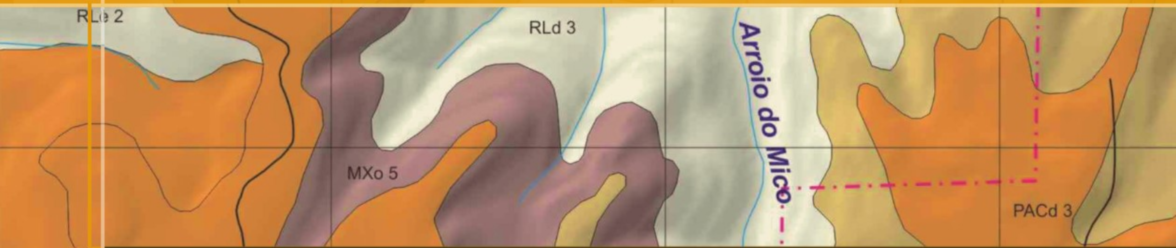
A.E. Hartemink
A. McBratney
M.L. Mendonça-Santos
(Eds.)



Digital Soil Mapping with Limited Data

Springer

Digital Soil Mapping with Limited Data

Alfred E. Hartemink · Alex McBratney ·
Maria de Lourdes Mendonça-Santos
Editors

Digital Soil Mapping with Limited Data

With a foreword by Robert J. Ahrens

Springer

Editors
Alfred E. Hartemink
ISRIC – World Soil Information
6700 AJ Wageningen
The Netherlands
alfred.hartemink@wur.nl

Maria de Lourdes Mendonça-Santos
EMBRAPA-Brazilian Agricultural Research
Corporation/CNPS-The National
Centre of Soil Research
Rua Jardim Botânico. 1024 CEP.22.460-000
Rio de Janeiro-RJ, Brazil
lou.mendonca@cnps.embrapa.br

Alex McBratney
University of Sydney
McMillan Building A05
Science Rd.
NSW, Camperdown, Sydney 2006
Fac. of Agriculture
Australia
A.McBratney@usyd.edu.au

ISBN: 978-1-4020-8591-8 e-ISBN: 978-1-4020-8592-5

Library of Congress Control Number: 2008927194

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Foreword

Significant technological advances have been few and far between in the past approximately one hundred years of soil survey activities. Perhaps one of the most innovative techniques in the history of soil survey was the introduction of aerial photographs as base maps for field mapping, which replaced the conventional base map laboriously prepared by planetable and alidade. Such a relatively simple idea by today's standards revolutionized soil surveys by vastly increasing the accuracy and efficiently. Yet, even this innovative approach did not gain universal acceptance immediately and was hampered by a lack of aerial coverage of the world, funds to cover the costs, and in some cases a reluctance by some soil mappers and cartographers to change.

Digital Soil Mapping (DSM), which is already being used and tested by groups of dedicated and innovative pedologists, is perhaps the next great advancement in delivering soil survey information. However, like many new technologies, it too has yet to gain universal acceptance and is hampered by ignorance on the part of some pedologists and other scientists.

DSM is a spatial soil information system created by numerical models that account for the spatial and temporal variations of soil properties based on soil information and related environmental variables (Lagacheric and McBratney, 2007). Pedologists working with DSM technology are in the process of addressing questions and concerns. Some of these questions include production and processing of covariates (soil forming factors derived from remote sensing and existing soil maps), the collection of soil data, the development of soil predictions based on numerical models, and the representation of digital soil maps.

Covariates include the traditional soil forming factors of parent material, topography, vegetation, and climate. Some of the more sophisticated remote sensing techniques help glean information on the mineralogy and specific properties of the surface layers or horizons. The ever expanding application of remote sensing and associated decrease in costs open the doors for advantageous development of stronger soil covariates and improvement to the predictive utility of DSM.

Traditional soil survey has always struggled with the collection of data. The amount of soil data and information required to justify the mapping product, how to interpolate data to similar areas, and how to incorporate older data are all challenges that need further discussion. Older data often were collected with antiquated or imprecise terms and must be cross-referenced to current standards, but the biggest obstacle in using older data is the lack of georeferencing. Traditional soil surveys have tried to write standards for data collection, but the practicality of applying the standards is difficult and not completely satisfactory. The amount of data is dependent on the complexity of the area and the experience of the mapper among other things. DSM also needs some guides or standards that will be difficult to cultivate to meet everyone's expectations and requirements. Some of the most ardent discussions in pedology center around standards, including different soil classification systems, and seemingly fail to concentrate and evaluate the end-product, which is the soil information provided to the user. DSM is a technological advancement that has the potential to be misunderstood and thus viewed with skepticism.

Numerical models are the functions that predict soil properties or soil classes. Most of these models have been calibrated with soil samplings and have been tested over small areas. The limitation of soil sampling dense enough to capture the spatial variability presently somewhat limit the use of numerical models to for large areas.

The world's overpopulation of the human race and associated pressures on resources, necessitate the immediate need for valuable soil information to make informed decisions about the soil resource, or, at the very least, make people aware of the problems and potential problems. We do not have the time or resources to canvass the earth and gather all the soil data and information needed to make soil surveys by our traditional methods. We need to look at the data that we do have and employ new methods and new technologies to deliver information on the soil resource. At the same time we should not be enamored solely on technology without an appreciation and understanding of soil-landscape relationships, which provide the predictive tools and foundations of soil survey.

DSM has the potential to deliver the needed information and in fact may provide better and more accurate information. However, the technology of DSM must overcome the skepticism associated with any new technology in the traditional world of soil survey where new technologies have been few and far between.

The purpose of this book is to present the latest technologies, challenges, and ideas related to DSM. Papers in this book were presented at the second Global Workshop on Digital Soil Mapping for Regions and Countries with Sparse Soil Spatial Data Infrastructures, which was held in Rio de Janeiro in July 2006. The EMBRAPA CNPS (Brazilian National Soil Research Centre) hosted the meeting, and the organizing committee was co-chaired by Dr. Lou Mendonça-Santos of EMBRAPA Solos and Prof. Alex McBratney of The University of Sydney, Australia. Chapters range from overviews of the DSM technology in general to specific applications in areas without much soil information or areas where specific parameters are investigated. Case studies in different parts of the world provide the opportunity to evaluate the information and test its utility. I invite you all to engage in this new technology, keep an open mind, continue the lively discussions that have always made pedology exciting and enjoyable, and in the process strive to save our most valuable resource, the soil.

USDA-NRCS Robert J. Ahrens

Reference

Lagacherie, P., McBratney, A.B., 2007. Chapter 1. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Editors), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science, volume 31. Elsevier, Amsterdam, pp. 3–24.

Preface

This book reports on the second of a series of global workshops on digital soil mapping held in Rio de Janeiro in July 2006 which coincided with the FIFA World Cup – so it was an exciting time to be in Brazil. The meeting was hosted by EMBRAPA Solos (Brazilian National Soil Research Centre) and the organizing committee was co-chaired by Dr. M.L. Mendonça-Santos of EMBRAPA Solos and Prof. Alex McBratney of the University of Sydney. The meeting was organised with the financial support of EMBRAPA, FAPERJ (Carlos Chagas Filho Foundation for Research Support of Rio de Janeiro State), CNPq (The National Council for Scientific and Technological Development) and CPRM (Brazilian Geological Service). There were some 100 participants from 20 countries.

The theme of the workshop was Digital Soil Mapping for Regions and Countries with Sparse Soil Spatial Data Infrastructures.

There has been considerable expansion in the use of digital soil mapping technologies and development of methodologies that improve digital soil mapping at all scales and levels of resolutions. These developments have occurred in all parts of the world in the past few years also in countries where it was previously absent. Much in the same way money and time are always short, there is almost always a shortage of data in soil research and its applications. That may lead to unsupported statements, sloppy statistics, misrepresentations and ultimately bad resource management. In digital soil mapping, maximum use is made of sparse data and this books contains several examples how that can be done.

From the Rio de Janeiro workshop we have selected 34 papers that focused on digital soil mapping methodologies and applications for areas where data are limited. The papers have been loosely grouped into the following sections (i) introductory papers, (ii) dealing with limited spatial data infrastructures, (iii) methodology development, and (iv) examples of digital soil mapping in various parts of the globe (including USA, Brazil, UK, France, Czech Republic, Honduras, Kenya, Australia). The last chapter summarises priorities for digital soil mapping. We feel this book is a logical development of the ideas presented in "Digital soil mapping – an introductory perspective", edited by Lagacherie et al., (2007) in the Developments in Soil Science Series.

Wageningen	A.E. Hartemink
Sydney	A.B. McBratney
Rio de Janeiro	M.L. Mendonça-Santos

Contents

Contributors

Anjos, L.H.C.
Universidade Federal Rural do Rio de Janeiro, BR 465 – Km47, Seropédica, RJ, Brazil, CEP 23890-000, lanjos@ufrrj.br

Arrouays, D.
INRA Unité INFOSOL, 2163 avenue de la pomme de pin, BP 20619 Ardon, F-45166 Olivet Cedex, France, Dominique.Arrouays@orleans.inra.fr

Arruda, W.
Instituto Nacional de Pesquisas da Amazônia – INPA – Manaus. Av. André Araújo – 2936, Manaus – AM, 69060-001

Bacic, I.L.Z.
Epagri/Ciram – Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S.A., Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina, Rodovia Admar Gonzaga 1347, C.P. 502. 88034-901, Florianópolis, Santa Catarina, Brazil, bacic@epagri.sc.gov.br

Bagnell, J.A.
Robotics Institute, Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA, bagnell2@andrew.cmu.edu

Bai, Z.G.
ISRIC – World Soil Information, P.O. Box 353, 6700 AJ, Wageningen, The Netherlands, zhanguo.bai@wur.nl

Baxter, S.J.
Department of Soil Science, The University of Reading, Whiteknights, P.O. Box 233, Reading, England, RG6 6DW, UK, sam.baxter2@googlemail.com

Behrens, T.
Department of Physical Geography, Institute of Geography, University of Tübingen, Tübingen, Germany, thorsten.behrens@uni-tuebingen.de

Bell, J.C.
Department of Soil, Water and Climate, Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, jaybell@umn.edu

Benites, V.M.
Embrapa Soils, Rua Jardim Botanico, 1024, Jardim Botanico, Rio de Janeiro, RJ, Brazil, CEP 22460-000, vinicius@cnps.embrapa.br

Berbara, R.L.L.
Universidade Federal Rural do Rio de Janeiro, BR 465 – Km47, Seropédica, RJ, Brazil, CEP 23890-000, berbara@ufrrj.br

Bodily, J.M.
USDA Natural Resources Conservation Service, 340 N 600 E, Richfield, UT 84701-2258, Jedd.Bodily@ut.gov.usda

Boettinger, J.L.
Department of Plants, Soils, and Climate, Utah State University, 4820 Old Main Hill, Logan, UT 84322-4820, USA, Janis.Boettinger@usu.edu

Boruvka, L.
Department of Soil Science and Geology, Czech University of Life Sciences in Prague, CZ-165 21, Prague 6 – Suchdol, Czech Republic, Boruvka@af.czu.cz

Brown, S.E.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, UK

Burt, J.E.
Department of Geography, University of Wisconsin, Madison, Wisconsin, USA, jeburt@wisc.edu

Carré, F.
European Commission, Directorate General Joint Research Centre, Land Management & Natural Hazards Unit, TP 280, 21020 Ispra (VA), Italy, Florence.Carre@jrc.it

Carvalho Junior, W.
Embrapa Solos, Rua Jardim Botânico, 1024. Rio de Janeiro, RJ, Brazil, CEP 22.460-000, waldircj@cnps.embrapa.br

Chagas, C.S.
Embrapa Solos, Rua Jardim Botânico, 1024. Rio de Janeiro, RJ, Brazil, CEP 22.460-000, chagas.rj@gmail.com

Clarke, R.T.
Hydraulic Research Institute (IPH), Federal University of RioGrande do Sul (UFRGS). Av. Bento Gonçalves, 9500, Porto Alegre, RS, Brazil, 91501-970, clarke@iph.ufrgs.br

Cole, N.J.
USDA Natural Resources Conservation Service, 621 W Fetterman St., Buffalo, WY 82834

Cook, S.E.
International Centre for Tropical Agricultura (CIAT), AA6713, Cali, Colombia, s.cook@cgiar.org

Cooke, H.J.
Cranfield University, National Soil Resources Institute, Building 37, Cranfield MK43 0AL, UK, h.cook@cranfield.ac.uk

Corpetti, T.
Equipe COSTEL, UMR CNRS 6554 (Littoral Environnement Télédétection et Géomatique), Université de Rennes 2, place du recteur H. le Moal, 35 043 Rennes cedex France, thomas.corpetti@uhb.fr

