general methods seem to offer significant opportunities for representing uncertainty in geostatistical analyses of these types of SMN data.

22.4.8 Uncertainty in Estimating Spatial Models

Parameters in spatial models can usually be separated into two sets: those that represent the expected value or a trend for the primary variable and those that represent its variance and correlation in space/time. Since Matheron's work (Matheron 1965), the uncertainty about trend parameters has been accounted for through the ordinary or universal kriging methodologies. For the variance parameters, the kriging methods have adopted a plug-in approach: first, the parameters are estimated, and then the estimated values plugged into the kriging equations to calculate the prediction and associated kriging variance. Hence the uncertainty of the estimated spatial model is ignored. Marchant and Lark (2004) used the Fisher information matrix to further include variance parameter uncertainty in the resulting spatial predictions in a maximum likelihood framework. Bayesian methods also incorporate fully the effects of variance parameter uncertainty through Markov chain Monte Carlo methods (Minasny et al. 2011) (see Sect. 14.4.2).

22.4.9 Inclusion of Temporal Variation

After more than one phase of a SMN has been completed, the model of variation must be modified to quantify temporal variation in addition to spatial variation. Then kriging algorithms can be used to map the change in indicators across the study region. Different spatio-temporal models have been applied in existing monitoring surveys. De Cesare et al. (2001) reviewed the use of space-time covariance models. Papritz and Flühler (1994) suggested that different phases of a survey can be treated as coregionalised variables, and Lark et al. (2006a) used robustly estimated coregionalisation models to determine the sampling requirements for mapping change in metals in the part of eastern England. Bellamy et al. (2005) included the rate of change of SOC as a parameter in their model of variation. The challenge is to determine the most appropriate model for a particular SMN. It is important to validate the model once it has been fitted so deficiencies can be identified and strategies introduced to rectify them. For example, in a monitoring survey of phosphorus enrichment in the Florida Everglades, Marchant et al. (2009) identified that phosphorus was more variable in parts of the study region adjacent to pumping stations which input agricultural runoff. They used remotely sensed data of dominant vegetation to automatically identify these regions and adjusted their model to accommodate the larger variability.

22.5 Laboratory Testing Methods

The question of which soil measurement methods to recommend is complex. Most countries have long-established SMNs and use specific testing methods. Changing these methods to a different one would impede data comparison with previous results, unless parallel analyses are performed, using both national and new reference methods. An important example relates to the assessment of global SOC stocks or changes. The analytical methods may often be different in space or time. For example, modern analytical methods such as dry combustion might be used instead of the more common Walkley and Black method (Meersmans et al. 2009). Therefore correction factors are needed to avoid methodological bias when comparing SOC data from sampling campaigns using different analytical procedures (e.g. Jolivet et al. 1998; Lettens et al. 2007; Meersmans et al. 2009).

Arrouays et al. (2008a) reported information on soil testing techniques gathered by partners from all the European member states. They found that, in some cases, the applied test procedures were not sufficiently detailed; the information provided was often very vague, even after several requests, with partners reporting only the type of extract or equipment used. Nevertheless, for SMNs for which this information was available, the testing methods showed numerous differences, indicating that the use of international standards (when they exist) is far from common. Indeed, as numerous international standards for soil analysis are still lacking, standardisation will be one of the main issues in setting up a SMN at an international level. Clearly, there is a widespread need for agreeing testing methods and ensuring that these are validated and conducted to produce data of known and documented quality.

As a minimum, for each testing method employed, the following is essential: a fully documented procedure with details of calibration methods that ensure traceability to international standards; data on the repeatability (within-batch error) and reproducibility (between-batch error) of the method, based on repeated analysis of standard samples (preferably certified reference materials); and a testing method detection limit, based on an agreed multiple of the standard error for whole procedure blanks. In addition, to support continuing quality assurance procedures, repeated analysis is required of standard samples included within each batch and analysis of the results using statistical process control charts. Participation in inter-laboratory proficiency exercises is critical. Although using a single laboratory to test all samples ensures consistency in the quality of results, it does not guarantee adequate quality, and in this case, as when several laboratories are participating in testing, it is imperative that inter-laboratory comparisons are conducted to support and demonstrate sufficient inter-comparability between laboratories.

Except for those parameters for which a consensus exists, the question of testing method harmonisation remains a very difficult issue. For several parameters, combining several techniques, on all samples or on a subset of samples, from archives can be a useful option to harmonise data obtained using different or inadequately validated testing methods. This can allow samplestaken in previous

campaigns to be used to detect changes and to establish pedotransfer functions for prediction of non-measured indicators (e.g. estimation of bulk density based on texture and organic matter content).

Generally, the main cost in soil monitoring is from sampling in the field, and adding more indicator measurements has a relatively low opportunity cost depending on the complexity of the testing procedure.

22.6 Archiving Samples

Soils are monitored through long-term networks that require long-term commitment from researchers and from funding agencies. In numerous countries, soils are monitored on the basis of national schemes. Despite these enormous efforts to characterise soils, it is striking that in the European Union, for instance, about 40% of the monitoring programmes do not archive soil samples that are collected and analysed (Arrouays et al. 2008a; Morvan et al. 2008). However there are very good reasons to retain samples.

We do not know what we will be interested in the future. When the Broadbalk experiment was established in 1843 at Rothamsted, UK, researchers were certainly not aware that their decision to carefully archive samples taken from the experimental plots would enable monitoring a posteriori of the levels of polychlorinated biphenyls in the environment (Alcock et al. 1993).

New analytical techniques will arise in the future. These will be more precise and/or will allow the use of new tracers of environmental or biogeochemical processes. A number of substances which cannot be detected using current testing methods will become quantifiable. Progresses in microbiology and molecular tools already enabled soil DNA libraries to be built to explore the micro-biodiversity and its long-term changes in relation to global change or other pressures (Dequiedt et al. 2009, 2011; Gardi et al. 2009; Ranjard et al. 2010; Bru et al. 2011). Techniques and standards for soil analyses are evolving continuously. Thus it is good practice to retain soil samples so that they can be tested in the future. However, archiving samples raises some scientific and technical issues concerning the effects of changes in sample properties with time: effect of drying (temperature) and ageing on numerous soil properties, e.g. volatilisation of persistent organic pollutants (Garmouma and Poissant 2004), changes in trace element speciation and bioaccessibility (Martens and Suarez 1997; Martinez et al. 2003; Furman et al. 2007), changes in pH (Prodromou and Pavlatou-Ve 1998), changes in phosphorus solubility (Styles and Coxon 2006) and changes in microbial communities identification (e.g. Clark and Hirsch 2008; Tzeneva et al. 2009).

22.7 Harmonisation Issues

The uniformity of methodologies and the scope of actual monitoring networks are variable between national systems. A recent review identified the differences between existing systems and described options for harmonising soil monitoring in the member states and some neighbouring countries of the European Union (Morvan et al. 2008). The present geographical coverage is uneven between and within countries (Morvan et al. 2008; Saby et al. 2008b). The most serious barrier identified, which limits the harmonisation of data from existing SMNs, is the wide variety of soil testing methods that have been employed historically. Harmonisation can be defined as the minimisation of systematic differences between different sources of environmental measures (Keune et al. 1991). There are some opportunities for harmonising data obtained using different testing methods, for example, by regression analysis, but these are limited. Recently, Baume et al. (2011) proposed a universal kriging approach that is able to deal with the issue of merging data from different monitoring networks. The establishment of benchmark sites devoted to harmonisation and inter-calibration has been advocated as a technical solution by many authors (e.g. Theocharopoulos et al. 2001; Wagner et al. 2001; Van-Camp et al. 2004; Kibblewhite et al. 2008; Morvan et al. 2008; Gardi et al. 2009). Cathcart et al. (2008) have recently set up 43 benchmark sites in Alberta, Canada, to monitor agricultural soil quality, and the sites were selected to be representative of a number of soil chemical and physical properties across the region. *However, at present, few studies have addressed crucial scientific issues such as how many calibration sites are necessary and how to choose them* (Louis et al. 2014).

22.8 Conclusions

Numerous scientific challenges arise when designing a SMN, especially when assessing large areas and several properties that are driven by numerous controlling factors of various origins and scales.

Three broad approaches to the establishment of SMNs can be distinguished, including:

1. The design and construction of purpose-built SMNs
2. Resampling of the soil at sites where measurements have previously been made for other purposes
3. Compilation and analysis of soil data that have previously been collected in other soil analysis exercises or experiments

It is essential to establish an adequate sampling protocol that can be applied rigorously at each sampling location and time. The organic layers at the soil surface should be sampled separately from the underlying organo-mineral soil, and

organo-mineral soils can be sampled both by depth increments in the site and by pedogenic horizons in soil pits, outside the monitoring area, but close to it. Different statistical methods should be associated with the different types of sampling design. Several studies propose new statistical methods that account for different sources of uncertainty (e.g. location errors, measurement error and detection limit, estimation of spatial model). Except for those parameters for which a consensus exists, the question of testing method harmonisation remains a very difficult issue. For several parameters, combining several techniques, on all samples or on a subset of samples, from archives can be a useful option to harmonise data obtained using different or inadequately validated testing methods. The establishment of benchmark sites devoted to harmonisation and inter-calibration has been advocated as a technical solution. However, no one has addressed crucial scientific issues such as how many calibration sites are necessary and how to locate them.

References

Alcock RE, Johnston AE, McGrath SP, Berrow ML, Jones KC (1993) Long-term changes in the polychlorinated biphenyl content of United Kingdom soils. Environ Sci Technol 27:1918–1923

Arrouays D, Mench M, Amans V, Gomez A (1996) Short-range variability of fallout Pb in a contaminated soil. Can J Soil Sci 76:73–81

Arrouays, D., Vogel, H., Eckelmann, W., Armstrong-Brown, S., Loveland, P., Coulter, B (1998) Soil monitoring networks in Europe. 16th World Congress of Soil Science. Montpellier, France

Arrouays D, Deslais W, Badeau V (2001) The carbon content of topsoil and its geographical distribution in France. Soil Use Manag 17:7–11

Arrouays D, Richer de Forges A, Morvan X, Saby NPA, Jones AR, Le Bas C (eds) (2008a) Environmental assessment of soil for monitoring: volume IIb: survey of national networks. Office for the Official Publications of the European Communities, Luxembourg. EUR 23490 EN/2A

Arrouays D, Morvan X, Saby NPA, Richer de Forges A, Le Bas C, Bellamy PH, Berényi Üveges J, Freudenschuß A, Jones RJA, Kibblewhite MG, Simota C, Verdoodt A, Verheijen FGA (eds) (2008b) Environmental assessment of soil for monitoring: volume IIa: inventory & monitoring. Office for the Official Publications of the European Communities, Luxembourg. EUR 23490 EN/2B

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modelling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton

Barnett V (2002) Sample survey: principles and methods, 3rd edn. Arnold, London

Baume O, Skøien JO, Heuvelink GBM, Pebesma EJ, Melles SJ (2011) A geostatistical approach to data harmonization – application to radioactivity exposure data. Int J Appl Earth Obs Geoinform 13:409–419

Baxter SJ, Oliver MA, Archer JR (2006) The representative soil sampling scheme of England and Wales: the spatial variation of topsoil nutrient status and pH between 1971 and 2001. Soil Use Manag 22:383–392

Bellamy PH, Loveland PJ, Bradley RI, Lark RM, Kirk GJD (2005) Carbon losses from all soils across England and Wales 1978-2003. Nature 437:245–248

Black H, Booth P, Bellamy P, Elston D, Frogbrook ZL, Hudson G, Lark RM, Marchant B, Potts J, Reynolds B, Lilly A (2008) Design and operation of a UK soil monitoring network. Environment Agency, Bristol. 060072/SR

Bru D, Ramette A, Saby NPA, Dequiedt S, Ranjard L, Jolivet C, Arrouays D, Philippot L (2011) Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. ISME J 5:532–542