Crawford, D.M.
Department of Primary Industries, Primary Industries Research Victoria, Werribee Centre, 621 Sneydes Road, Werribee, Victoria, 3030, Australia, Doug.Crawford@dpi.vic.gov.au

Daroussin, J.
Unité Science du Sol, INRA Centre d'Orléans, Avenue de la Pomme de Pin, BP 20619 Ardon, 45166 OLIVET Cedex, France, Joel.Daroussin@orleans.inra.fr

Dart, R.O.
Embrapa's fellowship grant, EMBRAPA Solos – Brazilian Agricultural Research Corporation, The National Centre of Soil Research, Rua Jardim Botânico 1024. Rio de Janeiro, RJ, Brazil, CEP 22.460-000, rdart81@yahoo.com.br

Dent, D.L.
ISRIC – World Soil Information, P.O. Box 353, 6700 AJ, Wageningen, The Netherlands, david.dent@wur.nl

Dias, M.B.
Robotics Institute, Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA, bdias@andrew.cmu.edu

Drabek, O.
Department of Soil Science and Geology, Czech University of Life Sciences in Prague, CZ-165 21, Prague 6 – Suchdol, Czech Republic, Drabek@af.czu.cz

English, E.
Department of Geography, University of Wisconsin, Madison, Wisconsin, USA, EEnglish@wisc.edu

Fasolo, P.J.
Embrapa Florestas, Estrada da Ribeira Km 11, CP. 319, CEP 83411-000, Colombo, PR, Brazil

Figueiredo, S.R.
Soil Science Graduate School of the Federal University of Rio Grande do Sul (UFRGS). Av. Bento Gonçalves, 7712, Porto Alegre, RS, Brazil, 90.001-970

Filho, E.I.F.
Universidade Federal de Viçosa. Avenida Peter Henry Rolfs, s/n, Campus Universitário. CEP 36570-000. VIÇOSA – MG. Brazil, elpidio@ufv.br

Flores, C.A.
Embrapa Clima Temperado, BR 392 Km 78, CP. 403, CEP 96001-970, Pelotas, RS, Brasil, flores@cpact.embrapa.br

Follain, S.
UMR LISAH SupAgro/INRA/IRD, 2 place Pierre Viala, F-34060 Montpellier Cedex 1, France, Follain@supagro.inra.fr

Giasson, E.
Department of Soils of the Federal University of Rio Grande do Sul (UFRGS). Av. Bento Gonçalves, 7712, Porto Alegre, RS, Brazil, 90.001-970, giasson@ufrgs.br

Golden, M.L.
USDA Natural Resources Conservation Service, Rm. 4252-S P.O. Box 2890, Washington, DC 20013-2890, Micheal.golden@wdc.usda.gov

Gonzalez, J.P.
Robotics Institute, Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA, jpgonzal@cmu.edu

Gouéry, P.
Equipe développement, Centre de Ressource Informatique, Université de Rennes 2, place du recteur H. le Moal, 35 043 Rennes cedex France, pascal.gouery@uhb.fr

Gregory, L.J.
CSIRO Land and Water, G.P.O. Box 1666, Canberra, ACT 2602, Australia

Hammer, R.D.
USDA Natural Resources Conservation Service, 100 Centennial Mall North, Federal Bldg, Rm 152, Lincoln, NE 68508, david.hammer@lin.usda.gov

Hartemink, A.E.
ISRIC – World Soil Information, P.O. Box 353, 6700 AJ, Wageningen, The Netherlands, alfred.hartemink@wur.nl

Hasenack, H.
Universidade Federal do Rio Grande do Sul, Centro de Ecologia, CP. 15007, CEP 91501-970, Porto Alegre, RS, Brasil, hasenack@ecologia.ufrgs.br

Haydu-Houdeshell, C.A.
USDA Natural Resources Conservation Service, 14393 Park Avenue, Suite 200 Victorville, CA 92392, Carrie-Ann.Houdeshell@ca.usda.gov

Hempel, J.W.
USDA Natural Resources Conservation Service, 157 Clark Hall Annex, P.O. Box 6301, Morgantown, WV 26506, jon.hempel@wv.usda.gov

Hott, M.C.
EMBRAPA National Dairy Cattle Research Center, 610 Eugênio do Nascimento Street, Juiz de Fora, MG, Brazil 36038-330, hott@cnpgl.embrapa.br

Howell, D.
USDA Natural Resources Conservation Service, 1125 16th Street, Room 219 Arcata, CA, USA 95521, David.Howell@ca.usda.gov

Iwata, S.
Petrobras – UNBSOL – Manaus. Rua Darcy Vargas, 645, Parque 10 de Novembro, Manaus – AM

Jacquier, D.W.
CSIRO Land and Water, G.P.O. Box 1666, Canberra, ACT 2602, Australia

Jarvis, A.
International Centre for Tropical Agricultura (CIAT), AA6713, Cali, Colombia; Bioversity International, Regional Office for the Americas, Cali, Colombia, a.jarvis@cgiar.org

Jordan, C.J.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, UK

Kessler, H.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, UK

Kienast-Brown, S.
USDA Natural Resources Conservation Service, Department of Plants, Soils, and Climate, Utah State University, 4820 Old Main Hill, Logan, UT 84322-4820, USA, Suzann.kienast@ut.usda.gov

Kim, Y.G.
Humboldt State University, Department of Mathematics, Arcata, CA, USA 95521, ygk1@humboldt.edu

Krol, B.G.C.M.
International Institute for Geo-Information Science and Earth Observation (ITC), Postbus 6, 7500 AA, Enschede, The Netherlands, krol@itc.nl

Kuras, O.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, UK

Lagacherie, P.
INRA Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), 2 place Viala 34060 Montpellier cedex 1, France, lagache@supagro.inra.fr

Lark, R.M.

Environmetrics Group, Biomathematics and Bioinformatics Division Rothamsted Research, Harpenden Hertfordshire AL5 2JQ, UK, murray.lark@bbsrc.ac.uk

Lawley, R.

British Geological Survey, Nottingham, NG12 5GG UK, rslaw@bgs.ac.uk

Le Du-Blayo, L.

Equipe COSTEL, UMR CNRS 6554 (Littoral Environnement Télédétection et Géomatique), Université de Rennes 2, place du recteur H. le Moal, 35 043 Rennes cedex France, laurence.ledu@uhb.fr

Lemercier, B.

UMR Sol-Agro et hydrosystème-Spatialisation INRA/Agrocampus Rennes, 65 rue de Saint-Brieuc, CS 84215, F-35042 Rennes Cedex, France, Blandine.Lemercier@agrocampus-rennes.fr

Li, B.

State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, libl@lreis.ac.cn

Lima, H.N.

Universidade Federal do Amazonas – UFAM – Manaus. Av. Gen. Rodrigo Octávio Jordão Ramos – 3000, Manaus – AM, 69077-000

Machado, P.L.O.A

Embrapa Rice and Beans, Rodovia GO-462, Km 12, CEP: 73075-000 Goiania, GO, Brazil pmachado@cnpaf.embrapa.br

MacMillan, R.A.

LandMapper Environmental Solutions Inc., 7415 118 A Street NW, Edmonton, Alberta, Canada, T6G 1V4, bobmacm@telusplanet.net

Martins, G.C.

Embrapa Amazônia Ocidental, Rod AM 010, Km 29 Manaus AM, 69011– 970, Brazil.lau@cpaa.embrapa.br

Mayr, T.R.

Cranfield University, National Soil Resources Institute, Building 37, Cranfield MK43 0AL, UK, t.mayr@cranfield.ac.uk

McBratney, A.B.

Faculty of Agriculture, Food & Natural Resources, The University of Sydney, JRA McMillan Building A05, NSW 2006, Australia, a.mcbratney@usyd.edu.au

McKenzie, N.J.
CSIRO Land and Water, G.P.O. Box 1666, Canberra, ACT 2602, Australia, Neil.McKenzie@csiro.au

Mendonça-Santos, M.L.
EMBRAPA – Brazilian Agricultural Research Corporation', The National Center of Soil Research – CNPS, Rua Jardim Botânico, 1024. Jardim Botânico, Rio de Janeiro, RJ, Brazil, CEP 22460-000. mendonca@cnps.embrapa.br

Michel, K.
UMR 5600 (Ville, Environnement, Société), Laboratoire de Biogéographie, Ecole Normale Supérieure des Lettres et Sciences Humaines, France, kristell.michel@ens-lsh.fr

Minasny, B.
Faculty of Agriculture, Food & Natural Resources, The University of Sydney, JRA McMillan Building A05, NSW 2006, Australia, b.minasny@usyd.edu.au

Moore, A.C.
USDA Natural Resources Conservation Service, 157 Clark Hall Annex, P.O. Box 6301 Morgantown, WV 26506, Amanda.moore@wv.usda.gov

Naumov, A.
International Potash Institute (IPI), P.O. Box 569, CH-8810, Horgen, Switzerland, alnaumov@geogr.msu.ru

Nield, S.J.
USDA Natural Resources Conservation Service, 100 E B St., Casper, WY 82601-1962, Shawn.Nield@wy.usda.gov

Oberthür, T.
International Centre for Tropical Agricultura (CIAT), AA6713, Cali, Colombia, t.oberthur@cgiar.org

Palmer, R.C.
Fircroft, Sandy Lane, Stockton-on-the-Forest, York, Y032 9UU, UK; Cranfield University, National Soil Resources Institute, Building 37, Cranfield MK43 0AL, UK, RC.Plamer@btinternet.com, r.c.palmer@cranfield.ac.uk

Pares, J.G.
Embrapa's fellowship grant, EMBRAPA Solos – Brazilian Agricultural Research Corporation, The National Centre of Soil Research, Rua Jardim Botânico 1024. Rio de Janeiro, RJ, Brazil, CEP 22.460-000, enhancer@terra.com.br

Pavlu, L.
Department of Soil Science and Geology, Czech University of Life Sciences in Prague, CZ-165 21, Prague 6 – Suchdol, Czech Republic, Pavlu@af.czu.cz

Penizek, V.
Department of Soil Science and Geology, Czech University of Life Sciences in Prague, CZ-165 21, Prague 6 – Suchdol, Czech Republic, Penizek@af.czu.cz

Polidoro, J.C.
Embrapa Soils, Rua Jardim Botanico, 1024, Jardim Botanico, Rio de Janeiro, RJ, Brazil, CEP 22460-000, polidoro@cnps.embrapa.br

Pötter, R.O.
Embrapa Florestas, Estrada da Ribeira Km 11, CP. 319, CEP 83411-000, Colombo, PR, Brazil

Prado, R.B.
Embrapa Soils, Rua Jardim Botanico, 1024, Jardim Botanico, Rio de Janeiro, RJ, Brazil, CEP 22460-000, rachel@cnps.embrapa.br

Qin, C.
State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, qincz@lreis.ac.cn

Ramsey, R.D.
Department of Wildland Resources, Utah State University, 5230 Old Main Hill, Logan, UT 84322-5230, USA, Doug.Ramsey@usu.edu

Reuter, H.I.
European Commission, Directorate General Joint Research Centre, Land Management & Natural Hazards Unit, TP 280, 21020 Ispra (VA), Italy, Hannes.Reuter@jrc.it

Rincon-Romero, M.
Faculty of Engineering, Universidad del Valle, Cali, Colombia, maurorin@univalle.edu.co

Viscarra, Rossel, R.A.
CSIRO Land and Water, Bruce E. Butler Laboratory, GPO Box 1666, Canberra ACT 2601, Australia, raphael.viscarra-rossel@csiro.au

Rossiter, D.G.
International Institute for Geo-information Science & Earth Observation (ITC), Postbus 6, 7500 AA Enschede, The Netherlands, rossiter@itc.nl

Saby, N.
INRA Unité INFOSOL, 2163 avenue de la pomme de pin, BP 20619 Ardon, F-45166 Olivet Cedex, France, Nicolas.Saly@orleans.inra.fr

Santos, H.G.
Researcher at EMBRAPA Solos –Brazilian Agricultural Research Corporation, The National Centre of Soil Research. Rua Jardim Botânico 1024. Rio de Janeiro, RJ, Brazil, CEP 22.460-000, Brazil.humberto@cnps.embrapa.br

Saunders, A.M.
USDA Natural Resources Conservation Service, 101 Aupuni St. Ste. 229, Hilo, HI 96720-4261, Amy.Saunders@hi.usda.gov

Schaefer, C.E.G.R.
Universidade Federal de Viçosa. Avenida Peter Henry Rolfs, s/n, Campus Universitário. CEP 36570-000. VIÇOSA – MG. Brazil, carlos.schaefer@ufv.br

Scheib, A.J.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, UK

Scheib, C.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, UK

Scheurer, O.
Département Agronomie, Institut Polytechnique LaSalle Beauvais, BP 30313, 60026 Beauvais Cedex, France, Olivier.Scheurer@lasalle-beauvais.fr