Brus DJ, de Gruijter JJ (1993) Design-based versus model-based estimates of spatial means: theory and application in environmental soil science. Environmetrics 4:123–152

Brus DJ, de Gruijter JJ (1997) Random sampling or geostatistical modeling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 60: 1–44

Brus, DJ, Knotters M (2008) Sampling design for compliance monitoring of surface water quality: a case study in a polder area. Water Resour Res 44: W11410, 10 pp. 10.1029/2007WR006123

Brus DJ, Noij IGAM (2008) Designing sampling schemes for effect monitoring of nutrient leaching from agricultural soils. Eur J Soil Sci 59(2):292–303

Brus DJ, de Gruijter JJ, de Vries W (2010) A sampling strategy for estimating plot average annual fluxes of chemical elements from forest soils. Geoderma 159:399–408

Cahoon LB, Ensign SH (2004) Spatial and temporal variability in excessive soil phosphorus levels in eastern North Carolina. Nutr Cycl Agroecosys 69:111–125

Cathcart J, Cannon K, Heinz J (2008) Selection and establishment of Alberta agricultural soil quality benchmark sites. Can J Soil Sci 88:399–408

Clark IM, Hirsch PR (2008) Survival of bacterial DNA and culturable bacteria in archived soils from the Rothamsted Broadbalk experiment. Soil Biol Biochem 40:1090–1102

Conen F, Yakutin MV, Sambuu AD (2003) Potential for detecting changes in soil organic carbon concentrations resulting from climate change. Glob Chang Biol 9:1515–1520

Conen F, Zerva A, Arrouays D, Jolivet C, Jarvis PG, Grace J, Mencuccini M (2004) The carbon balance of forest soils: detectability of changes in soil carbon stocks in temperate and boreal forests. In: Griffiths H, Jarvis PJ (eds) The carbon balance of forest biomes. BIOS Scientific Press, Oxford

Cressie N, Kornak J (2003) Spatial statistics in the presence of location error with an application to remote sensing of the environment. Stat Sci 18:436–456

De Cesare L, Myers DE, Posa D (2001) Estimating and modelling space – time correlation structures. Stat Probabil Lett 51:9–14

de Gruijter J, Brus DJ, Bierkens MFP, Knotters M (2006) Sampling for natural resource monitoring. Springer-Verlag, Heidelberg

De Oliveira V (2005) Bayesian inference and prediction of Gaussian random fields based on censored data. J Comput Graph Stat 14:95–115

Dequiedt S, Thioulouse J, Jolivet C, Saby NPA, Lelievre M, Maron P-A, Martin MP, Chemidlin Prévost-Bouré N, Toutain B, Arrouays D, Lemanceau P, Ranjard L (2009) Biogeographical patterns of soil bacterial communities. Environ Microbiol Rep 1:251–255

Dequiedt S, Saby NPA, Lelievre M, Jolivet C, Thioulouse J, Toutain B, Arrouays D, Bispo A, Lemanceau P, Ranjard L (2011) Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. Glob Ecol Biogeogr 20(4):641–652

Desaules A, Ammann S, Schwab P (2010) Advances in long-term soil-pollution monitoring of Switzerland. J Plant Nutr Soil Sci 173:525–535

Diggle, P. J., Ribeiro, P. J. (2007) Model-based geostatistics. Springer, New York, 230pp

Ellert BH, Bettany JR (1995) Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Can J Soil Sci 75:529–538

European Commission (2006) Thematic strategy for soil Protection (Communication). COM(2006)231 final. European Commission, Brussel

FAO/ECE (1994) International workshop on harmonisation of soil conservation monitoring systems. Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Budapest

Firbank LG, Barr CJ, Bunce RGH, Furse MT, Haines-Young RH, Hornung M, Howard DC, Sheail J, Sier ARJ, Smart SM (2003) Assessing stock and change in land cover and biodiversity in GB: an introduction to the countryside survey 2000. J Environ Manag 67:207–218

Fridley BL, Dixon P (2007) Data augmentation for a Bayesian spatial model involving censored observations. Environmetrics 18:107–123
Furman O, Strawn DG, McGeehan S (2007) Sample drying effects on lead bioaccessibility in reduced soil. J Environ Qual 36:899–903
Gardi C, Montanarella L, Arrouays D, Bispo A, Lemanceau P, Jolivet C, Mulder C, Ranjard L, Römbke J, Rutgers M, Menta C (2009) Soil biodiversity monitoring in Europe: ongoing activities and challenges. Eur J Soil Sci 60:807–819
Garmouma M, Poissant L (2004) Occurrence, temperature and seasonal trends of α- and γ-HCH in air (Québec, Canada). Atmos Environ 38:369–382
Garten CT, Wullscheleger SD (1999) Soil carbon inventories under a bioenergy crop (switchgrass): measurement limitations. J Environ Qual 28:1359–1365
Goidts E, van Wesemael B, Crucifix M (2009a) Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. Eur J Soil Sci 60:723–739
Goidts E, van Wesemael B, Van Oost K (2009b) Driving forces of soil organic carbon evolution at the landscape and regional scale using data from a stratified soil monitoring. Glob Chang Biol 15:2981–3000
Hawkins DM, Cressie N (1984) Robust Kriging – a PRoposal. J Int Ass Math Geol 16:3–18
Helsel DR (2006) Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. Chemosphere 65:2434–2439
Hungate BA, Jackson RB, Field CB, Chapin FS (1996) Detecting changes in soil carbon in CO_2 enrichment experiments. Plant Soil 187:135–145
Jolivet C, Arrouays D, Bernoux M (1998) Comparison between analytical methods for organic carbon and organic matter determination in sandy spodosols of France. Commun Soil Sci Plant Anal 29:2227–2233
Keune H, Murray AB, Benking H (1991) Harmonization of environmental measurement. GeoJournal 23:249–255
Kibblewhite MG, Jones RJA, Montanarella L, Baritz R, Huber S, Arrouays D, Micheli E, Stephens M (eds) (2008) Environmental assessment of soil for monitoring: volume VI: soil monitoring system for Europe. Office for Official Publications of the European Communities, Luxembourg. EUR 23490 EN/6
Kravchenko AN, Robertson GP (2011) Whole-profile soil carbon stocks: the danger of assuming too much from analyses of too little. Soil Sci Soc Am J 75:235–240
Krige DG (1966) Two-dimensional weighted moving average trend surfaces for ore-evaluation. J South Afr Inst Min Metall 66:13–38
Kyriakidis PC (2004) A geostatistical framework for area-to-point spatial interpolation. Geogr Anal 36:259–289
Lark RM (2009) Estimating the regional mean status and change of soil properties: two distinct objectives for soil survey. Eur J Soil Sci 60:748–756
Lark RM, Bellamy PH, Rawlins BG (2006a) Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. Geoderma 133:363–379
Lark RM, Cullis BR, Welham SJ (2006b) On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Eur J Soil Sci 57:787–799
Lemercier B, Gaudin L, Walter C, Aurousseau P, Arrouays D, Schvartz C, Saby NPA, Follain S, Abrassart J (2008) Soil phosphorus monitoring at the regional level by means of a soil test database. Soil Use Manag 24:131–138
Lettens S, De Vos B, Quataert P, Van Wesemael B, Muys B, Van Orshoven J (2007) Variable carbon recovery of Walkley-black analysis and implications for national soil organic carbon accounting. Eur J Soil Sci 58:1244–1253
Louis BP, Saby NPA, Orton TG, Lacarce E, Boulonne L, Jolivet CC, Ratié C, Arrouays D (2014) Statistical sampling design impact on predictive quality of harmonization functions between soil monitoring networks. Geoderma 213:133–143
Marchant BP, Lark RM (2004) Estimating Variogram uncertainty. Math Geol 36:867–898

Marchant BP, Newman S, Corstanje R, Reddy KR, Osborne TZ, Lark RM (2009) Spatial monitoring of a non-stationary soil property: phosphorus in a Florida water conservation area. Eur J Soil Sci 60:757–769

Marchant BP, Saby NPA, Lark RM, Bellamy PH, Jolivet CC, Arrouays D (2010) Robust analysis of soil properties at the national scale: cadmium content of French soils. Eur J Soil Sci 61:144–152

Marchant BP, Saby NPA, Jolivet CC, Arrouays D, Lark RM (2011a) Spatial prediction of soil properties with copulas. Geoderma 162(3–4):327–334

Marchant BP, Tye AM, Rawlins BG (2011b) The assessment of point-source and diffuse soil metal pollution using robust geostatistical methods: a case study in Swansea (Wales, UK). Eur J Soil Sci 62:346–358

Martens DA, Suarez DL (1997) Selenium speciation of marine shales, alluvial soils, and evaporation basin soils of California. J Environ Qual 26:424–432

Martinez CE, Jacobson AR, McBride MB (2003) Aging and temperature effects on DOC and elemental release from a metal contaminated soil. Environ Pollut 122:135–143

Matheron G (1965) Les Variables Régionalisées et Leur Estimation: Une Application de la Théorie des Fonctions Aléatoires aux Sciences de la Nature (in French). Masson et Cie, Paris

Meersmans J, Van Wesemael B, De Ridder F, Dotti MF, De Baets S, Van Molle M (2009) Changes in organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. Glob Chang Biol 15:2739–2750

Meersmans J, van Wesemael B, Goidts E, van Molle M, De Baets S, De Ridder F (2011) Spatial analysis of soil organic carbon evolution in Belgian croplands and grasslands. 1960–2006. Glob Chang Biol 17:466–479

Minasny B, Vrugt JA, McBratney AB (2011) Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation. Geoderma 163:150–162

Morvan X, Saby NPA, Arrouays D, Le Bas C, Jones RJA, Verheijen FGA, Bellamy PH, Stephens M, Kibblewhite MG (2008) Soil monitoring in Europe: a review of existing systems and requirements for harmonisation. Sci Total Environ 391:1–12

Orton TG, Rawlins BG, Lark RM (2009) Using measurements close to a detection limit in a geostatistical case study to predict selenium concentration in topsoil. Geoderma 152:269–282

Papritz A, Flühler H (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. Geoderma 62:29–43

Prodromou KP, Pavlatou-Ve AS (1998) Changes in soil pH due to the storage of soils. Soil Use Manag 14:182–183

Ranjard L, Dequiedt S, Jolivet C, Saby NPA, Thioulouse J, Harmand J, Loisel P, Rapaport A, Fall S, Simonet P, Joffre R, Bouré NC-P, Maron P-A, Mougel C, Martin MP, Toutain B, Arrouays D, Lemanceau P (2010) Biogeography of soil microbial communities: a review and a description of the ongoing french national initiative. Agron Sustain Dev 30:359–365

Rawlins BG, Scheib AJ, Lark RM, Lister TR (2009) Sampling and analytical plus subsampling variance components for five soil indicators observed at regional scale. Eur J Soil Sci 60:740–747

Reijneveld A, van Wensem J, Oenema O (2009) Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. Geoderma 152:231–238

Reijneveld JA, Ehlert PAI, Termorshuizen AJ, Oenema O (2010) Changes in the soil phosphorus status of agricultural land in the Netherlands during the 20th century. Soil Use Manag 26:399–411

Richer de Forges AC, Arrouays D (2010) Analysis of requests for information and data from a national soil data centre in France. Soil Use Manag 26:374–378

Saby N, Arrouays D (2004) Simulation of the use of a soil-monitoring network to verify carbon sequestration in soils: will changes in organic carbon stocks be detectable? Commun Soil Sci Plant Anal 35:2379–2396

Saby NPA, Arrouays D, Antoni V, Lemercier B, Follain S, Walter C, Schvartz C (2008a) Changes in soil organic carbon in a mountainous French region, 1990–2004. Soil Use Manag 24: 254–262

Saby NPA, Bellamy PH, Morvan X, Arrouays D, Jones RJA, Verheijen FGA, Kibblewhite MG, Verdoodt A, Berényiüveges J, Freudenschuß A, Simota C (2008b) Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? Glob Chang Biol 14:2432–2442

Saby NPA, Marchant BP, Lark RM, Jolivet CC, Arrouays D (2011) Robust geostatistical prediction of trace elements across France. Geoderma 162(3–4):303–311

Skinner RJ, Todd AD (1998) Twenty-five years of monitoring pH and nutrient status of soils in England and Wales. Soil Use Manag 14:162–169

Sleutel S, De Neve S, Hofmas G (2007) Assessing causes of recent organic carbon losses from cropland soils by means of regional-scales input balances for the case of Flanders (Belgium). Nutr Cycl Agroecosyst 78:265–278

Smith P (2004) How long before a change in soil organic carbon can be detected? Glob Chang Biol 10:1878–1883

Styles D, Coxon C (2006) Laboratory drying of organic-matter rich soils: phosphorus solubility effects, influence of soil characteristics, and consequences for environmental interpretation. Geoderma 136:120–135

Theocharopoulos SP, Wagner G, Sprengart J, Mohr M-E, Desaules A, Muntau H, Christou M, Quevauviller P (2001) Comparative soil sampling in the Dornach site (Switzerland) for soil three-dimensional pollution description. Sci Total Environ 264:63–72

Tzeneva VA, Salles JF, Naumova N, de Vos WM, Kuikman PJ, Dolfing J, Smidt H (2009) Effect of soil sample preservation, compared to the effect of other environmental variables, on bacterial and eukaryotic diversity. Res Microbiol 160:89–98

Van-Camp L, Bujarrabal B, Gentile AR, Jones RJA, Montanarella L, Olazabal C, Selvaradjou S-K (eds) (2004) Reports of the technical working groups established under the thematic strategy for soil protection. EUR 21319 EN/5. European Communities, Luxembourg

van Wesemael B, Paustian K, Andrén O, Cerri CEP, Dodd M, Etchevers J, Goidts E, Grace P, Kätterer T, McConkey BG, Ogle S, Pan GX, Siebner C (2011) How can soil monitoring networks be used to improve predictions of organic carbon pool dynamics and CO_2 fluxes in agricultural soils? Plant Soil 338:247–259

Viaud V, Angers DA, Walter C (2010) Toward landscape-scale modeling of soil organic matter dynamics in agroecosystems. Soil Sci Soc Am J 74:1847–1860

Wagner G, Desaules A, Muntau H, Theocharopoulos S, Quevauviller P (2001) Harmonisation and quality assurance in pre-analytical steps of soil contamination studies – conclusions and recommendations of the CEEM soil project. Sci Total Environ 264:103–118

Walter C, Rossel RAV, McBratney AB (2003) Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. Soil Sci Soc Am J 67:1477–1486

Wang DD, Shi XZ, Wang HJ, Weindorf DC, Yu DS, Sun WX, Ren HY, Zhao YC (2010) Scale effect of climate and soil texture on soil organic carbon in the uplands of Northeast China. Pedosphere 20(4):525–535

Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester

Wheeler DM, Sparling GP, Roberts AHC (2004) Trends in some soil test data over a 14-year period in New Zealand. New Zeal J Agr Res 47:155–166

Yu DS, Zhang ZQ, Yang H, Shi XZ, Tan MZ, Sun WX, Wang HJ (2011) Effect of soil sampling density on detected spatial variability of soil organic carbon in a red soil region of china. Pedosphere 21:207–213

Chapter 23
Farm-Scale Soil Carbon Auditing

Jaap J. de Gruijter, Alex. B. McBratney, Budiman Minasny, Ichsani Wheeler, Brendan P. Malone, and Uta Stockmann

"You have been weighed,
you have been measured,
and you have been
found wanting".

Adhemar (A Knights Tale, 2001)

23.1 Introduction

The soil system is recognized as a significant terrestrial sink of carbon. Estimates for the top meter of soil in the world range between 1,200 and 2,500 petagrams for organic C (Batjes 1996; Lal 2004). The reliable assessment and monitoring of soil carbon stocks is of key importance for soil conservation and in mitigation strategies for increased atmospheric carbon (Stockmann et al. 2013). Carbon credits are the heart of a cap-and-trade scheme, by offering a way to quantify carbon sequestered from the atmosphere; carbon credits gain a monetary value to offset

This chapter is a reprint from an original paper published in the journal: *Geoderma*. Changes from the original article include its formatting, and the addition of cross-references to other chapters of this book and some minor textural changes.

Citation of the original paper: de Gruijter, J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B., Stockmann, U., 2016. Farm-scale soil carbon auditing. Geoderma 265, 120–130.

J.J. de Gruijter (✉)
Alterra, Wageningen University and Research Centre, PO Box 32, NL-6700 AA Wageningen, The Netherlands
e-mail: jjdegr@kpnmail.nl

A.B. McBratney • B. Minasny • I. Wheeler • B.P. Malone • U. Stockmann
Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_23

a given amount of carbon dioxide releases (Paustian et al. 2009). The agricultural industry worldwide has the capacity to capture and store carbon emissions in soil (Paustian et al. 2000). However, there is still a debate on how soil can benefit for the offsets in the carbon economy because there is no good and efficient way of measuring soil carbon storage with appropriate statistical confidence (Post et al. 2001; Smith 2004b). A scheme that can measure and monitor soil carbon storage on a farm, which is crucial to the participation of the agricultural sector in the carbon economy, is essential.

There is a win-win position for increased carbon storage in soil. Soil organic carbon (SOC) provides benefits of enhanced soil fertility through improved soil structure, by promoting the agents and mechanisms of aggregation, and increased cation exchange capacity (Stockmann et al. 2013). Studies of Australian soil systems have shown that conversion of forested and grassland areas into cultivated agriculture has led to an overall decline in SOC stock in those soils (Dalal and Chan 2001; Luo et al. 2010). Conservation tillage, reforestation, and sustainable development practices are recognized methods to promote carbon storage. One mechanism that can facilitate the effective management of the soil carbon is to treat it as a tradeable resource or commodity. A monetary value has been assigned to carbon, in all its states and forms, which can allow for the trading and offsetting of carbon budgets. The development of carbon credit markets accessible to the private sector would allow for incentives such as government payments, tax credits, and/or emission trading, which can aid in overcoming farmer reluctance to adopting management strategies that increase soil carbon (Rosenberg and Izaurralde 2001).

There are two distinct approaches recognized to establish SOC stock with Tier 3 method (IPCC, 2006) including process-based models and inventory measurement systems. The choice between each approach depends largely on applicability to the situation, data availability, and cost-effectiveness. When considering the costs and low sequestration rates, process-based models may be favored (Conant and Paustian 2002; Smith 2004b); however, it is also challenging considering the diverse combinations of climate, soil type, and managements (Rabotyagov 2010). It is inevitable that not all combinations will be covered or parameterized and support for emerging managements will have a temporal lag in incorporation as data over time is required. Added to this, there are several other reasons to also develop Tier 3 direct measurement methods including:

(1) providing an independent verification tool applicable to emerging managements at the farm scale; (2) encompassing adaptive land management through independence from established management assumptions; (3) provision of site-specific feedback to landholders as well as data generation for wider purposes; (4) continual improvement of sample design efficiency through time.

One of the biggest problems in direct measurement of soil carbon is the expense of verification as we are dealing with the inherent variability of soil in the landscape (Allen et al. 2010; Smith 2004b), as discussed in Section E of this book. The amount of carbon stored in the soil per unit of land area is highly variable and depends on annual inputs, soil type, and the degradation rate of the soil C (Jandl et al. 2013).

Current methods for measuring, mapping, and quantifying soil carbon within an area are expensive and inefficient (Miklos et al. 2010). Furthermore, it is still not established how we can monitor changes in soil carbon efficiently and effectively with sufficient statistical confidence. A scheme that recognizes the whole farm as a system that can store carbon is crucial to the agricultural industry, particularly in the carbon economy. The applicability of direct measurement will likely expand with more efficient sampling and increasing analytical volumes through time.

Typically, repeat direct measurement of soil C for trend assessment has been carried out by revisiting intensively sampled plots (Bowman et al. 2002; Chappell et al. 2013) largely to reduce short-range variability and maximize change detection from repeat visits (Lark 2009). However, the optimal sample design for trend assessment of soil C due to management differences and establishing the soil C status of an area are two different objectives (de Gruijter et al. 2006; Lark 2009).

Traditionally, soil carbon stock inventory has been based on performing direct measurement of soil C content for a number of sampling sites within an area, then extrapolating the data to the desired extent. In Australia, the standard sampling unit for soil organic carbon is recommended as a 25 × 25 m square (McKenzie et al. 2000). The NSW Department of Environment, Climate Change and Water (DECCW) proposed that for a given unit of land, a 25 × 25 m subarea is chosen and divided into 10 × 10 equal sized quadrats. Ten samples are chosen at random from these 100 quadrats and the carbon concentration and soil bulk density estimated to a fixed depth of usually 50 cm. At some later date, the process is repeated in the same subarea. The difference in carbon content is calculated. The quadrat method has two drawbacks. First, the sampling area (i.e., the subarea) is known, which potentially can lead to dishonest practices where carbon may be deliberately sequestered in the known sampling area. More importantly, the extrapolation of the average carbon content from the smaller subarea to the larger unit of land under sequestration management leads to a large sampling variance resulting in an uncertain estimate of the change in carbon content (Singh et al. 2012).

Such limitations inherent with conventional methods of soil C stock assessment have prompted the development of new technologies that provide alternative methods of data acquisition, as will be discussed further in Chap. 13. Technologies such as remote sensing hyperspectral imagery (Denis et al. 2014; Stevens et al. 2010) and proximal sensing (Cremers et al. 2001; Gomez et al. 2008) have been proposed for estimating surface carbon content over large areas. However, such imagery only provides measurement on or near the soil surface; moreover the high cost of acquiring such data and the need for laboratory calibration limit its application in a routine auditing process. As discussed in Chaps. 5 and 7, field near-infrared spectroscopy can give estimates of carbon; however sampling is also required, and infrared estimation is based on calibration to standard analytical techniques (Stevens et al. 2010). Field prediction of SOC using NIR is still too uncertain, with a standard error of prediction ranging from 1.3 to 5.8 g/kg (Bellon-Maurel and McBratney 2011). In addition, correction of soil moisture from field-acquired spectra can also reduce the prediction accuracy (Ge et al. 2014). See Sect. 5.4.2.1 for further detail on

this too. A precise and cost-effective means of carbon storage measurement needs to be developed urgently in order to be able to credit farmers and landholders for their beneficial inputs in improving both our air and soil. This chapter will review various aspects on soil carbon auditing along with a case study on a farm in New South Wales, Australia.

23.2 Sampling Objective

Before designing an effective auditing scheme, the primary step needed is detailing the objectives in statistical terms (de Gruijter et al. 2006). Generally there has been not enough effort put into this part in order to design a cost-effective scheme. Most effort is directed into finding a cheaper analytical measurement for soil carbon; however, soil sampling is the largest cost in this activity. As a comparison, the analytical cost of total C by dry combustion is A$10 per sample, while sampling cost is on average A$100 per soil profile. Thus, a careful definition of the sampling objectives will allow effective and efficient sampling design and auditing process. The details of the objective that must be clearly defined in this work are (de Gruijter et al. 2006, p.29):

1. *Target universe*: boundaries of the universe in space and in time.
   In our case, the target universe in space is a farm, and along the time axis, the universe spans 3–7 years; the auditor will monitor the target variable at a particular time and repeat the same measure after 3–7 years to establish the change in the measure.
2. *Target variable*: the soil property to be measured on the sampling units.
   We take as target variable the soil organic carbon stock, in ton per ha, denoted by C. The C stock is measured to a fixed depth (up to a maximum depth of 1 m) or better to a fixed cumulative mass of soil, which deals with compaction/tillage issues (Wendt and Hauser 2013).
3. *Target parameter*: the statistic to be estimated from the sample data.
   In our case, the target parameter is the change in mean C stock between time t_1 and t_2, defined as

$$\Delta := \overline{C}_{t_1} - \overline{C}_{t_2}, \tag{23.1}$$

   where $\overline{C}_t$ is the spatially averaged C stock at time t, defined as

$$\overline{C}_t := \frac{1}{A} \int_{\mathbf{x} \in \mathcal{A}} C_t(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{23.2}$$

   where $C_t(\mathbf{x})$ is C stock up to 1 m, or to a fixed cumulative mass of soil, at location $\mathbf{x}$ at time t, $\mathcal{A}$ is the project area (the farm), and A is the surface area of the farm in ha.