Schmidt, K.
University of Tübingen, Institute of Geography, Department of Physical Geography, Tübingen, Germany, karsten.schmidt@uni-tuebingen.de

Scholten, T.
University of Tübingen, Institute of Geography, Department of Physical Geography, Tübingen, Germany, thomas.scholten@uni-tuebingen.de

Schvartz, C.
ISA Lille, Laboratoire Sols et Environnement, 48 Boulevard Vauban, F-59046 Lille Cedex, France, c.Schvartz@isa.lille.fr

Smith, B.
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG UK, b.smith@bgs.ac.uk

Stum, A.K.
USDA Natural Resources Conservation Service, 340 N 600 E, Richfield, UT 84701-2258, Alex.Stum@ut.usda.gov

Teixeira, W.G.
Embrapa Amazônia Ocidental – Rod AM 010 – Km 29 – Manaus – AM, 69011 – 970, wenceslau@cpaa.embrapa.br

Thompson, J.A.
West Virginia University, P.O. Box 6108, Morgantown, WV 26506-6108, james.thompson@mail.wvu.edu

Tornquist, C.G.
Soil Science Graduate School of the Federal University of Rio Grande do Sul (UFRGS). Av. Bento Gonçalves, 7712, Porto Alegre, RS, Brazil, 90.001-970, carlostornquist@terra.com.br

Turetta, A.P.D.
Brazilian Agricultural Research Corporation – EMBRAPA Solos, Rua Jardim Botânico, 1024. Jardim Botânico, Rio de Janeiro, RJ. Brazil, CEP 22460-000, anaturetta@cnps.embrapa.br

Valladares, G.S.
EMBRAPA National Satellite Monitoring Research Center, 303 Soldado Passarinho Avenue, Campinas, SP, Brazil, 13070–115, gustavo@cnpm.embrapa.br

Vasat, R.
Department of Soil Science and Geology, Czech University of Life Sciences in Prague, CZ-165 21, Prague 6 – Suchdol, Czech Republic, Vasat@af.czu.cz

Vieira, C.A.O.
Universidade Federal de Viçosa. Avenida Peter Henry Rolfs, s/n, Campus Universitário. CEP 36570-000. VIÇOSA – MG. Brazil, carlos.vieira@ufv.br

Walter, C.
UMR Sol-Agronomie-Spatialisation INRA/Agrocampus Rennes, 65 rue de Saint-Brieuc, CS 84215, F-35042 Rennes Cedex, France, Christian.Walter@agrocampus-rennes.fr

Weber, E.
Universidade Federal do Rio Grande do Sul, Centro de Ecologia, CP. 15007, CEP 91501-970, Porto Alegre, RS, Brasil, eweber@portoweb.com.br

Yang, L.
State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, yangl@lreis.ac.cn

Zhou, C.
State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, zhouch@lreis.ac.cn

Zhu, A.-X.
State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China; Department of Geography, University of Wisconsin, Madison, Wisconsin, USA, azhu@wisc.edu, axing@lreis.ac.cn

Part I
Introduction

Chapter 1
Digital Soil Mapping: A State of the Art

P. Lagacherie

Abstract Digital Soil Mapping (DSM) can be defined as the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables. DSM is now moving toward the operational production of soil maps thanks to a set of researches that have been carried out for the last fifteen years. These researches dealt with various topics: the production and processing of soil covariates, the collection of soil data, the development of numerical models of soil prediction, the evaluation of the quality of digital soil maps and the representation of digital soil maps. The recent advances and open questions within each of these topics are successively examined.

The emergence of DSM as a credible alternative to fulfill the increasing worldwide demand in spatial soil data is conditioned to its ability to (i) increase spatial resolutions and enlarge extents and (ii) deliver a relevant information. The former challenge requires to develop a specific spatial data infrastructure for Digital Soil Mapping, to grasp Digital Soil Mapping onto existing soil survey programs and to develop soil spatial inference systems. The latter challenge requires to map soil function and threats (and not only "primary" soil properties), to develop a framework for the accuracy assessment of DSM products and to introduce the time dimension.

1.1 Introduction

Digital Soil Mapping (DSM) can be defined as "the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables" (Lagacherie and McBratney, 2007). The formulation of this recent concept is an attempt to federate toward an operational

P. Lagacherie
INRA Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), 2 place Viala 34060 Montpellier cedex 1, France
e-mail: lagache@supagro.inra.fr

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

perspective the somewhat heteroclite set of researches on soil variability that have been undertaken these last 15 years. By gathering 80 scientists from 17 different countries, the first workshop devoted to Digital Soil Mapping, held in Montpellier in September 2004, provided the opportunity to put in common a wide range of skills and tools that have a role to play in the future of DSM: soil surveying, soil information systems, expert systems, GIS, pedometric techniques, data mining techniques and remote sensing procedures.

The state of the art of Digital Soil Mapping has been already the purpose of exhaustive reviews that were written just before (McBratney et al., 2003) and after (DSM Working group of ESBN, 2006) the Montpellier workshop. The purpose of this paper is to present a complementary view that includes the main learning and open questions on Digital Soil Mapping that emerged from the Montpellier workshop. They concern the various topics that compose Digital Soil Mapping i.e. production and processing of soil covariates, collection of soil data, development of numerical models of soil predictions, evaluation of quality of DSM outputs, representation of digital soil maps. It will then be put forward the two key challenges that Digital Soil Mapping will face within these next years: (i) increasing spatial extents and resolutions, (ii) delivering a relevant soil information.

1.2 Production and Processing of Soil Covariates

The soil covariates are the spatial data available over large areas which can be used as inputs of Digital Soil Mapping procedures. They represent the environmental factors that are recognized as governing the soil formation (e.g. parent material, relief vegetation, climate) and the spatial soil information that can be retrieved from remote sensing images or small scale soil maps of the region of interest. The key points that emerged from the papers and the discussions of the Montpellier workshop were the following:

- A great variety of environmental covariates are now used as input for Digital Soil Mapping. The first source of environment covariates is still DEM but tests of other sources were presented too. The most promising evolution is the use of new remote sensing images that provide high resolution maps of some topsoil properties: Hyperspectral images (Ben-Dor and Patkin, 1999; Madeira et al., 2007) or Gamma-Ray spectrometry (Wilford and Minty, 2007)
- Soil covariate are often simple and current environmental variables such as elevation, slope, vegetation index etc...However several pre-processing of soil covariates are also performed with a view to produce more sophisticated covariates that would represent more accurately the soil variations. Three types of pre-processing are considered: (i) derivation of soil covariates representing the spatial variability of a given pedological process (e.g. Mérot et al., 2003) (ii) identification of structuring elements of the soil cover e.g. landscape units (Dobos and Montanarella, 2007, Robbez-Masson, 2007) or regolith-catenary units (Thwaites, 2007), (iii) The decomposition of the initial image of a soil

covariate into several spatial components of decreasing spatial resolution to deal with multi-scaled soil landscape relations (Mendonça-Santos et al., 2007),
– Some authors promoted knowledge-driven approaches based on the expression of qualitative mental models of pedogenesis in an explicit form that allows to select and build the most appropriate soil covariates (Walter et al., 2007, McKenzie and Gallant, 2007). These approaches are expected to better respect the principle of parsimony and provide more robust soil predictions than the usual data-driven approaches.

With the exponential development of the spatial data infrastructures over the world, it is expected that useful spatial data related with soil variations will be made more and more cost-effective in the future. Digital Soil Mapping community has to be aware of this evolution and must develop new tools and spatial prediction techniques to take advantage of any new soil covariate that may improve soil prediction over large areas.

1.3 Collection of Soil Data

It was largely recognised in the discussions of the Montpellier workshop that a good soil dataset is a key factor to build an accurate DSM function and to evaluate the quality of its outputs. However the collection of soil data has been (and still remain) a limiting factor that can severely brake the Digital Soil Mapping progresses. To overcome this problem, three complementary ways can be explored, (i) develop optimal sampling methods (ii) use as much as possible legacy soil data and (iii) develop new soil sensors for accurate and cost-effective estimation of soil properties.

The development of optimal sampling methods was the subject of few papers in the Montpellier workshop (and in the literature). The methods presented in Montpellier aimed to optimise the coverage of the geographical space (Brus et al., 2007), the coverage of the soil covariate space (Minasny and McBratney, 2007) or both (Heuvelink et al., 2007, Dobermann and Simbahan, 2007). These methods are derived from well known statistical and geostatistical techniques. They do not take into account more sophisticated sampling criteria that are often considered in classical soil surveying e.g. the optimisation of real field costs of soil data collection (e.g. accessibility) or the use of a priori hypothesis about the expected patterns of soil variations (e.g. introduction of sampling density constraints to capture an expected short-range variability, identifications of representative transects or areas). Much work is needed to integrate these criteria too.

The use of legacy data was only evoked in one paper of the Montpellier meeting (Bui et al., 2007). However legacy soil data represent in many countries a huge reservoir of soil information that can serve as input of Digital Soil Mapping procedures or as validation sets. Two types of legacy data must be distinguished, existing soil maps and soil profiles. The former provide a continuous representation of the soil pattern. They could be used as soil covariates in case of a small scale soil map covering the studied region (Mayr and Palmer, 2007) or as a source for calibrating Digital Soil

Mapping procedures that take into account the soil surveyor knowledge if they are sampled to be representative of a larger region (e.g. Lagacherie et al. 1995). Existing soil profiles provide local information on many soil properties at different depths. They are often used as inputs of many statistical and geostatistical procedures (Carré and Girard, 2002, Hengl et al. 2004). However the use of legacy data is hampered by a number of problems like unavailability of numeric data, lack of harmonisation and imprecision of soil descriptions, imprecise georeferencing, non optimal location of soil data etc... Beside the necessary integration of soil data in coherent spatial data infrastructures that has been undertaken by some soil agencies (Daroussin et al., 2007, Dusart, 2007, Feuerherdt, 2007, ESBN DSM working group, 2006), the Digital Soil Mapping community must adapt its models so that the intrinsic imprecision of these data can be handled properly.

With the technological impulse given by the development of precision agriculture, we can expect in the near future the wide development of more time and cost-efficient methodologies for soil observation and analysis than the classical methods of soil surveying. For example, reflectance spectrometry was proved as being a good alternative to the costly soil physical and chemical laboratory soil analysis for the estimation of a large range of soil properties (Shepherd and Walsh, 2002, Viscarra Rossel et al. 2006). The increasing use of such methodologies will surely boost the development of Digital Soil Mapping since it may provide a large amount of good quality and inexpensive soil data for imputing the DSM procedures.

1.4 Development of Numerical Models of Soil Prediction

Numerical models of soil prediction are the functions that predict soil classes or soil properties from soil covariates (class scorpan functions and property scorpan function) or from available soil data (soil allocation functions, class-to-property functions and pedotransfer functions) (Lagacherie and McBratney, 2007). Most of these functions are calibrated with soil samplings over which both the inputs and the outputs of the functions are known.

The development of numerical models of soil prediction is the most developed topic of Digital Soil Mapping as shown by the reviews on Digital Soil Mapping. The main points that emerged from the Montpellier workshop are listed hereafter.

– The Digital Soil Mapping functions that were presented were almost equally distributed between class scorpan functions and property scorpan functions. Two contributions only (Boruvka and Penizek, 2007; Albrecht et al. 2007) dealt with another type of function, namely soil allocation functions.
– A distinction appeared clearly between "quantitative soil surveyor approaches" that capture the knowledge of soil surveyors to build the DSM functions (Walter et al. 2007, McKenzie and Gallant, 2007, Mayr and Palmer, 2007, Cole and Boettinger, 2007, Saunders and Boettinger, 2007) and "classical pedometrician approaches" that apply existing data mining or geostatistical models to the set of available data. Few "hybrid" approaches were presented (Walter et al. 2007).

– Most of the presented "classical pedometrician approaches" were still based on non-spatial statistical models: Multilinear regression (Behrens and Scholten, 2007, Odeh et al. 2007, Dobos et al. 2007), Generalised linear models (Howell et al. 2007), discriminant analysis (Taylor and Odeh, 2007) decision trees (Behrens and Scholten, 2007, Hollingsworth et al. 2007), specific data mining algorithms including or neural networks (Behrens and Scholten, 2007). A few examples illustrated the use of geostatistical models, i.e. ordinary kriging (Odeh et al. 2007, Baxter et al. 2007, Bernoux et al. 2007), multiple indicator kriging (Taylor and Odeh, 2007) and regression kriging (Odeh et al. 2007, Dobos et al. 2007). The application of these latter models to large areas appears still limited by soil samplings that are not dense enough to capture the patterns of the soil variations. Further examples dealing with large spatial extents are needed to identify in which pedological contexts and for which data configurations these models may outclass non-spatial ones.
– None of the proposed functions dealt with multiscale variations of the soil cover although multiscale variations are often observed at the large spatial extents considered in Digital Soil Mapping (Lagacherie and McBratney, 2007, Chaplot and Walter, 2007). Theoretical models dealing with these variations have been presented (Lark, 2007). However their use for building effective DSM functions are still not straightforward. More work is certainly needed on this topic.

The development of numerical models of soil prediction has reached the point where a lot of functions are now available. However there is not enough comparisons of their performances over the variety of situations that may be practically encountered in Digital Soil Mapping. These comparisons would provide the necessary expertise in selecting the most appropriate function given the objective of the study (spatial extent, resolution) the data context (soil covariates, soil sampling, available pedological knowledge and skills), and the nature of the soil variation. Beside, more sophisticated Digital Soil Mapping functions than the one presented are certainly required (i) to enhance the synergy between soil knowledge and pedometric models (ii) to tackle multiscaled variations of the soil cover and (iii) to deal with evolutions of soil properties with time that are now monitored extensively.

1.5 Evaluation of the Quality of Digital Soil Maps

Digital soil maps are an essential part of a soil assessment framework which supports soil-related decision- and policy-making and therefore it is of crucial importance that DSM products are of known quality. Questions on quality and accuracy assessment of DSM products arose in almost every sessions of the Montpellier workshop which demonstrates the importance of this topic. Meanwhile, it was stated that the validation phase is rarely funded and undertaken in applied DSM projects. Furthermore, the small number of contributions devoted specifically to this subject in Montpellier is a symptom of the immaturity of our collective experience. Therefore defining and applying a common accuracy assessment framework is probably the greatest challenge of Digital Soil Mapping on its way toward practical applications. Some points must be taken in consideration:

- It is important to define precisely which type of quality indicator are needed (Finke, 2007). Generic indicators from the map maker point of view are well known (e.g. attribute accuracy, positional uncertainty, logical consistency,...). However quality must be also assessed from the user point of view i.e. by taking into account how the Digital Soil Map is effectively used.
- The advantage of Digital Soil Mapping approaches over classical surveys is that they can predict the quality of the outputs (e.g. error variance for kriging, coefficient of determination for regression, ...). However it is important to realise that (i) these quality predictions are based on model assumptions that may not hold in reality and (ii) that they are often calculated from the same data that were used to build the DSM function. Validations from independent samples are thus highly preferable whenever possible (DSM working group of ESBN, 2006).
- We must be especially careful to validate like data at like scales, e.g. point-scale validation for point-scale prediction, area-scale validation for area-scale prediction. This means that new error metrics must be proposed to compare areal soil predictions with the "kind of truth" that represents a soil map used as validation data (e.g. Greve and Greve, 2007).

1.6 Representation of Digital Soil Maps

Visualization techniques are expected to provide the indispensable insights into the complexity of the soil cover that are required by both digital soil surveyors and end-users. Before the computational era, the choropleth map was probably the best way to summarize on a sheet of paper the complex information resulting from a soil survey. The appearance of computerized techniques dealing with spatial data have dramatically modified this situation by allowing visualization of more sophisticated conceptualizations of the soil cover such as the ones handled in Digital Soil Mapping (Burrough, 2007).

New progresses can be made thanks to the development of scientific visualization tools and virtual reality techniques. The interactivity and the availability of multiple visualizations available from such techniques can stimulate our understanding of complex environment systems, and can disseminate Digital Soil Mapping outputs to a large array of potential end-users (Grunwald et al. 2007). A more complete review of the possibilities given by these new tools is available in the report of the ESBN working group on Digital Soil Mapping (2006).

1.7 Challenges for the Next Years

1.7.1 Increasing Spatial Resolution and Enlarging Extent

Within the two past decades, Digital Soil Mapping has been moving from the research phase to productions of map over regions, catchments and whole countries.

By extrapolating this past evolution, we predicted that a global digital soil properties map at 100 m resolution could be available in 2027 (Lagacherie and McBratney, 2007). Although this perspective is by far more favorable than the classical soil surveys one, it is seen as still insufficient to face the huge demand in soil information. A realistic challenge could be to obtain the same map within the next decade. In this perspective the following points will have to be taken in consideration:

– ***Develop a specific spatial data infrastructure for Digital Soil Mapping***. Research works clearly demonstrated that Digital Soil Mapping can be boosted in the near future by a set of innovative data such as precise and high resolution DEM (e.g. LIDAR), remote sensors mapping soil surface properties, e.g. hyperspectral images or gamma-radiometric (see also Section 2.2) or proxy sensors measuring local soil properties (e.g. Spectrometry, see Chapter 13). These data should be made extensively available as soon as possible to the DSM community.
– ***Grasp Digital Soil Mapping onto existing soil survey programs*** Digital Soil Mapping is a recent scientific field that has been sometimes presented and perceived as a radical alternative to conventional soil survey techniques for building soil spatial infrastructures. However for many organizational and scientific reasons, it is unrealistic and unproductive that DSM applications could be undertaken independently of the existing soil survey programs and knowledges. Similarly, new soil survey programs cannot be thought without any Digital Soil Mapping input because the conventional soil survey methods are simply too slow and expensive to fulfill the huge worldwide demand of soil information. A synergy is thus to be found, including the systematic use of legacy data (soil profiles and soil maps) in DSM procedures (see Chapters 25 and 27), the incorporation of local soil knowledges in DSM models (Walter et al. 2007), and the extensive use of DSM approaches in current soil survey programs (see Chapters 4 and 6).
– ***Develop soil spatial inference systems***. Dealing with large extents and fine resolutions increases the risk of facing with complex patterns of soil variation. It therefore seems very unrealistic to gamble on the emergence of a kind of best DSM model that would run properly whatever the study area and the data configuration. There is a need of a Spatial Soil Inference System (Lagacherie and McBratney, 2007) that would make several DSM models cooperate to produce the best possible soil map of a given area i.e. the optimal response to a given soil-user request considering a given set of available input data and a given pedological context.

1.7.2 Delivering a Relevant Soil Information

In their review of soil science developments over the past 40 years, Mermut and Eswaran, (2000) underlined that the demands from the society to the soil science community has dramatically increased which has led to the emergence of new areas of interest such as land and soil quality, recognition of problems of land degradation and desertification, cycling of bio-geochemicals, soil pollution assessment and monitoring. These have been added to the old topics traditionally investigated by soil science such as soil fertility assessment or land management. This phenomenon has been amplified by the paramount developments of soil databases (Rossiter, 2004) which has attracted new users of soil information. Consequently, the output of soil survey, and of Digital Soil Mapping, is becoming less and less under the control of soil surveyors themselves. Therefore, the soil properties that must be predicted are no longer the few selected by soil surveyors because of their relative accessibility, but can be imposed by other specialists for their own models (see Chapter 3). In this perspective the following objectives will have to be considered:

– ***Map soil function and threats.*** The predicted soil properties provided by DSM models must be coupled with other environmental data (e.g. relief, climate,...) and process-based models to produce more comprehensive outputs for soil users (see Chapters 21 and 26). Several examples of such outputs (mapping land qualities, leaching soil potential, erosion risk, heavy metal contaminations,...). were provided in the report of the ESBN working group on Digital Soil Mapping (2006). Note that a prerequisite is the development of pedotransfer functions to estimate uneasy-to measure soil properties.
– ***Develop a framework for the accuracy assessment of DSM products***. Delivering DSM products without any estimation of their accuracy may lead to erroneous decisions. It is thus important to convert the body of knowledge on accuracy assessment that exists in the literature (Finke, 2007, ESBN DSMWG, 2006) into a set of comprehensive procedures that can be routinely applied whatever the data configurations and the DSM outputs (see also Chapters 7 and 11).
– ***Introduce the time dimension.*** Time has long been recognized as a major factor of soil variations. This is all the more true that human activity may induce dramatic changes of the soil properties e.g. heavy metal contamination around urbanized areas, decrease of carbon content due to climate and/or land use changes. Monitoring systems have been set to follow these evolutions in several parts of the world (e.g. Saby et al. 2007) and they begin to be exploited (Chapters 22 and 23). However these systems will have to be completed by spatialised models of soil evolution in view of extrapolating their results. Although the development of such models is still in infancy (Minasny and McBratney, 2007), they also appears as promising tools to obtain more generic and robust Digital Soil Mapping models.

1.8 Conclusion

Digital Soil Mapping is on its way toward maturity. A set of recent papers and workshops contributed to clarify the concepts, to identify some potentially useful data layers and predictive methodologies issued from a dispersed set of research, and to show the first applications of these. As shown by this paper, much more has still to be done. However I am convinced, with others, that DSM give us a chance to satisfying the worldwide demand of soil information.

References

Albrecht, C., Huwe, B., Jahn, R., 2007. Chapter 31. Comparison of approaches for automated soil identification. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 425–436.

Baxter, S., Oliver, M., Archer, J.R., 2007. Chapter 35. The spatial distribution and variation of available P in agricultural topsoil in England & Wales in 1971, 1981, 1991 and 2001. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 477–486.

Behrens, T., Scholten, T., 2007. Chapter 25. A comparison of data-mining techniques in predictive soil mapping. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 353–364.

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. International Journal of Remote Sensing, 23(6), 1043–1062.

Bernoux, M., Arrouays, D., Cerri, C.E.P., Cerri C.C., 2007. Chapter 37. Regional organic carbon storage maps of the western Brazilian Amazon based on prior soil maps and geostatistical interpolation. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 497–506.

Boruvka, L., Penizek, V., 2007. Chapter 30. A test of an artificial neural network allocation procedure using the Czech Soil Survey of Agricultural Land data. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 415–424.

Brus, D.J., de Gruijter, J.J., Van Groenigen, J.W., 2007. Chapter 14. Designing spatial coverage samples using the k-means clustering algorithm. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 183–192.

Bui, E., Simon, D., Schoknecht, N., Payne, A., 2007. Chapter 15. Adequate prior sampling is everything: lessons from the Ord River basin, Australia. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 193–206.

Burrough, P., 2007. Chapter 41. The display of digital soil data, 1976–2004. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 555–570.

Carré, F., Girard, M.C., 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110(3,4), 241–263.

Chaplot, V., Walter, C., 2007. Chapter 38. Improving the spatial prediction of soils at local and regional levels through a better understanding of soil-landscape relationships: soil hydromorphy in the Armorican Massif of Western France. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 507–522.

Cole, N., Boettinger, J., 2007. Chapter 27. Pedogenic understanding raster classification methodology for mapping soils, Powder River Basin, Wyoming. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 377–388.

Daroussin, J., King, D., Le Bas, C., Vrscaj, B., Montanarella, L., 2007. Chapter 4. The Soil Geographical Database of Eurasia at scale 1:1 000 000: history and perspective in Digital Soil Mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 55–66.

DSM ESBN Working group 2006. Digital Soil Mapping as the support of Production of Functional Soil Maps. Dobos, E., Carré, F., Hengl, T., Reuter, H.I., Tóth, G. (eds). EUR 22123 EN, 68 pp. Office for Official Publications of the European Communities, Luxemburg.

Dobermann, A., Simbahan G.C., 2007. Chapter 13. Methodology for using secondary information in sampling optimisation for making fine-resolution maps of soil organic carbon. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 167–182.

Dobos, E., Montanarella, L., 2007. Chapter 9. The development of a quantitative procedure for soilscape delineation using digital elevation data for Europe. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 107–118.

Dobos, E., Micheli, E., Montanarella, L., 2007. Chapter 36. The population of a 500-m resolution soil organic matter spatial information system for Hungary. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 487–496.

Dusart, J., 2007 Chapter 6. Adapting soil mapping practices to the proposed EU INSPIRE directive. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 77–86.

Feuerherdt, C., Robinson, N., Williams, S., 2007. Chapter 7. Storage, maintenance and extraction of digital soil data. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 87–96.

Finke, P., 2007. Chapter 39. Quality assessment of digital soil maps: producers and users perspectives. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 523–542.

Greve, M.H., Greve, M.B., 2007. Chapter 40. Using soil covariates to evaluate and represent the fuzziness of soil map boundaries. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 543–554.

Grunwald, S., Ramasundaram, V., Comerford, N.V., Bliss, C.M., 2007. Chapter 42. Are current scientific visualization and virtual reality techniques capable to represent real soil-landscapes? In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 571–580.

Hengl, T., Heuvelink, G., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120(1–2), 75–93.

Heuvelink, G.B.M., Brus, D.J., de Gruijter J.J., 2007. Chapter 11. Optimization of sample configurations for digital mapping of soil properties with universal kriging. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 137–152.

Hollingsworth, I., Bui, E., Odeh, I., Ludwig, J., McLeod, P., 2007. Chapter 29. Rule-based land-unit mapping of the Tiwi Islands, Northern Territory, Australia. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 401–414.