In this study, we discretize the area by superimposing a fine grid, with N grid points serving as sampling units. The mean C stock is then redefined as the average over the grid points:

$$\overline{C}_t := \frac{1}{N} \sum_{i=1}^{N} C_{i,t} \tag{23.3}$$

where $C_{i,t}$ is the measured C stock at the ith grid point at time t.

23.3 Major Design Decisions

Three major design decisions are to be made: the sampling approach, the type of sampling design, and whether to bulk samples or not.

23.3.1 Sampling Approach

As repeated many times in the literature and discussed in Sect. 11.6, there are two primary approaches in sampling: design-based and model-based. Their ideal applications are for global estimates of a target variable and mapping the target variable, respectively (Brus and de Gruijter 1997). The design-based approach is therefore the most appropriate for C stock auditing. It implies that sampling locations are selected by probability sampling and that inference (e.g., estimation) is based on the sampling design used to select the sampling locations. A strong advantage in the context of carbon auditing is that there is no need to make model assumptions, which always remain questionable.

23.3.2 Type of Sampling Design

The simplest method in probability sampling is by simple random sampling (SRS): a fixed number of sample points are drawn at random and independently from each other within the target area. However, the efficiency of SRS can often be markedly improved by stratified simple random sampling (StSRS). This divides an area into subareas called strata, and SRS is applied in each stratum.

Appropriate stratification and allocation of sample sizes to the strata will usually lead to higher precision and lower cost of estimation (Cochran 1977). For a single sampling round, the precision can be expressed as the sampling variance of the estimated mean:

$$V(\hat{\overline{z}}) = \sum_{h=1}^{H} \left(\frac{N_h}{N} \right)^2 \frac{S_h^2}{n_h}, \tag{23.4}$$

where H is the number of strata, N_h is the size (number of grid points) of stratum h, S_h is the standard deviation of z in stratum h, and n_h is the sample size allocated to stratum h.

Given a stratification and a total sample size n, optimal allocation of sample sizes to the strata, in the sense of minimal sampling variance of the mean, can be realized by so-called Neyman allocation (Dalenius and Hodges 1959; Cochran 1977). The optimal sample size for stratum h is then given by:

$$n'_h = n \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} . \tag{23.5}$$

The change in total soil carbon between time t_1 and t_2, Δ, can be estimated as the difference between the estimates at time t_1 and t_2:

$$\widehat{\Delta} = \widehat{\overline{C}}_{t_1} - \widehat{\overline{C}}_{t_1} . \tag{23.6}$$

If measurement is based on sampling at the same locations in both rounds, then the variance of the estimated difference is given by

$$V(\widehat{\Delta}) = V(\widehat{\overline{C}}_{t_1} - \widehat{\overline{C}}_{t_1}) = V(\widehat{\overline{C}}_{t_1}) + V(\widehat{\overline{C}}_{t_2}) - 2\rho\sqrt{V(\widehat{\overline{C}}_{t_1})V(\widehat{\overline{C}}_{t_2})} , \tag{23.7}$$

where ρ is the temporal autocorrelation coefficient of the measurements at t_1 and t_2. However, to avoid the risk of fraudulent practices, it is prudent to sample at the second round at random locations that are selected independently from the first round. This is what we propose to do. So in our case, the covariance term in (23.7) equals zero, so

$$V(\widehat{\Delta}) = V(\widehat{\overline{C}}_{t_1}) + V(\widehat{\overline{C}}_{t_2}) . \tag{23.8}$$

Clearly, the certainty thus achieved may have a price. No advantage is taken from a possibly positive temporal autocorrelation, which would render a higher precision. On the other hand, improving the stratification for the second round by using new prior information (at least the first round sample data) will reduce $V(\widehat{\Delta})$. This advantage would be lost if the sample sites are revisited, because then the stratification must be kept as it is.

23.3.3 Method of Measurement

There are various methods to determine the C stock of a core sample, each having its own pros and cons in terms of costs and accuracy. The choice has consequences for the cost-effectiveness of the monitoring scheme as a whole and should be considered in relation with the decision on bulking (Sect. 23.3.4).

The C stock of a core is usually measured indirectly, by taking subsamples and measuring their water content and C concentration. The C stock is then calculated from C concentration (kg/kg) × bulk density (kg/m^3) × depth of core (m). Clearly, subsampling errors add to the analytical errors, thus increasing the total measurement error. To address this problem, Pallasser et al. (2015) developed a carbon determination system for whole soil cores. Their method seems promising for routine application in soil carbon auditing. It has the additional advantage of a direct measurement of both C stock and C concentration.

23.3.4 Sample Compositing

Stratified random sampling allows sample bulking or compositing in order to reduce laboratory costs. Soil materials sampled at the locations (aliquots) during a sampling round can be bulked and mixed thoroughly to form a composite or bulk sample. Each composite is formed by bulking a random selection of a pre-chosen number of aliquots, either within or across strata. Bulking is a common practice in sampling for soil testing. The basic assumption is that soil carbon is additive, and analyzing a bulk sample gives the same result as averaging the values of the individual aliquots (de Gruijter et al. 2006). Bulking across evenly sized strata has been successfully applied in estimation of mean soil P concentration in the Netherlands (Brus and Noij 2008) and was proposed for probability sampling in C stock estimation (Chappell et al. 2013).

The one and only advantage of compositing is cost reduction; however, there are four disadvantages. First, the cost reduction is due to a smaller number of measurements, which increases the contribution of measurement errors to the total estimation error of the mean. Second, sample sizes (total and within strata) can no longer be chosen freely, as there have to be multiples of 2, 3, or higher factors, depending on how many aliquots are bulked. (For instance, with compositing across ten strata, the sample size has to be a multiple of ten). This implies that the sampling variance cannot be minimized as effectively as in non-composite sampling. Third, as composites are larger than single aliquots, mixing and/or subsampling error will generally have a larger effect on the total estimation error than in non-composite sampling. Fourth, updating the prediction field using data from the first sampling round will be problematic because the geographical coordinates of the sample points are related only with averages over points. Local compositing (i.e., using an aggregate sample support) is a different matter and may be advantageous where there is substantial local variability (Lark 2012).

We carried out a preliminary analysis (not reported here) on the cost-effectiveness of compositing in our circumstances, based on assumed values of sampling costs, measurement costs, measurement and mixing error, and spatial variability of SOC. Our temporary conclusion is that the advantage of compositing will not outweigh its disadvantages for monitoring SOC at farm scale.

23.4 Stratification Methods

There are various ways to stratify a farm, which includes compact geographical stratification, stratification by ancillary variables, or stratification by a map of predictions of the target variable. A novel method is stratification by a map of predictions with uncertainties.

23.4.1 Compact Geographical Stratification

Compact geographical stratification works best when no information on the farm is available. It is just stratification on the basis of spatial coordinates to ensure that the target area is fully covered spatially. Typically a farm is represented by a regular grid. The spatial coordinates of the grid points can be stratified into compact geographical strata by minimizing the within strata sum of squared distances:

$$O_{\mathrm{MSSD}} = \sum_{i=1}^{N} \sum_{h=1}^{H} d_{ih}^2 \,, \tag{23.9}$$

where d_{ih} is the Euclidean distance between location i (x_i, y_i) and the mean of stratum h (b_h, c_h):

$$d_{ih}^2 = (x_i - b_h)^2 + (y_i - c_h)^2 \,. \tag{23.10}$$

The assumption is that the variable is spatially correlated so that the variation within each subregion is smaller than the global variation. An example of the compact geographical stratification for a farm is given in Fig. 23.1.

23.4.2 Stratification by Ancillary Variables

In most situations, ancillary variables will be available for the farm. Here, we can recognize the readily (and cheaply) available information: digital elevation models and aerial photography or satellite imagery. The second option is to survey the area using proximal soil sensors, such as electromagnetic induction and gamma radiometrics. Land use is particularly important in auditing soil carbon and needs to be incorporated in the stratification. We can perform k-means or fuzzy k-means clustering (McBratney and de Gruijter 1992) of the ancillary variables to come up with the strata. Miklos et al. (2010) utilized radiometric surveys to stratify the study area (farm) to estimate soil carbon stock. Similarly, Simbahan et al. (2006)

Fig. 23.1 Compact geographical stratification of the case-study farm (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

used ancillary variables to target sampling for digital soil mapping. However, both of these studies used model-based sampling strategy for the purpose of mapping soil carbon stock. In addition, collection of high-resolution ancillary variables using proximal soil sensing is expensive and would not be feasible for C stock auditing. k-means clustering minimizes the mean squared distance between the grid of ancillary multivariables and their nearest centroid. The clusters are represented by their centroids or means. In k-means the objective function is

$$O_{\mathrm{KM}} = \sum_{i=1}^{N} \sum_{h=1}^{H} d_{ih}^2 \,. \tag{23.11}$$

This is similar to (23.9), except that d here is the component of a distance matrix, calculated as

$$d_{ih}^2 = (\mathbf{x}_i - \mathbf{c}_h)' \mathbf{A} (\mathbf{x}_i - \mathbf{c}_h) \,, \tag{23.12}$$

where $\mathbf{c}_h$ is the class center (centroid) of class h and $\mathbf{A}$ is the distance norm matrix, which can be the inverse of variance-covariance matrix of $\mathbf{X}'$ or called Mahalanobis distance. An example in Fig. 23.2 shows the farm stratified in 12 clusters based on ancillary variables: elevation and total gamma radiometric counts for the farm in the case study. The assumption is that the selected ancillary variables are known to control the distribution of soil carbon, and the weights of these variables are equal.

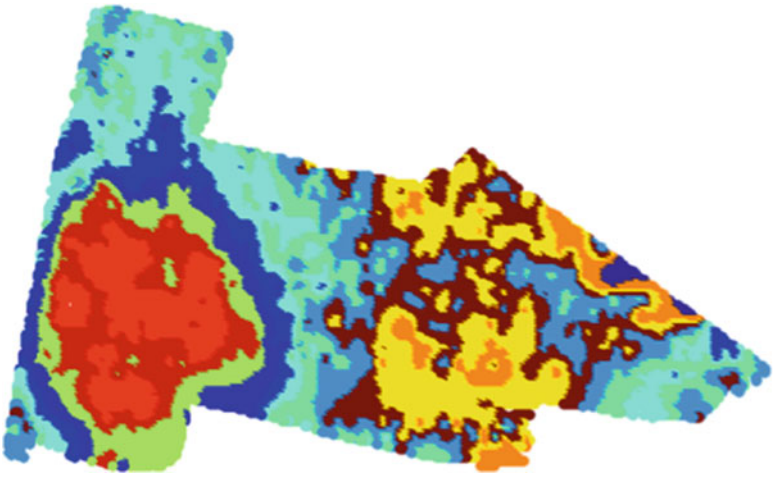


Fig. 23.2 Stratification by k-means clustering of ancillary variables of the case-study farm (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

23.4.3 Stratification by a Map of Predictions

SOC prediction fields generated from digital soil mapping (as outlined in Chap. 12 of this book) can be used as a source of univariate information for stratification (McBratney et al. 2011; Wheeler et al. 2012; Wheeler 2014). This approach uses the available quantitative knowledge, the relationships between the covariates (ancillary variables) and soil carbon, the knowledge being embodied in the model used to generate the spatial distribution of estimated carbon content. Effective stratification involves locating stratum boundaries along its target variable distribution and allocation of sample sizes to each stratum in a manner that increases the efficiency of the survey leading to higher precision or lower costs (de Gruijter et al. 2006).