Howell, D., Kim, Y., Haydu-Houdeshell, C., Clemmer, P., Almaraz, R., Ballmer, M., 2007. Chapter 34. Fitting soil property spatial distribution models in the Mojave Desert for Digital Soil Mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 465–476.

Lagacherie, P., Legros, J.P., Burrough, P., 1995. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. Geoderma 65, 283–301.

Lagacherie,P., McBratney, A.B., 2007. Chapter 1. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 3–24.

Lark, R.M., 2007. Chapter 23. Decomposing digital soil information by spatial scale. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 301–326.

Madeira Netto, J.S., Robbez-Masson, J.-M., Martins, E., 2007. Chapter 17. Visible-NIR hyperspectal imagery for discriminating soil types in the la Peyne watershed, France. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 219–234.

Mayr, T., Palmer, R., 2007. Chapter 26. Digital Soil Mapping: An England & Wales perspective. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 365–376.

Mendonça-Santos, M.L., McBratney, A.B., Minasny, B., 2007. Chapter 21. Soil prediction with spatially decomposed environmental factors. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 269–280.

Mermut A.R., Eswaran H., 2000. Some major developments in soil science since the mid-1960s Geoderma 100 (3–4), 403–426.

Mérot, P., Squividant, H., Aurousseau, P., Hefting, P., Burt, T., Maitre, V., Kruk, M., Butturini, A., Thenail, C., Viaud, V. 2003. Testing a climato-topographic index for predicting wetlands distribution along an European climate gradient. Ecological Modelling 163(1–2), 51–71.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On Digital Soil Mapping. Geoderma 117, 3–52.

McKenzie, N., Gallant, J., 2007. Chapter 24. Digital Soil Mapping with improved environmental predictors and models of pedogenesis. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 327–352

Minasny, B., McBratney, A.B., 2007. Mechanistic soil–landscape modelling as an approach to developing pedogenetic classifications. Geoderma 133(1–2), 138–149.

Minasny, B., McBratney, A.B., 2007. Chapter 12. Latin hypercube sampling as a tool for Digital Soil Mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 153–166.

Odeh, I.O.A., Crawford, M., McBratney, A.B., 2007. Chapter 32. Digital mapping of soil attributes for regional and catchment modelling, using covariates, and statistical and geostatistical techniques. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, volume 31. Elsevier, Amsterdam, pp. 437–454.

Robbez-Masson, J.M., 2007. Chapter 19. Producing dynamic cartographic sketches of soilscapes by contextual image processing in order to improve efficiency of pedological survey. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 245–256.

Rossiter, D.G., 2004. Digital soil resource inventories: status and prospects. Soil Use & Management 20(3), 296–301.

Saby, N., Arrouays, D., Boulonne, L., Jolivet,C., Pochot, A. 2007. Geostatistical assessment of Pb in soil around Paris, France Science of The Total Environment 367(1,15), 212–221.

Saunders, A., Boettinger, J., 2007. Chapter 28. Incorporating Classification Trees into a Pedogenic Understanding Raster Classification Methodology, Green River Basin, Wyoming. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 389–400.

Shepherd K.D., Walsh M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society America Journal 66, 988–998.
Taylor, J.A., Odeh, I.O.A., 2007. Chapter 33. Comparing discriminant analysis with binomial logistic regression, regression kriging and multi-indicator kriging for mapping salinity risk in northwest New South Wales, Australia. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 455–464.
Thwaites, R.N., 2007. Chapter 20. Conceptual and digital soil-landscape mapping using Regolith-Catenary Units. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 257–268.
Viscarra Rossel, R.A., Walwoort, D.J.J., Mc Bratney, A.B., Janik, L.K., Skjemstad., J.O. 2006. Visible, near infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.
Walter, C., Lagacherie, P., Follain, S., 2007. Chapter 22. Integrating pedological knowledge into Digital Soil Mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 281–300.
Wilford, J., Minty, B., 2007. Chapter 16. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in soil science, vol. 31. Elsevier, Amsterdam, pp. 207–218.

Chapter 2
Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures

Budiman Minasny, Alex. B. McBratney and R. Murray Lark

Abstract This chapter reviews some hardware and software for digital soil mapping. By *hardware* we mean various kinds of sensor and instrument which can give us better soil and *scorpan* data, and by *software* we mean mathematical or statistical models that can improve our spatial predictions. There are two approaches for the development of hardware for acquiring soil information: the top-down, and the bottom-up. The top-down approach asks which technologies are available and which variables can we measure that are related to *scorpan* factors. The bottom-up approach starts from a problem that we systematically analyse so as to identify the information that is needed to solve it. We then tackle the technical problems of collecting this information, and only at the end move to developing the field technology. We evaluate various software approaches to improve spatial prediction of soil properties or soil classes. Finally, the implication of using data-mining tools for the production of digital soil maps is discussed.

2.1 Introduction

There is a global need for quantitative soil information for environmental monitoring and modelling. One response to this demand is digital soil mapping, where soil maps are produced digitally based on environmental variables (McBratney et al., 2003). The environmental or so-called *scorpan* factors (*scorpan* is a mnemonic for factors for prediction of soil attributes: soil, climate, organisms, parent materials, age, and spatial position proposed by McBratney et al., 2003) derived from various sources (digital elevation models, remote sensing images, existing soil maps), and available in digital form, are used to generate soil information in the form of a database where most of the information consists of predictions that are statistically optimal. Fig. 2.1 summarises the process of digital soil mapping, where geo-referenced soil

B. Minasny
Faculty of Agriculture, Food & Natural Resources, The University of Sydney, JRA McMillan Building A05, NSW 2006, Australia
e-mail: b.minasny@usyd.edu.au

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

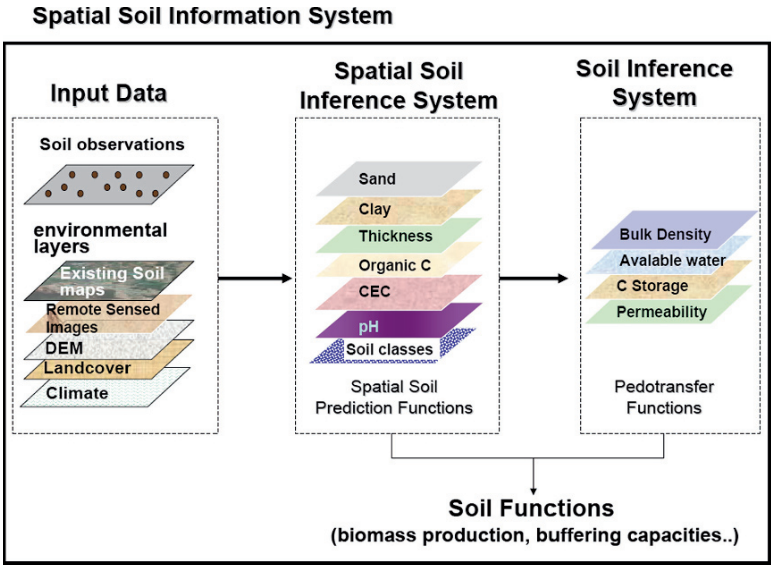


Fig. 2.1 Principles of digital soil mapping

observations coupled with environmental variables form the input data. Under a spatial soil inference system, soil properties over the whole area can be predicted and mapped using spatial soil prediction functions (such as regression, kriging, or a combination of both). This prediction is based on correlations between the environmental variables and soil attributes, as well as the spatial autocorrelation of the attributes themselves. These spatially inferred soil properties can be used to predict more difficult-to-measure functional soil properties, such as field capacity, available water capacity using pedotransfer functions under soil inference system. All of the predicted soil properties can be used to evaluate soil functions. The details on digital soil mapping is covered in Chapter 1.

In this chapter, we review some technologies for digital soil mapping, and consider in particularly the case of countries with sparse data infrastructure. Technology can be hardware or software. By *hardware* we mean various kinds of sensor and instrument which can give better soil and *scorpan* data, and by *software* we mean mathematical or statistical models which can improve spatial predictions. Firstly, various *hardware* and *software* that can be used to obtain soil information are reviewed, followed by a discussion on the implication of using data-mining tools for the production of digital soil maps.

2.2 Hardware

Figure 2.2 shows the electromagnetic spectrum, highlighting those parts where soil information can be obtained. Matter emits electromagnetic radiation in different parts of the spectrum, and this radiation can be measured by different types of

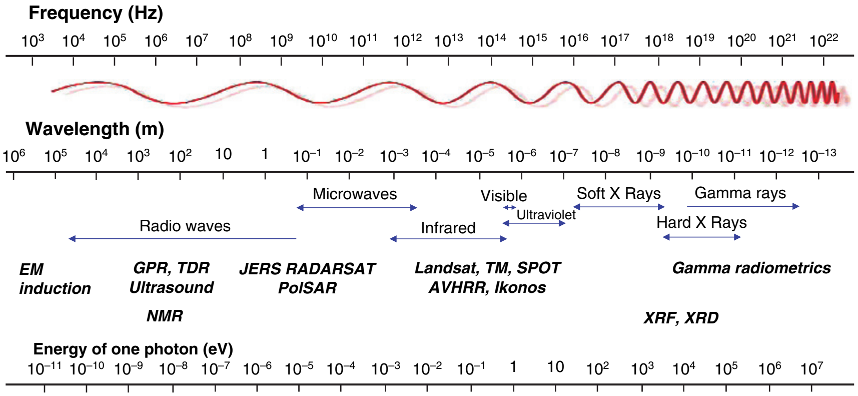


Fig. 2.2 Electromagnetic spectrum, highlighting instruments for obtaining soil information

spectroscopy depending on the wavelength. It provides a basis for remote sensing of the properties of matter. A sensing system might measure the radiation emitted by an object after the object has itself been irradiated. Two examples of this are the optical remote sensing systems that measure the solar radiation reflected by an object, and the synthetic aperture radar systems (SAR) that measure deliberately long-wave radiation backscattered by an object. Alternatively, it may be possible to measure radiation emitted by an object because of its temperature (emitted in thermal infra-red frequencies) or because of radioactive decay (decay of uranium thorium and potassium isotopes are widely measured by “passive” gamma radiometry in geophysics). The electromagnetic radiation emitted from an object will therefore depend on its physico-chemical properties, some of which are of direct interest in soil studies such as temperature, mineralogy, organic content, physical structure, or the chlorophyll content of the vegetation.
Some examples of the instruments used in soil science, grouped by their wavelengths are:

– Radiowaves, wavelengths about 300,000–0.3 m (Fig. 2.3).
  This includes sensors in the low frequency (about 10 kHz), e.g. electromagnetic induction (EMI), and high frequency (about 100 MHz), e.g. ground penetrating radar (GPR), time domain reflectometer (TDR), frequency domain moisture sensors (FD), which detect variations in soil dielectric constant. Radiowave or microwave radiation can be applied in the presence of a magnetic field to excite nuclear and electron magnetic resonances (e.g., Nuclear Magnetic Resonance (NMR) at radiowaves, and Electron Spin Resonance (ESR) at microwaves) that are sensitive to the surrounding molecules, from which information about local bonding of atoms can be obtained (O’Day, 1999). Another application in the radio frequency band is transmission of information via wireless sensor networks (Wang et al., 2006).

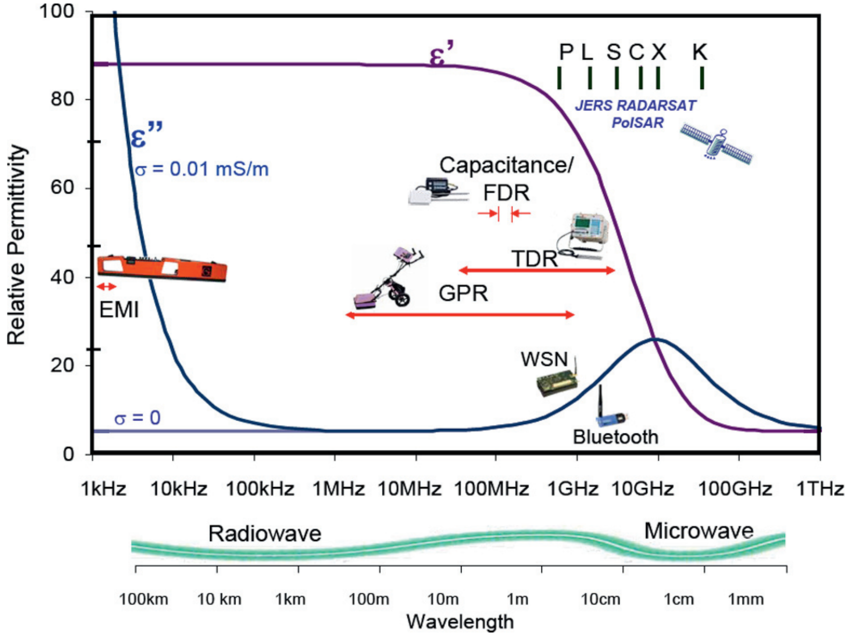


Fig. 2.3 Real ε' and imaginary ε'' permittivity as a function of frequency in the radio-microwave spectrum

- Microwaves, wavelengths 0.3–0.0003 m (Fig. 2.3).
  Information can be obtained by radar especially if there is a contrast change in dielectric constant. Passive and active microwave imaging systems have been built and experimented with for imaging. The application of radar to active imaging systems has resulted in SAR (Synthetic Aperture RADAR). SAR has been used in mapping soil with rough or impenetrable terrain such as the Amazon Basin (EMBRAPA, 1981).
- Infrared (0.7–300 μm wavelength), visible light (about 400 nm to about 700 nm), ultraviolet (3–400 nm).
  Soil scientists in the field mostly used the visible light spectrum through the Munsell soil colour chart to determine soil colour and the presence of some pedological features. Infrared energy radiation may cause vibrational excitation of covalently bonded atoms, and give rise to absorption spectra. Lasers, which emit radiation at a single frequency, are commonly used as a radiation source in the ultraviolet, visible, and infrared regions. Absorption spectra of compounds are a unique reflection of their molecular structure. Recent work has shown how diffuse reflectance spectroscopy in the visual-near-infrared and mid-infrared regions can be used to predict various soil physical, and chemical properties (Viscarra Rossel et al., 2006). Chapter 13 shows the use of diffuse reflectance spectroscopy for rapid acquisition of soil information.

– X-rays (wavelengths 10–0.01 nm).
  Radiation in this frequency band causes atomic level excitations (absorption and emission of radiation) that are used to probe bonding states (core or valence level) around an atom. This is mainly used in the laboratory for characterisation of mineral structure, such as X-ray diffraction, XAFS (x-ray absorption fine-structure spectroscopy) for probing physical and chemical structure of minerals at an atomic scale.
– Gamma rays (wavelengths shorter than about 0.01 nm).
  All soils contain concentrations of naturally occurring radioisotopes which can decay to produce gamma rays. Gamma-ray spectrometry measures the natural emission of gamma-ray radiation of the earth's surface, it estimates the abundances of potassium (^{40}K), thorium (^{232}Th) and uranium (^{238}U) (Cook et al., 1996; Wilford et al., 1997). Gamma-ray spectroscopy mapping can be conducted using remote sensing (low flying aircraft, helicopter), or proximal sensing (vehicle mounted).

There are two general models for the development of "technologies" for acquiring soil information. These are:

1. Top-Down
   We first seek variables that can be measured that might be related to *scorpan* factors or target variables by looking to other disciplines such as geophysics, and explore the possibilities of extracting information from these technologies that is pertinent to practical problems. A classical example (where this has been reasonably successful) is EMI technology and a more recent example is the use of gamma radiometrics, which has been used in geological prospecting for over 30 years to detect anomalies associated with exploitable ore deposits.
2. Bottom-up
   We start from a problem that we analyse systematically so as to identify the information that is really needed to solve it. We then tackle the general technical problems of collecting this information, and only at the end move to developing the field technology. An excellent example of this is the development of the on-the-go pH and lime-requirement sensing system at Australian Centre for Precision Agriculture. The approach is now mostly proximal sensors developed for precision agriculture, to obtain high resolution soil information (Adamchuk et al., 2004) such as soil moisture, mechanical resistance, organic matter content, soil texture, and nutrients concentration.

Both approaches have been successful, although the bottom-up approach is more intellectually satisfying. Currently, the bottom-up approach is mainly for developing proximal sensors for high resolution digital soil mapping. High resolution soil data are often needed in areas where the land has high value or poses high risk. The applications are precision agriculture, assessment of contamination sites, and in urban and industrial areas where land is valuable. It falls into soil mapping category D1 of McBratney et al. (2003) with a pixel resolution of (1 m $\times$ 1 m) to (10 m × 10 m). A bottom-up approach also works at coarser-resolution studies for D3 and D4, catchment or environmental mapping. While it is not soil mapping, the use of remote sensor imagery to parameterize soil-vegetation-atmosphere transfer models illustrates how process-based understanding might be used to extract soil information from remote sensor data at coarse scales (e.g. Verhoef, 2004).

2.2.1 Data Sources for Scorpan

The *scorpan* factors can be obtained from various sensors, either remotely or proximally sensed. Remote sensing for soil properties is reviewed by Ben-Dor (2002), while proximal sensing is given by Adamchuk et al. (2004). Here we list some sensors, based on their platform, that are commonly used for digital soil mapping:

Satellite based (Fig. 2.4):

- Hyperion http://eo1.usgs.gov/hyperion.php
  The Hyperion from EO-1 satellite provides a high resolution hyperspectral imager capable of resolving 220 spectral bands from 400 to 2500 nm with a 30 m spatial resolution, and image swath width 7.5 km. Hyperspectral images measure reflected radiation at a series of narrow and contiguous wavelength bands. Its use for digital soil mapping is still limited (Datt et al., 2003) and can be challenging as noise of the spectra and the influence of vegetation.
- Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) http://asterweb.jpl.nasa.gov/

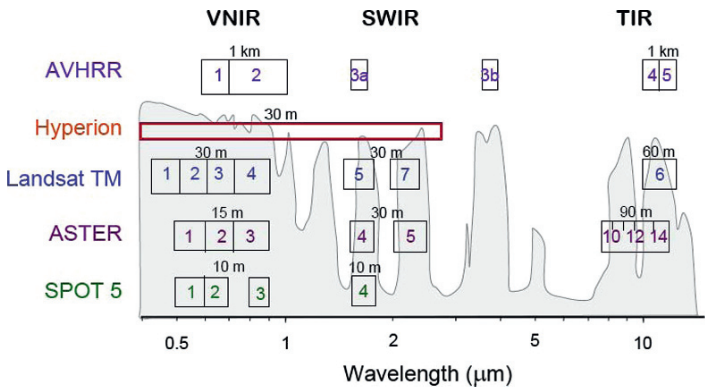


Fig. 2.4 Satellite based remote sensing instruments as a function of wavelengths. The gray curves represents atmospheric electromagnetic opacity

ASTER is a multispectral imaging system (Yamaguchi et al., 1998). Multispectral imagers measure radiation reflected from a surface at a few wide, separated wavelength bands. ASTER measures visible reflected radiation in three spectral bands (VNIR between 0.52 and 0.86 μm, with 15-m spatial resolution), and infrared reflected radiation in six spectral bands (SWIR between 1.6 and 2.43 μm, with 30-m spatial resolution). In addition, ASTER records the data in band 3B (0.76–0.86 μm) with a backward looking that enables the calculation of digital elevation model (DEM). ASTER also receives emitted radiation in five spectral bands (TIR between 8.125 and 11.65 μm, with 90-m spatial resolution). It has been used for mapping geological units (Gomez et al., 2004), and areas of degraded land (Chikhaoui et al., 2006).

– Landsat TM, and Enhanced Thematic Mapper Plus (ETM+)
  http://landsat7.usgs .gov/
  The Enhanced Thematic Mapper Plus (ETM+) is a multispectral scanning radiometer that is carried on board the Landsat 7 satellite. It provides images with spatial resolution of 30 m for the visible and near-infrared, and 60 m for the thermal infrared, and 15 m for the panchromatic. Landsat has been used most often in digital soil mapping. Chapter 16 and Cole and Boettinger (2007) discussed its practical use for digital soil mapping. Chapter 22 illustrates its application for land-use mapping.
– Satellites Pour l'Observation de la Terre or Earth-observing Satellites (SPOT)
  http://www.spot.com
  SPOT provides high-resolution multispectral images with resolution of 10 m in the visible and near infra-red (0.50–0.89 μm), and 20 m in the short wave infrared (1.58–1.75 μm). Barnes and Baker (2000) investigated its use for mapping soil texture class.
– Advanced Very High Resolution Radiometer (AVHRR) http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html
  The AVHRR provides four to six bands of multispectral images (visible red, near infra-red, short-wave infra-red, and thermal infra-red) with 1.1 km resolution from the NOAA polar-orbiting satellite series. The AVHRR data have been collected to monitor global change information, however the data can be used as a cost-effective way of estimating soil properties at regional level (Odeh and McBratney, 2000). Its use is illustrated in Chapter 21.
– Moderate Resolution Imaging Spectroradiometer (MODIS)
  (http://modis.gsfc.nasa.gov/)
  MODIS is an instrument aboard the Terra and Aqua satellites. Terra's orbit around the Earth passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1–2 days, acquiring data in 36 spectral bands at a resolution of 250 m (620–876 nm), 500 m (459–2155 nm), and 1000 m (405–14385 nm). Its use mainly for monitoring vegetation activity via NDVI (Huete et al., 1994). Tsvetsinskaya et al. (2002) used MODIS data to derive surface albedo for the arid areas of Northern Africa and

the Arabian peninsula. This enabled them to relate the surface albedo statistics to FAO soil groups. See also Chapter 30.

Airborne based:

– HyMap^{TM} (Hyperspectral Mapping) http://www.intspec.com/
  is an airborne imaging VNIR-SWIR spectrometer, with 450–2500 nm spectral coverage, 128 spectral bands of 10–20 nm bandwiths. Examples of its use for mapping soil are Madeira Netto et al. (2007) and Selige et al. (2006).
– AVIRIS (Airborne Visible Infrared Imaging Spectrometer) http://aviris.jpl.nasa.gov/ is an airborne imaging instrument producing 224 spectral bands ranging from 400 to 2500 nm, with a spatial resolution of 20 m. Palacios-Orueta and Ustin (1996) showed that AVIRIS spectra can be used to discriminate between soil types.
– Airborne gamma radiometrics
  Variations of gamma radiation has been found to correspond with the distribution of soil-forming materials over the landscape (Cook et al., 1996).
– Aerial photography
  This technique, providing images in the visual light, is still being used in soil surveys and with interpretation is used to generate soil maps (Bie and Beckett, 1973).

Proximal, ground-based:

– Electrical magnetic induction (EMI) (http://www.geonics.com/) or electrical resistance measurement. These instruments measure the bulk soil electrical conductivity, it has been successful for high resolution digital soil mapping for properties such as clay and water content (Corwin and Lesch, 2005).
– Gamma radiometrics
  Gamma-ray spectrometers can measure an energy spectrum ranging from 0 to 3 MeV. The value of gamma-ray spectrometry lies due to the different rock types contain varying amounts of radioisotopes of K, U and Th. Ground-based gamma-ray spectrometers have been used for mapping soil properties (Viscarra Rossel et al., 2007; Wong and Harper, 1999).

2.2.2 Better Soil Data

By *better soil data* we mean data obtained more efficiently, so that a larger number of samples are analysed at lower costs, in less time and with higher accuracy. Spectroscopic techniques are being used and explored as possible alternatives to enhance or replace conventional laboratory methods of soil analysis. Viscarra Rossel et al. (2006) provides a review on visible, near-infrared, and mid-infrared diffuse reflectance spectroscopy for simultaneous assessment of various soil properties in laboratory, thus will not be reviewed here (see also Chapter 13). Several instruments that can be used for field measurement of soil properties:

- FieldSpec FR® spectroradiometer (Analytical Spectral Devices Inc., Boulder, Colorado http://www.asdi.com) provides diffuse reflectance spectrospcopy of soil samples from 350 to 2500 nm and sampling resolution of 1 nm. Waiser et al. (2007) found that this field instrument produces acceptable estimation of clay content at various water contents and parent materials.
- Spectral core scanners (Spectral Imaging Ltd., Finland, http://www.specim.fi) is an imaging spectrograph which produces the image of a soil core at the visible and near infrared regions. It is still under experimental testing for use in soil sensing.
- Electrochemical methods have been successfully used to directly evaluate soil fertility (Adamchuk et al., 2005). This is done by using an ion-selective electrode (glass or polymer membrane), or an ion-selective field effect transistor (ISFET). The principle involved measurement of potential difference (voltage) between sensor and reference parts of the system is related to the concentration of specific ions (e.g. H^+, NO^{3-}).

Table 2.1 gives a summary of the hardware and its use in digital soil mapping.

2.2.3 Current Problems

While the above methods are becoming available to provide better soil data in future surveys, we often have to start with legacy data. Legacy soil data arise from traditional soil survey (Bui and Moran, 2001). Methods of soil survey are generally empirical and based on a conceptual model developed by the surveyor, correlating soil with underlying geology, landforms, vegetation and air-photo interpretation. Under traditional free survey samples are located to confirm the surveyor's interpretation of the landscape and not in accordance with a statistical design. This will lead to bias in the areas that are sampled. Carré et al. (2007) examined this problem in more detail. It should be noted that, while soil observations collected in free survey pose a problem to the statistician, as soil scientists we recognize that the conceptual (or mental) models developed by soil surveyors in the past (and represented in map legends, map memoirs and map boundaries) can be highly informative. The main problem that legacy data pose us is how to ensure that this information is transferred effectively into the digital soil mapping framework (e.g. see Chapters 25 and 27).