An approximate solution to the problem is provided by the cumulative square root of the frequency (cum-root-f) method of Dalenius and Hodges (1959). To deal with the frequency distribution of skewed variables, the geometric stratification method (Gunning and Horgan 2004) or the method of Lavallee and Hidiroglou (1988) can be used to determine optimal stratum boundaries. This method attempts to minimize the sampling variance of the mean $V(\hat{\bar{z}})$. However, a disadvantage is that it assumes implicitly that the predictions have only negligible errors. An example in Fig. 23.3 shows the farm stratified in 12 strata by the cum-root-f method.

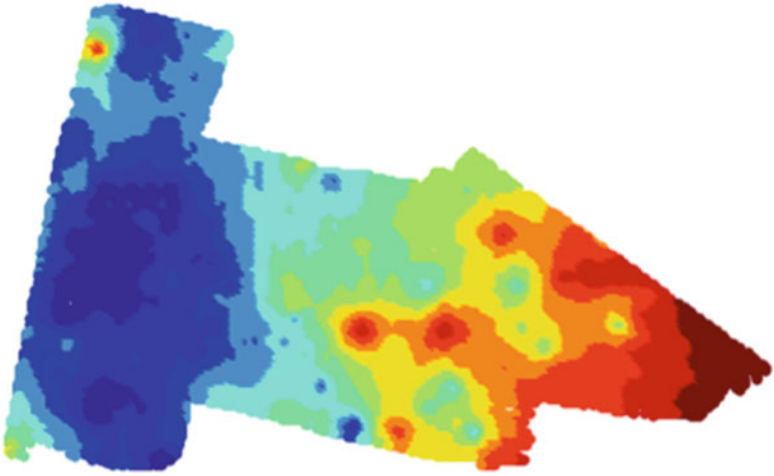


Fig. 23.3 Stratification of the carbon map of the case-study farm (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

23.4.4 Stratification by a Map of Predictions with Uncertainties

The cum-root-f method and its variants on the one hand and the compact geographic stratification method on the other hand represent two extreme solutions in the sense that they assume errorless predictions and absence of useful prediction, respectively. Realizing the limitations of the current methods, de Gruijter et al. (2015) proposed a new stratification method (*Ospats*), which uses a raster of predicted values with associated error variances. By taking prediction errors into account, this method can produce stratifications that represent transitions between the two extremes. The resulting stratification is optimized by minimizing the expected sampling variance, assuming optimal (Neyman) allocation of sample sizes to strata. This is implemented by minimizing:

$$O = \sum_{h=1}^{H} \left\{ \sum_{i=1}^{N_h - 1} \sum_{j=i+1}^{N_h} d_{ij}^2 \right\}^{1/2} , \tag{23.13}$$

where d_{ij}^2 here is the squared difference between the true values of the target variable C, which are unknown. However, d_{ij}^2 can be predicted using the predictions and their error variances. The prediction of d_{ij}^2 is obtained by taking its expectation, conditional on $\widetilde{C}_1 \cdots \widetilde{C}_N$:

$$D_{ij}^2 := \mathrm{E}_\xi (d_{ij}^2 | \widetilde{C}_i : i = 1, \ldots, N) , \tag{23.14}$$

which can be written out as

$$D_{ij}^2 = (\widetilde{C}_i - \widetilde{C}_j)^2 + V(e_i) + V(e_j) - 2\mathrm{Cov}(e_i, e_j)\,, \tag{23.15}$$

where $\widetilde{C}_i$ is the prediction of C_i, with prediction error e_i.

The objective function O is minimized by an iterative reallocation algorithm similar to those for k-means, except that here mutual distances are taken between grid points instead of distances between grid points and centroids. See de Gruijter et al. (2015) for further details.

The value of O resulting after minimization can be used to predict the sampling variance of the estimated mean under Neyman allocation, for a given total sample size n (cf. formula 18 in de Gruijter et al. (2015)):

$$\widetilde{V}(\widehat{\overline{C}}) = \frac{\overline{O}^2}{n}\,, \tag{23.16}$$

where $\overline{O} = O/N$.

23.5 Optimization Criterion

The total sample size is usually determined, either as the maximum affordable or via a targeted minimum detectable difference (MDD, the smallest detectable difference between means when the variation, significance level, statistical power, and sample size are specified) based on a prior estimate of the spatial variability of SOC (Garten and Wullschleger 1999; Saby et al. 2008; Singh et al. 2012; Smith 2004a). As a step toward further rationalization, we follow a value of information (VOI) approach, not a statistical one. The sample size will be determined so as to maximize the expected profit for the farmer, by financial quantification of the value of the sample data and the costs to collect the data. The VOI approach to decide rationally on the level of research investment is generally considered as superior to more or less arbitrary statistical criteria such as variance, power, or MDD. See Morgan et al. (1990) for a discussion of the VOI approach.

In the early 1970s, many quantitative studies have been devoted to the relation between quality and production costs of soil maps, notably by Philip Beckett and co-workers (Beckett and Burrough 1971). However, the problem is often too complex to apply the VOI approach, mainly because the financial consequences of differences in data quality are hard to quantify. This is probably the reason why the VOI approach has found little application in soil survey so far. See Bie and Ulph (1972) for an early application and more recently Giasson et al. (2000) and Knotters et al. (2010). As explained below, we can apply the VOI approach here, because we can quantify the value of sample data before they are collected.

The data value (*DV*) of a set of SOC sample data depends on the precision of the estimated sequestration $\widehat{\Delta}$. This is because the farmer will not be able to trade his sequestration on the basis of $\widehat{\Delta}$ alone, without accounting for uncertainty of the estimate. Therefore, we wish to determine a tradeable amount of sequestration *tp*, such that there is large probability γ (say 95%) that the future sequestration will amount to *tp* or more. To that order, we define *tp* as the lower boundary of a one-sided prediction interval around the predicted difference in estimated mean carbon contents, $\widehat{\Delta} = \widehat{\overline{C}}_{t_1} - \widehat{\overline{C}}_{t_1}$:

$$tp := \widehat{\Delta} - Z_\gamma \sqrt{\widetilde{V}(\widehat{\overline{C}}_{t_1}) + \widetilde{V}(\widehat{\overline{C}}_{t_2})}\,, \tag{23.17}$$

where Z_γ is the γ quantile of the standard normal distribution.

Of course, after the second round, *tp* will be calculated from the sample data as the lower boundary of the one-sided confidence interval:

$$tp = \widehat{\Delta} - t_{\alpha,\nu} \sqrt{\widehat{V}(\widehat{\overline{C}}_{t_1}) + \widehat{V}(\widehat{\overline{C}}_{t_2})}\,, \tag{23.18}$$

where $t_{\alpha,\nu}$ is the $(1-\alpha)$ quantile of the Student distribution with $\alpha = \gamma$ and $\nu = n_1 - H_1 + n_2 - H_2$ degrees of freedom (from both rounds).

Optimized stratification by *Ospats* and value of information as optimization criterion are core elements of the auditing process that we propose. The process as a whole is schematically presented in Table 23.1.

23.6 Optimization of the First Sampling Round Design

Using *Ospats* means that, for a given number of strata, the resulting stratification is optimized for any sample size, assuming optimal (Neyman) allocation of sample sizes to the strata. Thus, there are two design parameters left that are still to be optimized: the number of strata and the total sample size for that number.

First we optimize the sample size for a given number of strata. To do this, we need to make an assumption about the sampling variance from the second round relative to the first round, because the data value *DV* depends on the sum of these variances; see (23.17). For simplicity we assume that both variances will be equal. (This will happen, for instance, when the same stratification and the same sample size are used for both rounds, and the spatial variances within the strata do not change.) Under this assumption, (23.17) simplifies to

$$tp = \widehat{\Delta} - Z_\gamma \sqrt{2\widetilde{V}(\widehat{\overline{C}}_t)}\,, \tag{23.19}$$

where $\widetilde{V}(\widehat{\overline{C}}_t)$ follows from (23.16) and depends now only on n.

Table 23.1 Schematic overview of the auditing procedure

Step	Action
1	Preparation:
1a	Delineate the area
1b	Superimpose a grid with predictions and error variances
1c	Determine cost per grid point and carbon offset price
2	Optimize design for the first sampling round:
2a	Choose allowed minimum sample size within strata (e.g., 3)
2b	Choose a feasible range of strata numbers
2c	For each strata number in the range, calculate stratification (*Ospats*), total sample size (Eq. 23.21), and sample sizes within strata (Eq. 23.5)
2d	Select the design with the largest strata number that still fulfills the condition of step 2a
2e	Draw a stratified random sample according to the design from step 2d
3	Execute the first sampling round:
3a	Collect samples at the locations from step 2e, and take laboratory measurements to determine the carbon stock for each location
3b	Estimate the total carbon stock (Eq. 23.6) and its variance (Eq. 23.8)
4	Optimize design for the second sampling round:
4a	Update the predictions and error variances using the sample data from the first round
4b	Repeat step 2
5	Execute the second sampling round: repeat step 3
6	Finish: calculate the confidence interval for the total amount of sequestered carbon (Eq. 23.18)

We define the expected financial gain (G) as the data value DV minus the data costs DC, so $G := DV - DC$. In order to maximize G, we need a cost function for DC. We assume that a simple linear function for the variable costs of field and laboratory work suffices here, so $DC = f \cdot n$, where f is the average cost of obtaining the data per grid point. (The average traveling time between sample sites will decrease with increasing sample size. However, we expect that this will have only a minor effect, at least at farm scale.)

For the data value, we have $DV = CP \cdot A \cdot tp$, where CP is the price of sequestered carbon per mass unit and A is the surface area of the farm. So we wish to find the optimal sample size n' that maximizes G, using (23.19) and (23.16):

$$\begin{aligned} G &= DV - DC = CP \cdot A \cdot tp - f \cdot n \\ &= CP \left\{ A \cdot \widehat{\Delta} - A \cdot z_\gamma \sqrt{2\widetilde{V}(\widehat{\overline{C}}_t)} \right\} - f \cdot n \\ &= CP \cdot A \cdot \widehat{\Delta} - CP \cdot A \cdot z_\gamma \cdot \overline{O} \sqrt{\frac{2}{n}} - f \cdot n \end{aligned} \tag{23.20}$$

Equating the derivative $\mathrm{d}G/\mathrm{d}n$ to zero, renders

$$n' = \left(\frac{CP \cdot A \cdot Z_\gamma \cdot \overline{O}}{f\sqrt{2}} \right)^{2/3} , \tag{23.21}$$

a real which should be rounded to the nearest integer. The optimal sample sizes for the strata follow by imputing n' in (23.5) and rounding to the nearest integer. Note that the underlying assumption here is that tp will not exceed the sequestration capacity of the farm during the contract period, or a possible limit set in advance by the auditor or a regulator.

To see how, for a given number of strata, the gain changes with increasing sample size, we cannot use G as in (23.20) because that depends on $\widehat{\Delta}$, which is unknown before sampling. Therefore, we use an incremental gain, relative to $n = 1$:

$$G_{\mathrm{inc}}(n) := G(n) - G(1) . \tag{23.22}$$

Applying (23.20), this boils down to

$$G_{\mathrm{inc}}(n) = CP \cdot A \cdot Z_\gamma \cdot \overline{O}\sqrt{2} \left(1 - \sqrt{\frac{1}{n}} \right) - f(n-1) . \tag{23.23}$$

As $G(1)$ is negligible, $G_{\mathrm{inc}}(n)$ is practically equal to $G(n)$.

Having analyzed above how the sample size can be optimized given the number of strata, we turn now to the optimization of the number of strata itself, where we assume that for each possible number of strata, the optimal sample size will be chosen. Here, we have the inconvenience that (23.23) cannot be used for gain comparisons between design alternatives with different numbers of strata, as O differs between stratifications. To enable these comparisons, we redefine the incremental gain as a function of H, relative to $H = 1$ (i.e., no stratification) while standardizing the sample size on the optimal size for the given number of strata, n'_H:

$$G_{\mathrm{inc}}(H) := G(H, n'_H) - G(1, n'_1) , \tag{23.24}$$

where $G(H, n'_H)$ is the gain with H strata and the optimal sample size for H. Again applying (23.20), this boils down to

$$G_{\mathrm{inc}}(H) = CP \cdot A \cdot Z_\gamma \sqrt{2} \left(\overline{O}_1 \sqrt{\frac{1}{n'_1}} - \overline{O}_H \sqrt{\frac{1}{n'_H}} \right) + f(n'_1 - n'_H) , \tag{23.25}$$

where $\overline{O}_H$ is the value of $\overline{O}$ for the *Ospats* stratification with H strata. (Note that $\overline{O}_1$ follows from (23.13) without the need for iteration.)

23.7 Optimization of the Second Sampling Round Design

Optimization of the sampling design for the second round is similar to the first round, with one exception. The sample data from the first round are used to estimate the mean carbon content *and* the variance of that estimate. So the predicted sampling variance $\widetilde{V}(\widehat{\overline{C}}_{t_1})$ in the trading point equation (23.17) is replaced by the estimated sampling variance $\widehat{V}(\widehat{\overline{C}}_{t_1})$. So (23.17) is updated as

$$tp := \widehat{\Delta} - Z_\gamma \sqrt{\widehat{V}(\widehat{\overline{C}}_{t_1}) + \widetilde{V}(\widehat{\overline{C}}_{t_2})}\,, \tag{23.26}$$

where the first variance is a fixed quantity and the second is the result of optimizing H and n. As these variances cannot be assumed equal as for the first round, the gain equation (23.20) is updated as

$$\begin{aligned} G = DV - DC &= CP \cdot A \cdot tp - f \cdot n_2 \\ &= CP \cdot A \left\{ \widehat{\Delta} - Z_\gamma \sqrt{V_1 + \widetilde{V}(\widehat{\overline{C}}_{t_2})} \right\} - f \cdot n_2 \\ &= CP \cdot A \cdot \widehat{\Delta} - CP \cdot A \cdot Z_\gamma \sqrt{V_1 + \frac{\overline{O}_2^2}{n_2}} - f \cdot n_2 \end{aligned} \tag{23.27}$$

where n_2 is the sample size in the second round, $\overline{O}_2$ is the objective function value from the *Ospats* stratification used for the second round, and V_1 is shorthand for $\widehat{V}(\widehat{\overline{C}}_{t_1})$.

The optimal sample size is again found by equating the derivative of G with respect to n_2 to zero:

$$G' = \frac{CP \cdot A \cdot Z_\gamma \cdot \overline{O}_2^2}{2n_2^2 \sqrt{V_1 + \overline{O}_2^2 / n_2}} - f = 0 \tag{23.28}$$

which can be solved numerically by evaluation of G' for a range of n_2.

Equation (23.25) for the incremental gain as function of H is updated as:

$$G_{\text{inc}}(H) = CP \cdot A \cdot Z_\gamma \left(\sqrt{V_1 + \overline{O}_1^2 / n_1'} - \sqrt{V_1 + \overline{O}_H^2 / n_H'} \right) + f(n_1' - n_H') \tag{23.29}$$

23.8 Case Study

A study was performed at the University of Sydney E. J. Holtsbaum Agricultural Research property also known as the Nowley farm, located in North West Slopes and Plains of NSW, with an area of 23 km^2. Due to budget constraints, we could only

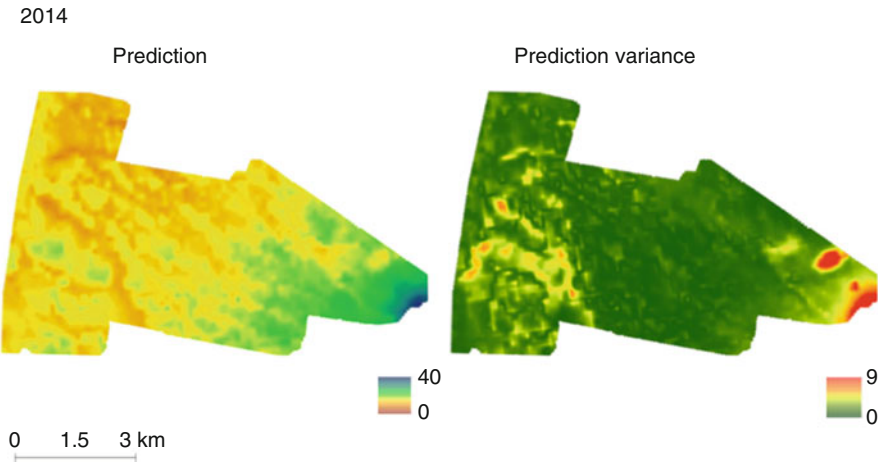


Fig. 23.4 Prediction and prediction variance of C stock in topsoil of Nowley farm (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

afford to take 50 samples. Within this limitation, we wanted to maximize the number of strata, which led to ten strata with five samples in each. Although the first round sampling design that we realized in this case study was predefined by the available budget, we still can demonstrate our data value approach to optimization, because this only needs a grid with predictions and a grid with associated uncertainties.

23.8.1 Optimizing the First Round Sampling

A digital map of C stock in the upper 7.5 cm was generated for the whole farm at a grid spacing of 10 × 10 m, along with its uncertainty (Fig. 23.4). The map was made using stepwise multiple linear regression, with covariates: elevation, terrain wetness index, gamma radiometric K, gamma radiometrics Th, and Landsat bands 2, 4, and 5.

$$\begin{aligned} C = {} & 4.34 + 0.09 \times Elevation - 4.56 \times GammaK + 1.22 \times GammaTh \\ & -1.62 \times LSBand2 + 0.17 \times LSBand4 + 0.37 \times LSBand5 \\ & -1.1 \times WetnessIndex \end{aligned} \tag{23.30}$$

The model was calibrated with data from 80 samples collected in 2014 in the same area. Model residuals showed no spatial autocorrelation. Leave-one-out cross validation gives RMSE = 5.1 and $R^2 = 0.42$.

Fig. 23.5 Stratification based on predictions and error variances in Fig. 23.4 (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

As indicated above, in this case study, we used sample data to calibrate a prediction model, which was then used to generate a carbon map. However, we expect that in practice our optimization system will be bootstrapped directly from a carbon map, so that there will be no need to collect data. Carbon maps at reasonably high resolution are becoming available. These have associated uncertainty and could be downscaled if necessary, for instance, Soil and Landscape Grid of Australia (Grundy et al. 2015) (http://www.clw.csiro.au/aclep/soilandlandscapegrid). See also Fig. 9 in Kidd et al. (2015) and Fig. 5 in Liddicoat et al. (2015).

The map of C stock and its uncertainty were used for stratifications with *Ospats*. Figure 23.5 shows the *Ospats* stratification with ten strata. Because of the current computational limit of the *Ospats* algorithm on grid size, the maps were coarse-gridded to 30 × 30 m. The resulting 25,955 grid data were used to optimize the number of strata, the total sample size, and the allocation of sample sizes to the strata.

In order to demonstrate the data value approach to optimization of the sample size for a given number of strata, we used the following parameters:

- Carbon offset price: $CP = A\$\,2.7$ per Mg CO_2 $= A\$\,10\ \mathrm{Mg}^{-1}$ C
- Cost of obtaining data per grid point: $f = A\$\,120$
- Surface area of the farm: $A = 2{,}336$ ha
- Number of grid points: $N = 25{,}955$
- 95% quantile of the standard normal distribution: $Z_\gamma = 1.645$
- Objective function value resulting from *Ospats*: depends on the number of strata. For instance, in the first round as realized (Sect. 23.8.2), we used ten strata, resulting in $\overline{O} = 6.33\ \mathrm{Mg} \cdot \mathrm{ha}^{-1}$

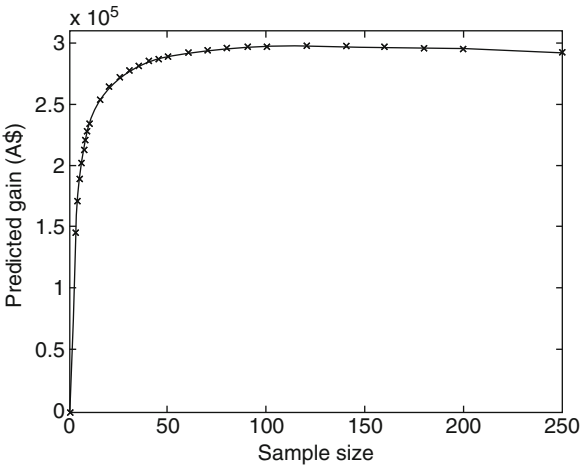


Fig. 23.6 Predicted gain [A$] as a function of first round sample size, using *Ospats* stratification with ten strata (see Fig. 23.5) (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

Table 23.2 Statistics of C sample data from Nowley farm, based on the *Ospats* stratification with ten strata and five samples per stratum

Strat.	Relative size	Mean	St. error estimated	St. error predicted	Optimal sample size
1	7.52	11.56	0.83	3.34	11
2	9.01	11.27	1.27	3.08	12
3	12.48	11.70	1.16	2.64	15
4	16.10	13.37	1.44	2.55	18
5	7.63	10.98	1.25	2.99	10
6	15.98	17.06	2.13	2.75	20
7	12.16	14.35	2.01	2.68	15
8	7.89	16.72	1.24	2.93	10
9	7.46	28.12	4.00	2.85	10
10	3.75	25.19	2.54	3.34	6
Farm	100.00	14.82	0.62	0.90	127

The optimal sample size, assuming Neyman allocation, follows from (23.21) and turns out to be 127. To see how the gain varies with sample size, we calculated the incremental gain G_{inc} using (23.23) for a range of sample sizes; see Fig. 23.6. The figure shows that the net return from sampling investments declines somewhat beyond $n = 127$, but the curve is surprisingly flat between $n = 50$ and 250. Table 23.2 shows the Neyman allocation for $n = 127$. The variation of the allocated sample sizes is largely due to the variation in surface areas between the strata, as their standard deviations are fairly even.

To optimize the number of strata, we maximize the predicted gain, under the condition that none of the sample sizes in the strata as determined by Neyman allocation is smaller than a pre-chosen minimum. The bare minimum is 2, to enable estimation of the sampling variance. However, in view of the possible loss of a

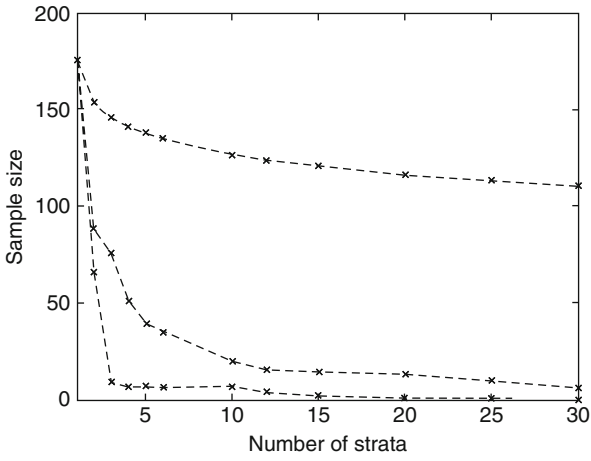


Fig. 23.7 Optimal total sample size (*upper line*), maximum sample size per stratum (*middle line*), and minimum sample size per stratum (*lower line*) as a function of first round stratum number (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

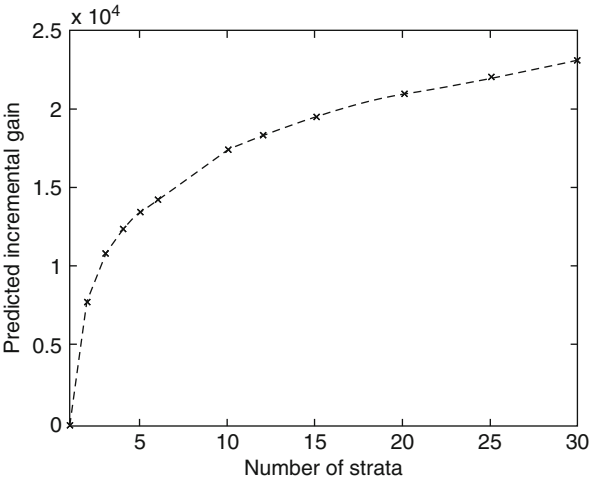


Fig. 23.8 Predicted incremental gain as a function of first round stratum number (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

sample, it is prudent to maintain a higher minimum, for instance, 3 or 4. So, for a range of possible stratum numbers, we calculate both the predicted incremental gain and the minimum sample size per stratum, given the optimal total sample size. Figure 23.7 shows how the optimal total sample size and the maximum and minimum allocated sample sizes change with increasing stratum number.

To optimize the number of strata, we applied (23.25) to a range of stratum numbers. See Fig. 23.8 for the predicted incremental gain as function of H, relative to $H = 1$ (no stratification). This figure shows that the more strata, the higher gain is predicted, although at an ever slower rate. For instance, using 12 strata instead of 10 renders an increase of A$ 970. It appears that 12 is the highest number of strata whereby all strata receive at least 3 samples via Neyman allocation. So in this case, 12 turns out to be the optimal number of strata under the condition that 3 is the allowed minimum.

23.8.2 Realization of the First Round Sampling

Ospats was used to stratify the farm into ten strata (Fig. 23.5), and five sample points were selected at random from each stratum. The survey was conducted in 2015, with locations determined using a GPS. Topsoil samples were collected at each location using a core with a diameter of 72 mm and height of 75 mm. Subsamples were air-dried and ground, and C concentration was determined using a vario MAX CN analyzer. C stock was then calculated as C concentration times bulk density times core height.

The statistics of the sample data are given in Table 23.2. The estimated mean carbon stock for the farm is $15.17\,\mathrm{Mg}\cdot\mathrm{ha}^{-1}$, with an estimated standard error of $0.62\,\mathrm{Mg}\cdot\mathrm{ha}^{-1}$. The standard error predicted by *Ospats* is $0.90\,\mathrm{Mg}\cdot\mathrm{ha}^{-1}$. Ideally, if both the estimate and the prediction were errorless, they should have been equal, but both figures have their uncertainty.

The data can also be used to estimate the sampling variance when simple random sampling (SRS) would have been applied with the same sample size. The spatial variance of the area $S^2(C)$ was estimated by:

$$\widehat{S^2}(C) = \widehat{\overline{C^2}} - \left(\widehat{\overline{C}}\right)^2 + \widehat{V}\left(\widehat{\overline{C}}\right) \tag{23.31}$$

cf. Equation 7.16 in de Gruijter et al. (2006). Divided by the sample size (50), this yielded a sampling variance for SRS of 0.752. The relative efficiency as compared to SRS equals 1.96, which is equivalent to 98 samples if it was conducted by SRS.

In conclusion, the *Ospats* stratification based on the available digital map of C stock was very efficient.

23.8.3 Optimizing the Second Round Sampling

Given the methods of optimal stratification and allocation, respectively, *Ospats* and Neyman, the remaining design parameters to be optimized are the number of strata H and the total sample size n. To update the stratification for the second round, one may improve the predictions by more advanced modeling and by using more predictive prior information. In this case, we calculated an updated carbon map using a multiple linear regression model (like for the first round), now predicting C content from elevation, gamma radiometric K, terrain wetness index, and weathering index. This model was calibrated with the sample data collected in 2014 (80) and 2015 (50). Again, the residuals showed no spatial autocorrelation. See Fig. 23.9 for the updated maps of predictions and prediction variances. The updated *Ospats* stratification with ten strata is displayed in Fig. 23.10.

To find the optimal sample size for ten strata, we used the gain formula (23.27), $V_1 = 0.3844$ as the variance estimated from the sample data, and $\overline{O}_2 = 1.0942$ resulting from the updated stratification. The gain as function of the sample size is

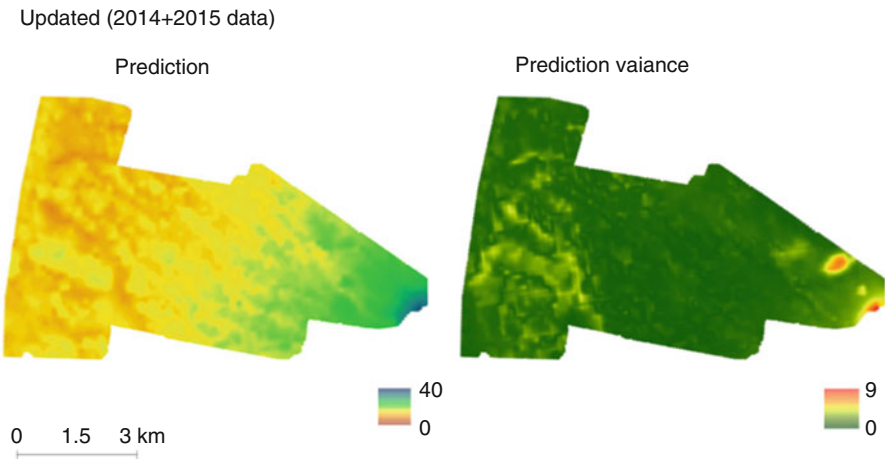


Fig. 23.9 Updated prediction and prediction variance of C stock in topsoil of Nowley farm (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

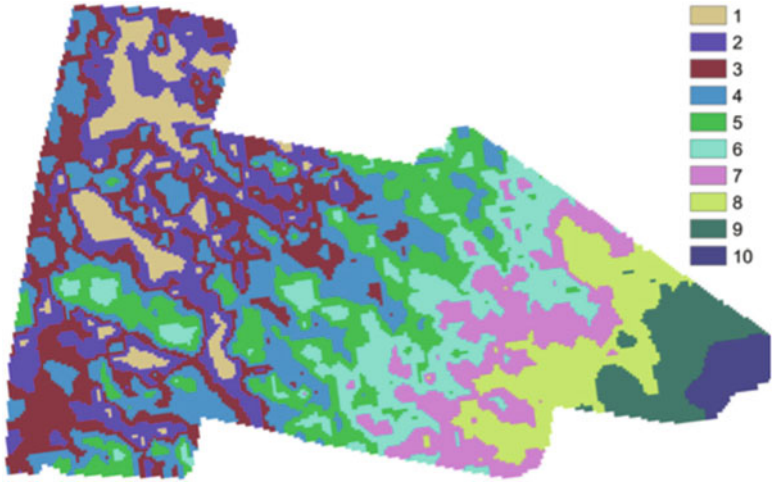


Fig. 23.10 Updated stratification based on predictions and error variances in Fig. 23.9 (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

shown in Fig. 23.11. It appears that the optimal sample size is only 15, implying that 10 strata is too many under the condition that each stratum receives at least 3 samples. Therefore, we determined the optimal sample size and Neyman allocation for a range of lower stratum numbers; see Fig. 23.12. This led to a maximum of 6 strata and a sample size of 22.

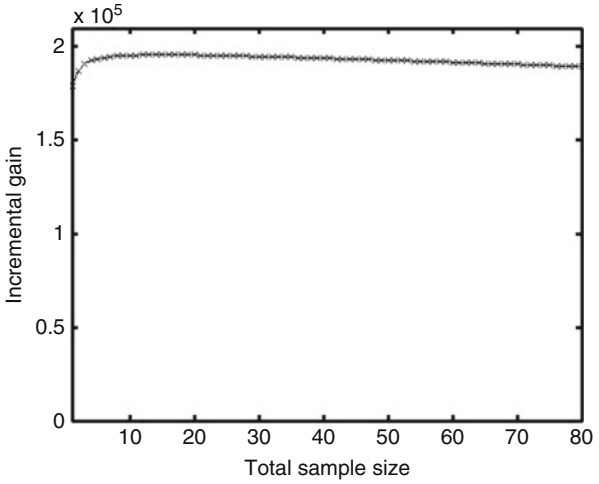


Fig. 23.11 Predicted gain [A$] as a function of second round sample size, using *Ospats* stratification with ten strata (see Fig. 23.10) (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

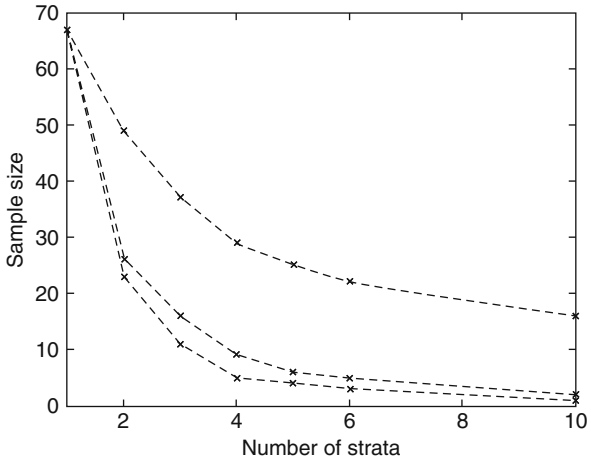


Fig. 23.12 Total sample size (*upper line*), maximum sample size per stratum (*middle line*), and minimum sample size per stratum (*lower line*), optimized for the second sampling round, as function of stratum number (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

The reason why only 6 instead of 10 strata and 22 samples instead of 50 suffice for the second round is that updating the carbon map with the first round sample data has enabled an even more efficient stratification than in the first round. This is shown in Fig. 23.13, presenting $\overline{O}$ as a function of stratum number for both rounds.

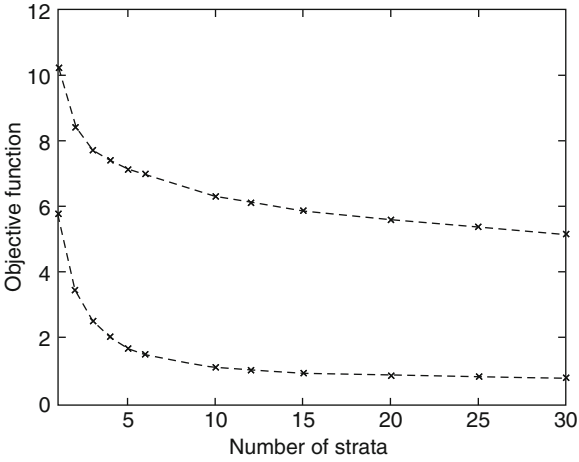


Fig. 23.13 Objective function $\overline{O}$ of stratifications for first round (*upper line*) and second round (*lower line*) as function of stratum number (Reprinted from de Gruijter J, McBratney AB, Minasny B, Wheeler I, Malone B, Stockmann U (2016) Farmscale soil carbon auditing. Geoderma 265, 120–130)

23.9 Discussion

The method devised contains two novel elements. First, a new method (*Ospats*) for optimal stratification (de Gruijter et al. 2015) is used here for soil carbon auditing. It facilitates effective exploitation of all relevant prior information about soil carbon in the project area, as condensed in a carbon map with associated uncertainty. Second, we developed a data value technique to optimize the number of strata as well as the sample size, which is novel in the context of carbon auditing. The optimization criterion of our technique, i.e., the predicted financial gain from a demonstrated amount of sequestration, should lead to decisions on sampling that are more rational than via general statistical criteria.

The information needed to start the optimization process, i.e., optimization for the first sampling round, is a carbon map with associated uncertainty. As carbon maps at reasonably high resolution are becoming available, prior data collection in the field will not be necessary. For instance, soil and landscape grids of Australia have associated uncertainty and could be downscaled if necessary.

The method is intended for auditing, not for monitoring nor management. For the latter purposes, there is generally no need for the restriction not to return to the same sites in subsequent sampling rounds. Efficiency of long-term monitoring may well profit from returning to (part of) the sites; see Brus and de Gruijter (2013). Also, the intended usage of monitoring results is usually complex, which makes the data value approach unfeasible.

As shown by the case study, updating the map with first round sample data may improve the resulting stratification considerably and hence reduce sampling costs. However, this updating option does not exist if the sample sites are revisited, because then the stratification too must stay unchanged.

In the case study, only the top 7.5 cm was sampled. This is shallower than usual in soil carbon auditing, but it is enough to demonstrate the method, as this can be applied likewise in projects involving deeper sampling.

The method has the following limitations and underlying assumptions:

1. The data value approach as devised here does only account for the costs that vary with the sample size, i.e., the costs of taking samples and laboratory analysis. Costs of travel to the farm and office work (administration, GIS, and computing) are not accounted for. This means that the gain predictions as calculated by the method can only serve to optimize a sampling design, not to decide on starting an audited sequestration project.
2. The gain predictions may be hypothetical because underlying assumptions are that the amount of carbon that will be sequestered by the farm does not reach a physical limit and that there is no regulatory set bound to it. However, if necessary the method can be easily adapted to account for a known limit.
3. The optimality of stratifications calculated by *Ospats* is only warranted as far as the uncertainty of the predictions is correctly quantified. At present we do not known how sensitive the stratification quality is for misrepresentation of the uncertainty.
4. Two sources of error are disregarded in this study: error in locating the sampling sites in the field and measurement errors in the laboratory. We expect that the disturbing effects of these error sources on the optimization will be moderate in general.

23.10 Conclusions

In this chapter, we presented a novel method for soil carbon auditing, which uses prior information in the form of a carbon map with associated uncertainty. The method is based on stratified random sampling and design-based inference about the amount of sequestered carbon. Stratification, total sample size, and sample sizes per stratum are mathematically optimized in conjunction. The criterion used is the expected financial gain (excluding fixed costs) for the farmer. This is maximized by a data value technique on the basis of assumptions about the costs of sampling and measurement and the price of sequestered carbon, given a required level of certainty about the amount of sequestered carbon.

An application on an Australia farm has shown that soil carbon changes across farms and regions can be audited effectively using the proposed stratification method and data value technique. The stratification method implies that strata will be created that are typically of unequal size and spatially non-contiguous. The former means that also the optimized sample sizes per stratum are unequal, and as a consequence, sample bulking across strata is unfeasible.

Updating the initial carbon map with sample data from the first round may considerably improve the efficiency of the stratification for the second round. As this stratification differs from the initial one, returning to the same sampling sites is unfeasible for design-based inference. Another reason not to return to the same sites is that it is not recommendable for auditing purposes, in order to avoid possible fraudulent practices and disturbance of sites.

Future research may focus on sensitivity of the auditing method for incorrect quantification of the uncertainty of the initial carbon map. Other research issues are the effects of spatial location error and measurement error on the optimization process and how these error sources could be accounted for.

Acknowledgements We acknowledge the support of the Australian Research Council via Linkage Project LP0989825 entitled "The auditability of soil carbon." We thank our linkage partners 3DAg Pty Ltd. for their continued support. Two anonymous reviewers gave helpful suggestions.

References

Allen DE, Pringle MJ, Page KL, Dalal RC (2010) A review of sampling designs for the measurement of soil organic carbon in Australian grazing lands. Rangel J 32:227–246

Batjes NH (1996) Total carbon and nitrogen in the soils of the world. Eur J Soil Sci 47:151–163

Beckett PHT, Burrough PA (1971) The relation between cost and utility in soil survey. IV. Comparison of the utilities of soil maps produced by different survey procedures, and to different scales. J Soil Sci 22:466–480

Bellon-Maurel V, McBratney AB (2011) Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – critical review and research perspectives. Soil Biol Biochem 43:1398–1410

Bie SW, Ulph A (1972) The economic value of soil survey information. J Agric Econ 23:285–297

Bowman RA, Reeder JD, Wienhold BJ (2002) Quantifying laboratory and field variability to assess potential for carbon sequestration. Commun Soil Sci Plant Anal 33:1629–1642

Brus DJ, de Gruijter JJ (1997) Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma 80:1–44

Brus DJ, de Gruijter JJ (2013) Effects of spatial pattern persistence on the performance of sampling designs for regional trend monitoring analyzed by simulation of space–time fields. Comput Geosci 61:175–183

Brus DJ, Noij IGAM (2008) Designing sampling schemes for effect monitoring of nutrient leaching from agricultural soils. Eur J Soil Sci 59:292–303

Chappell A, Baldock J, Viscarra Rossel R (2013) Sampling soil organic carbon to detect change over time. CSIRO Australia

Cochran W (1977) Sampling techniques. Wiley, New York

Conant RT, Paustian K (2002) Spatial variability of soil organic carbon in grasslands: implications for detecting change at different scales. Environ Pollut 116:S127–S135

Cremers DA, Ebinger MH, Breshears DD, Unkefer PJ, Kammerdiener SA, Ferris MJ, Catlett KM, Brown JR (2001) Measuring total soil carbon with laser-induced breakdown spectroscopy (LIBS). J Environ Qual 30:2202–2206

Dalal RC, Chan KY (2001) Soil organic matter in rainfed cropping systems of the Australian cereal belt. Aust J Soil Res 9:435–464

Dalenius T, Hodges JL (1959) Minimum variance stratification. J Am Stat Assoc 54:88–101

de Gruijter JJ, Brus DJ, Bierkens MFP, Knotters M (2006) Sampling for natural resource monitoring. Springer, Berlin

de Gruijter JJ, Minasny B, McBratney AB (2015) Optimizing stratification and allocation for design-based estimation of spatial means using predictions with error. J Surv Stat Methodol 3:19–42

Denis A, Stevens A, van Wesemael B, Udelhoven T, Tychon B (2014) Soil organic carbon assessment by field and airborne spectrometry in bare croplands: accounting for soil surface roughness. Geoderma 226–227:94–102

Garten CT Jr, Wullschleger SD (1999) Soil carbon inventories under a bioenergy crop (Switchgrass): measurement limitations. J Environ Qual 28:1359–1365

Ge Y, Morgan CLS, Ackerson JP (2014) VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. Geoderma 213:61–69

Giasson E, van Es C, van Wambeke A, Bryant RB (2000) Assessing the economic value of soil information using decision analysis techniques. Soil Sci 165:971–978

Gomez C, Viscarra Rossel RA, McBratney AB (2008) Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: an Australian case study. Geoderma 146:403–411

Grundy MJ, Viscarra Rossel RA, Searle RD, Wilson PL, Chen C, Gregory LJ (2015) Soil and landscape grid of Australia. Soil Res 53:835–844

Gunning P, Horgan JM (2004) A simple algorithm for stratifying skewed populations. Surv Methodol 30:159–185

Jandl R, Rodeghiero M, Martinez C, Cotrufo MF, Bampa F, van Wesemael B, Harrison RB, Guerrini IA, Richter DD, Rustad L, Lorenz K, Chabbi A, Miglietta F (2013) Current status, uncertainty and future needs in soil organic carbon monitoring. Sci Total Environ 468–469:376–383

Kidd D, Webb M, Malone B, Minasny B, McBratney A (2015) Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. Soil Res 53:932–955

Knotters M, Vroon H, van Kekem A, Hoogland T (2010) Deciding on the detail of soil survey in estimating crop yield reduction due to groundwater withdrawal. In: Devillers R, Goodchild H (eds) Spatial data quality: from process to decisions. CRC Press/Taylor Francis Group, Boca Raton, pp 117–125

Lal R (2004) Soil carbon sequestration impacts on global climate change and food security. Science 304:1623–1627

Lark RM (2009) Estimating the regional mean status and change of soil properties: two distinct objectives for soil survey. Eur J Soil Sci 60:748–756

Lark RM (2012) Some considerations on aggregate sample supports for soil inventory and monitoring. Eur J Soil Sci 63:86–95

Lavallee P, Hidiroglou M (1988) On the stratification of skewed populations. Surv Methodol 14:33–43

Liddicoat C, Maschmedt D, Clifford D, Searle R, Herrmann T, Macdonald LM, Baldock J (2015) Predictive mapping of soil organic carbon stocks in South Australia's agricultural zone. Soil Res 53:956–973

Luo Z, Wang E, Sun OJ (2010) Soil carbon change and its responses to agricultural practices in Australian agro-ecosystems: a review and synthesis. Geoderma 155:211–223

McBratney AB, de Gruijter JJ (1992) A continuum approach to soil classification by modified fuzzy k-means with extragrades. J Soil Sci 43:159–175

McBratney AB, Minasny B, de Gruijter JJ, Mulvey PJ (2011) A method of quantifying soil carbon. Australian Patent 20112651179

McKenzie N, Ryan P, Fogarty P, Wood J (2000) Sampling, measurement and analytical protocols for carbon estimation in soil, litter and coarse woody debris. National carbon accounting system technical report 14, Australian Greenhouse Office, Canberra

Miklos M, Short MG, McBratney AB, Minasny B (2010) Mapping and comparing the distribution of soil carbon under cropping and grazing management practices in Narrabri, North-West New South Wales. Aust J Soil Res 48:248–257

Morgan MG, Henrion M, Small M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge

Pallasser R, Minasny B, McBratney AB (2015) Carbon determination system for whole soil cores. Commun Soil Sci Plant Anal 46:221–234

Paustian K, Six J, Elliott ET, Hunt HW (2000) Management options for reducing CO2 emissions from agricultural soils. Biogeochemistry 48:147–163

Paustian K, Brenner J, Easter M, Killian K, Ogle S, Olson C, Schuler J, Vining R, Williams S (2009) Counting carbon on the farm: reaping the benefits of carbon offset programs. J Soil Water Conserv 64:36A–40A

Post WM, Izaurralde RC, Mann LK, Bliss N (2001) Monitoring and verifying changes of organic carbon in soil. Clim Chang 51:73–99

Rabotyagov SS (2010) Ecosystem services under benefit and cost uncertainty: an application to soil carbon sequestration. Land Econ 86:668–686

Rosenberg NJ, Izaurralde RC (2001) Storing carbon in agricultural soils to help head-off a global warming: guest editorial. Clim Chang 51:1–10

Saby NPA, Bellamy PH, Morvan X, Arrouays D, Jones RJA, Verheijen FGA, Kibblewhite MG, Verdoodt A, Üveges JB, Freudenschuß A, Simota C (2008) Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? Glob Chang Biol 14:2432–2442

Simbahan GC, Dobermann A, Goovaerts P, Ping J, Haddix ML (2006) Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132:471–489

Singh K, Murphy BW, Marchant BP (2012) Towards cost-effective estimation of soil carbon stocks at the field scale. Soil Res 50:672–684

Smith P (2004a) How long before a change in soil organic carbon can be detected? Glob Chang Biol 10:1878–1883

Smith P (2004b) Monitoring and verification of soil carbon changes under Article 3.4 of the Kyoto Protocol. Soil Use Manag 20:264–270

Stevens A, Udelhoven T, Denis A, Tychon B, Lioy R, Hoffmann L, van Wesemael B (2010) Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158:32–45

Stockmann U, Adams MA, Crawford JW, Field DJ, Henakaarchchi N, Jenkins M, Minasny B, McBratney AB, Courcelles VDRD, Singh K, Wheeler I, Abbott L, Angers DA, Baldock J, Bird M, Brookes PC, Chenu C, Jastrow JD, Lal R, Lehmann J, O'Donnell AG, Parton WJ, Whitehead D, Zimmermann M (2013) The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agric Ecosyst Environ 164:80–99

Wendt JW, Hauser S (2013) An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. Eur J Soil Sci 64:58–65

Wheeler I (2014) Monitoring total soil organic carbon at farm scale. PhD thesis, The University of Sydney

Wheeler I, McBratney AB, Minasny B, de Gruijter JJ (2012) Digital soil mapping to inform design-based sampling strategies for estimating total organic carbon stocks at the farm scale. In: Digital soil assessments and beyond – proceedings of the fifth global workshop on digital soil mapping, Sydney, pp 257–262