Another set of variables that is missed from remotely or proximally-sensed instruments are soil properties at depth. Remote-sensing images only tells us the surface condition, while soil is a three-dimensional body, and in many cases the properties in the subsoil hold lots of information we wish to know. Most satellite images working in the visible and infrared regions are influenced by crop cover (see Chapter 30). The use of instruments such as the spectral core scanners may be useful to quick acquisition of soil profile information.

Table 2.1 Example of hardware used for digital soil mapping

Instrument	Platform, wavelength, measurement	Typical spatial resolution	Example of application
Radarsat	Satellitte, Radar CB and	8–100 m	Li and Chen (2005)
EM-38	Ground-based, 14.6 kHz, bulk EC	1 m^2 Depth exploration: 0.75 & 1.5 m	Corwin and Lesch (2005)
EM-31	Ground-based, 9.8 kHz, bulk EC	4 m^2 Depth exploration: 6 m	Corwin and Lesch (2005)
GPR	Ground-based, Microwave (100–1000 MHz), dielectric constant	0.25–1 m^2	Davis and Annan (1989)
Gamma radiometrics	Aerial or ground based, Gamma rays (0–3 MeV), Radiometric K, U, Th	Aerial: 100 m Ground based: 10 m	Wilford and Minty (2007) Viscarra Rossel et al. (2007)
Aerial Photography	Visible, R, G, B Channel	5–10 m	Bie and Beckett (1973)
SPOT	Satellite, Multispectral VNIR, SWIR	10 m	Agbu et al. (1990)
Landsat	Satellite, Multispectral VNIR, SWIR, TIR	30 m (VNIR, SWIR) 60 m (TIR)	Chapter 16
ASTER	Satellite, Multispectral, VNIR, SWIR, TIR	15 m (VNIR), 30 m (SWIR), 90 m (TIR)	Chapter 4
MODIS	Satellite, Multispectral 405–14385 nm	250 m (620–876 nm), 500 m (459–2155 nm), 1000 m (405–14385 nm)	Tsvetsinskaya et al. (2002) Chapter 30
Hyperion	Satellite, Hyperspectral, 400–2500 nm	30 m	
Hymap	Airborne, Hyperspectral, 450–2500 nm	5 m	Madeira Netto et al. (2007)
AVIRIS	Airborne, Hyperspectral, 400–2500 nm	20 m	Palacios-Orueta and Ustin (1996)
FieldSpec	Handheld, 350–2500 nm	$\sim 0.001\ m^2$	Waiser et al. (2007) Chapter 13

2.3 Software

2.3.1 Improved Sampling Methods

In digital soil mapping, the primary assumption is that the *scorpan* factors are correlated with soil properties. To develop an effective model, we need to sample the soil for each different type of terrain, land cover, and others based on the available *scorpan* factors. This is further complicated because the pattern of spatial dependence will change from one area to another. One method is to stratify the area into classes of similar attributes, and then random selection of samples within the class (McKenzie and Ryan, 1999). Minasny and McBratney (2006) proposed Latin hypercube sampling (LHS) on the covariates. LHS follows the idea of a Latin square where there is only one sample in each row and each column. Latin hypercube generalises this concept to an arbitrary number of dimensions, where each of the variables is represented in a fully stratified manner. Thus, this sampling scheme does not require more samples for more dimensions (variables).

2.3.2 Improved Prediction Methods

Some spatial tools that are developing and in use in the past 5 years include:

- The use of REML (residual maximum likelihood) for better estimation of spatial covariance functions (Lark et al., 2006).
- Bayesian maximum entropy (BME) (Christakos, 2000). This approach allows the incorporation of a wide variety of hard and soft data in a spatial estimation context. The data sources may come in various forms, such as intervals of values, probability density functions or physical laws.
- Kalman filter (Webster and Heuvelink, 2006), for space-time interpolation.

Non-spatial statistical models, such as (Hastie et al., 2001):

- Classification and regression trees (see Chapter 34)
- Fuzzy classification (see Chapter 26)
- Bayesian Belief Networks (see Chapter 25)
- Support vector machines
- Neural networks
- Gaussian processes (see Chapters 3 and 33)

There is an increasing use of data-mining techniques for soil prediction. Furthermore, there is also the use of combined or ensemble models, such as boosting and bagging, which generate multiple models or classifiers (for prediction or classification), and to combine the predictions from those models into a single prediction or predicted classification (e.g. Brown et al., 2006).

Data-mining tools are usually designed to explore large amounts of data, and generate models with many parameters. Soil data are usually scarce, and often users do not realise that there are more parameters in their models than there are soil data. This often happens with neural networks. Users rely on software without understanding its principles, and assume that the number of parameters is equal to the number of predictors (input variables). The actual number of parameters may be much larger since it is the sum of the number of inputs, hidden units and outputs. Similarly, the number of parameters in a classification or regression tree is the number of nodes in the tree, not merely the number of inputs. The issue of overfitting is not always adequately discussed. Table 2.2 gives a summary of the influence of data size on the selection of model and data.

Breiman (2001) identified two cultures in statistical modelling to draw conclusions from data. The first one is *data modelling* which assumes a model, attempts to infer some mechanism and assumes that the data are a stochastic realisation of the model. And the other is the *black-box* approach which uses algorithmic models (data-mining tools) and treats the mechanism by which the observations arise as unknown. Breiman (2001) argued that with increasing number of data, sometimes it is impossible to draw a mechanism from the data, and if our goal is to use data to solve problems, then we should adopt the black-box approach. Breiman contended that nature's mechanisms are generally complex and cannot be summarised by simple models (such as linear or logistic regression). Thus, inferring a mechanism can be risky, a deceptively simple picture of the inside. The downside of the black-box approach is that, because the procedure is not upfront about the model structure, we can end up in a situation where we are, perhaps without realizing it, in effect using a model with more parameters than we would have data.

Fortunately, in the *scorpan* framework, most digital soil mapping practitioners adhere to the data-model approach. But there has been growing tendency to make

Table 2.2 An assessment of the how size of data sets may effect model prediction and assessments of the accuracy of estimates (based on Maindonald, 1999)

	Small data sets (< 200 samples)	Medium size (200–1000 samples)	Large data sets (>1000 samples)
Need for efficient prediction?	Strong	Reduced	Much reduced.
What model structure is desirable?	Linear model	Nonlinear model, and some data mining tools.	Data mining tools.
How can we get internal estimates of accuracy? (i.e. prediction to a similarly drawn sample)	Estimates must be model-based, with strong assumptions. Use leave-one-out cross validation.	Alternatives are: model-based estimates, or use cross-validation, or training/test sets	Use training/test set, with random splitting of the data.
How can we get external estimates of accuracy? (i. e. prediction outside of sample population)	No good alternative to reliance on model-based assessments	Use training/test set, with purposive choice of test set.	Use training/test set, with purposive choice of test set.

more use of black-box approaches especially to generate pedotransfer functions (Pachepsky and Rawls, 2004). Perhaps we should be looking to hybrid methods that combine the advantages of black-box approaches (exploiting the power of computers to identify complex relationships between variables) with the data-modelling approach (use of scientific understanding, and controlling the modelling process to avoid overfitting). One step towards this approach is outlined in the next section.

A good example of the hybrid approach is by Bui et al. (2006) who produced digital soil properties maps across the agricultural zone on the Australian continent using classification and regression trees. First all *scorpan* factors were taken and decision tree algorithms revealed an emerging pattern. The rules in the trees were evaluated and interpreted on the basis of general principles of soil genesis.

2.3.3 Selection Methods

Often in digital soil mapping all the predictor covariates are used with a nod to *scorpan* (if we are ingenious enough most available covariates can be equated with at least one of the *scorpan* factors), and then either put all the predictors into our design matrix (data modelling), or use automated selection (stepwise regression methods, best subsets etc.), or use regression trees and neural networks (black box). Hence, we use all the available technology, including the processing power of computers.

Is this satisfactory? Consider a not untypical situation where we have six bands of remote-sensor imagery (from the Landsat Thematic Mapper), a variable derived from these (NDVI), ten variables derived from a digital elevation model (e.g. elevation, slope, aspect, topographic wetness index, distance from stream, etc.) and data from three channels of gamma radiometry (K, U, Th). We have 20 potential predictors, and use all of them in a linear regression or regression tree, not to mention polynomial functions of these variables if there is reason to believe their effect is non-linear. Is this richness of data a problem or a virtue? Under the black-box view it is a virtue. Breiman (2001) states that the more predictor variables we have, the more information that can be extracted, and also the more information is available in various combinations of the predictors. This situation worries the data modeller. When we add a new predictor to a classical linear model we almost invariably improve its fit (as measured by the residual mean square) but we add the estimation error of the coefficient to the error of our prediction at a new target site. Variable selection procedures (e.g. stepwise regression) are sometimes used to identify a subset of predictors, but these may compound the problem. The larger our data set (in terms of the number of variables) the more likely we are to find some model that gives a good description of the variations of the particular data set, but rather poor predictions at new sites. As has often been stated "If you torture the data long enough, they will confess to anything".

We do not believe that soil scientists should abandon scientific judgement and statistical scruples and embrace a pure black-box approach. However, where many data are available there is no doubt that the power of the computer offers us a way of identifying previously unknown relationships between variables. One possible approach is to use automated model selection (e.g. stepwise regression) which is constrained so as to control the false discovery rate (FDR, loosely, the expected proportion of rejected null hypotheses that are actually true). Lark et al. (2007) propose an approach to model selection that uses the FDR, in combination with an independent prior selection of variables based on expert knowledge. Expert judgement is used both to determine the size of the pool of models that is searched (matching it to the strength of evidence for the existence of good models) and to ensure that the searched subset of possible models includes those that make sense given our knowledge of the soil. In trials it was found that this method selected good predictor models, and avoided overfitting in cases where uncontrolled model selection methods fell into this particular trap.

Another example is the incorporation of the soil taxonomic distance in the data-mining algorithm for spatial prediction of soil classes (Minasny and McBratney, 2007). Current methods for predicting soil classes only consider the minimisation of the misclassification error. Soil classes at any taxonomic level have taxonomic relationships between each other, and no statistical procedure so far has been utilised to account for these relationships. Incorporating taxonomic distance between soil classes in a supervised classification routine such as decision tree, allows more meaningful prediction and effective integration of soil science knowledge.

2.4 Discussion

Hardware offers us a growing range of covariates for use in digital soil mapping. The current limitation is that remote or proximal sensors generally measure only the soil surface properties. While hardware is growing in use and producing more data, the integration with software needs to catch up. We need an approach that is more disciplined, and with a firmer grounding in statistical theory, but also one that genuinely uses the *scorpan* approach, not just treating it as a fig-leaf to justify our use of whatever covariates are available. At the same time, however, we do want to exploit the power of the computer to search for structure in large data sets.

In conclusion, there is a range of technology available that potentially can be used for more efficient measurement of soil properties, and we need to make better use of them. An approach to integrating soil science knowledge (semi-formalising the *scorpan* approach) with computational power is needed for a better digital soil mapping model.

References

Adamchuk, V.I., Hummel, J.W., Morgan, M.T., Upadhyaya, S.K., 2004. On-the-go soil sensors for precision agriculture. Computers and Electronics in Agriculture 44, 71–91.

Adamchuk, V.I., Lund, E.D., Sethuramasamyraja, B., Morgan, M.T., Dobermann, A., Marx, D.B., 2005. Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. Computers and Electronics in Agriculture 48, 272–294.

Agbu, P.A., Feherenbacker, D.F., Jansen, I.J., 1990. Soil property relationships with SPOT satellite digital data in East Central Illinois. Soil Science Society of America Journal 54, 807–812.

Barnes E.M., Baker, M.G., 2000. Multispectral data for mapping soil texture: Possibilities and limitations. Applied Engineering in Agriculture 16, 731–741.

Ben-Dor, E., 2002. Quantitative remote sensing of soil properties. Advances in Agronomy 75, 173–243.

Bie, S.W., Beckett, P.H.T., 1973. Comparison of four independent soil series by air photo interpretation, Paphos Area (Cyprus). Photogrammetria 29, 189–202.

Breiman, L., 2001. Statistical modeling: The two cultures. With discussion. Statistical Science 16, 199–231.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273–290.

Bui, E.N., Henderson, B.L., Viergever, K., 2006. Knowledge discovery from models of soil properties developed through data mining. Ecological Modelling, 191, 431–446.

Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. Geoderma 103, 79–94.

Carré, F., McBratney, A.B., Minasny, B., 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141, 1–14.

Chikhaoui, M., Bonn, F., Bokoye, A.I., Merzouk A., 2006. Comparison of the ASTER and ETM plus captors for the cartography of the deterioration of soils with the help of the LDI index. Canadian Journal of Remote Sensing 32, 74–83.

Christakos, G., 2000. Modern Spatiotemporal Geostatistics. Oxford University. Press, New York.

Cole, N.J., Boettinger, J.L., 2007. Pedogenic understanding raster classification methodology for mapping soils, Powder River Basin, Wyoming, USA. In: Lagacherie, P., McBraney, A.B., Voltz, M. (Eds.), Developments in Soil Science 31. Digital Soil Mapping: An Introductory Perspective. pp. 389–400. Elsevier, Amsterdam.

Cook, S.E., Corner, R., Groves, P.R., Grealish, G.J., 1996. Use of airborne gamma radiometric data for soil mapping. Australian Journal of Soil Research 34, 183–194.

Corwin D.L., Lesch, S.M., 2005. Apparent soil electrical conductivity measurements in agriculture. Computers and Electronics in Agriculture 46, 11–43.

Datt, B., McVicar, T.R., Van Niel, T.G., Jupp, D.L.B., Pearlman, J.S., 2003. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. IEEE Transactions on Geoscience and Remote Sensing 41, 1246–1259.

Davis, J.L., Annan, P., 1989. Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy. Geophysical Prospecting 37, 531–551.

EMBRAPA. 1981. Mapa de Solos do Brasil, escala 1:5,000,000. Serviço Nacional de Levantamento e Conservação de Solos, Rio de Janeiro.

Gomez, C., Delacourt, C., Allemand, P., Ledru, P., Wackerle, R., 2004. Using ASTER remote sensing data set for geological mapping, in Namibia. Physics and Chemistry of the Earth 30, 97–108.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics. Springer-Verlag, New York.

Huete, A., Justice, C., Liu, H., 1994. Development of vegetation and soil indices for MODIS-EOS. Remote Sensing of Environment 49, 224–234.

Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. European Journal of Soil Science 57, 787–799.

Lark. R.M., Webster, R., Bishop, T.F.A., 2007. Using expert knowledge with control of false discovery rate to select regressors for prediction of soil properties. Geoderma 138, 65–78.

Li, J.H., Chen, W.J., 2005. A rule-based method for mapping Canada's wetlands using optical, radar and DEM data. International Journal of Remote Sensing 26, 5051–5069.

Madeira Netto, J.S., Robbez-Masson, J-M., Martins, E., 2007. Visible–Nir hyperspectral imagery for discriminating soil types in the La Peyne watershed (France). In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Developments in Soil Science 31. Digital Soil Mapping: An Introductory Perspective. pp. 225–240. Elsevier, Amsterdam.

Maindonald, J., 1999. An inquiry into data mining. The New Zealand Statistical Association Conference, Victoria University of Wellington, 4–7 July 1999. http://www.maths.anu.edu.au/~johnm/dm/dmpaper.html
McBratney, A.B., Mendonça-Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.
Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences 32, 1378–1388.
Minasny, B., McBratney, A.B., 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. Geoderma 142, 285–293.
O'Day, P.A., 1999. Molecular environmental chemistry. Reviews of Geophysics 37, 249–274.
Odeh, I.O.A., McBratney, A.B., 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi valley of eastern Australia. Geoderma 97, 237–254.
Pachepsky, Y.A., Rawls, W.J. (Eds.), 2004. Development of Pedotransfer Functions in Soil Hydrology. Developments in Soil Science 30. Elsevier.
Palacios-Orueta, A., Ustin, S.L., 1996. Multivariate statistical classification of soil spectra. Remote Sensing of Environment 57, 108–118.
Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136, 235–244.
Tsvetsinskaya, E.A., Schaaf, C.B., Gao, F., Strahler, A.H., Dickinson, R.E., Zeng, X., Lucht, W., 2002. Relating MODIS-derived surface albedo to soils and rock types over Northern Africa and the Arabian peninsula. Geophysical Research Letters 29, 1353, 10.1029/2001GL014096.
Verhoef, A., 2004. Remote estimation of thermal inertia and soil heat flux for bare soil Agricultural and Forest Meteorology 123, 221–236.
Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.
Viscarra Rossel, R.A., Taylor, H.J., McBratney, A.B., 2007. Multivariate calibration of hyperspectral γ-ray energy spectra for proximal soil sensing. European Journal of Soil Science 58, 343–353.
Waiser, T.H., Morgan, C.L.S., Brown, D.J., Hallmark, C.T., 2007. In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. Soil Science Society of America Journal 71, 389–396.
Wang, N., Zhang, N., Wang, M., 2006. Wireless sensors in agriculture and food industry – Recent development and future perspective. Computers and Electronics in Agriculture 50, 1–14.
Webster, R., Heuvelink, G.B.M., 2006. The Kalman filter for the pedologist's tool kit. European Journal of Soil Science 57, 758–773.
Wilford, J.R., Bierwirth, P.N., Craig, M.A., 1997. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. AGSO Journal of Australian Geology & Geophysics 17, 201–216.
Wong, M.T.F., Harper, R.J., 1999. Use of on-ground gamma-ray spectrometry to measure plant-available potassium and other topsoil attributes, Australian Journal of Soil Research, 37, 267–277.
Yamaguchi, Y. Kahle, A.B. Tsu, H. Kawakami, T. Pniel, M., 1998. Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). IEEE Transactions on Geoscience and Remote Sensing 36, 1062–1071.

Chapter 3
A New Global Demand for Digital Soil Information

S.E. Cook, A. Jarvis and J.P. Gonzalez

Abstract The question has to be asked why – given the substantial advances in quantitative techniques over the years – 'full' Digital Soil Mapping has not been mainstreamed further and harnessed to the problems soil information can help address. This paper suggests some reasons for a slow adoption, causes for optimism for a wider adoption than at present and – using a case study – demonstrates the ease of further development at national scale. Finally, we propose how a major effort of digital soil mapping could support development in Africa, outlining the opportunities and obstacles that await contributors.

3.1 History of Quantitative Soil Information

3.1.1 From Geostatistics Through Soil-Landscape Mapping to Gaussian Processes

Quantitative soil mapping originated in the 1970's following a frustration with the limitations of conventional soil maps to provide quantitative information about soil properties that could be accommodated in 'normal' scientific thinking. Major problems had been pointed out in the transmission of information from conventional (choropleth) maps. These problems related to both the classification process (Webster, 1968) and spatial representation using conventional surveyor procedures (e.g. Valentine, 1983), since conventional soil survey methods used a wealth of tacit understanding that proved difficult for other users to re-interpret (Hudson, 1992). The products – soil maps, their legends and classification – though useful, could not be progressed further. Digital soil mapping offered a way out of this bottleneck by providing an explicit, quantitative expression of soil property variation. Thirty years later it appears to be in a strong position to deliver.

S.E. Cook
International Centre for Tropical Agricultura (CIAT), AA6713, Cali, Colombia
e-mail: s.cook@cgiar.org

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

Many of the current approaches to quantitative soil prediction are based on kriging. Ordinary Kriging is a form of weighted local spatial interpolation that uses a Gaussian model to derive spatial estimates of variables supported by a data-set for the area being analyzed. Its main drawback is that it does not explicitly use knowledge of soil materials or soil formation processes that explain variation, hence relies largely on the support of samples in order to produce satisfactory results. There are extensions to this method that allow the use of ancillary data, but they are difficult (if not impossible) to extend to more than one ancillary variable. Some of the most promising approaches to predictive soil mapping are expert systems and regression trees. Expert systems use expert knowledge to establish rule-based relationships between environment and soil properties (Cook et al., 1996). They may not depend on soil data to determine soil-landscape relationships, but some approaches do. Regression Trees are decision trees with linear models in the leaves. They create a piecewise linear representation of the predicted variable. Using this method Henderson et al. (2001) obtained the best results in the literature, which are able to explain more than 50% of the variance of several soil properties such as pH, clay content and sand content.

While acknowledging the value of digitised 'conventional' soil information, such as the Digital Soil Map of the World (FAO, 2000), the question has to be asked why – given the substantial advances in quantitative techniques over the years – 'full' digital soil mapping has not been mainstreamed further and harnessed to the problems soil information can help address. This paper suggests some reasons for a slow adoption, causes for optimism that digital soil mapping could be much more widely adopted than at present and – using a case study from Honduras – demonstrates the ease of further development at national scale. Finally, we propose how a major effort of digital soil mapping could support development in Africa, outlining the opportunities and obstacles that await contributors.

3.1.2 Advanced Mapping Techniques: Supply-Driven or Demand-Driven

Despite the advances in quantitative soil mapping techniques, most soil maps continued to be produced using conventional techniques. Soil information is predominantly in the form of conventional soil maps, albeit often digitised and with expanded legends. A major reason for this seems to be that – as with many new techniques – research focuses on the search for new methods more than the demand for the information they produce. Experience with development of innovative techniques suggests that a period is required in which promising methods are proposed, trialed and improved in an iterative process of continuous development. The demand during this period also expands as the benefits are articulated more clearly.

Prior to this expression of demand, effort in digital soil mapping has tended to respond to the 'supply' of capability. Without a strong external demand for specific products, method development has tended to focus on case-studies where large sample data sets are already available, rather than by a purposeful development to meet a new 'demand'.

Notable example of demand-led digital soil mapping include high-risk engineering applications demanded the accuracy that only geostatistical estimation could provide. This expresses the second reason for the slow uptake of digital soil mapping – the perceived value of better soil information may be quite small compared with other sources of uncertainty in agricultural decision-making. It is perhaps instructive to recall that for many agricultural ministries, the imperative to undertake soil mapping derives right back to well-publicised disasters. Many other agencies in both the developing and developed worlds have commissioned soil survey on the basis of a general expectation of value rather than a clearly specified demand for accuracy. For reasons explained below, we believe that the time is right to re-examine the demand to meet the challenges of agricultural development in the many countries that still lack detailed soil information.

3.1.3 Programs in Many Countries are Considered 'Complete'

In many developed countries, soil survey has been 'completed', meaning that information at 1:50 000, 1:25 000 or even better is already available. It is increasingly difficult in developed economies to argue that agricultural production requires more systematic survey when the perception of policy-makers and key decision-makers is that adequate soil information is already available. Initiatives to improve the provision of new information by quantitative methods will prove a 'difficult sale' under such conditions. In Western Australia in the early 1990's, the realization that the agricultural economy was facing a widespread threat of land degradation triggered a program of soil mapping to guarantee soil information coverage of 1:250 000 or better, aiming for 1:100 000 or 1:50 000 in high value agricultural areas. With few exceptions, information was provided by conventional soil survey.

3.2 A New Demand for Global Soil Information

The lessons above suggest to us that the first requirement of digital soil mapping is to identify the clear demand for the information it provides. Without this, effort is likely to be inappropriate to its final use, or under-resourced and restricted to 'speculative' research of indeterminate value.

The basic rationale for soil mapping is to provide information to reduce uncertainty. Improved accuracy of soil measurement is only one form of uncertainty – metric uncertainty – that is removed for decision-makers. Others – explained below – are described by Rowe (1994) as temporal, structural and translational. Structural and translational uncertainty can be particularly difficult to appreciate, but in this case they could be taken to describe firstly, the importance of soil variation in relation to other biophysical factors; and secondly the value that decision-makers place on the improvement that such information enables. To reduce structural uncertainty it is necessary to show that soil variation is considered to be a prime source of uncertainty to a solvable problem. To reduce translational uncertainty it is necessary to show that this problem is considered to be 'important' by key stakeholders around the problem.

The future of digital soil mapping therefore seems to lie more in answering questions about the potential value of information as much as answering those about methodological capability. By comparison with some major challenges facing agricultural development in Africa, we demonstrate a four-stage test of demand that should help providers clarify what information is required, and why digital methods are necessary to acquire it, and then compare these against the current situation. For digital soil mapping to be recognised as a necessity, it should pass tests of significance, novelty, actionability and delivery.

3.2.1 Is the Soil Information Significant to the Problem?

The spatial soil information provided by digital soil mapping must be perceived as highly significant to major investors to compel its acquisition. That is, it must show that digital soil mapping will remove a major source of uncertainty. Further, the advantage of digital soil mapping over conventional methods must be apparent.

Decades of research, at a range of scales shows that soil variation impacts significantly on agricultural and environmental processes. This means that statements made about processes are imprecise to the degree that the effect of soil variation is not explicitly accounted for. Yet, site variation remains unexplained in agronomic experimentation, while other sources of variation are pursued to a level which is of little practical significance. Over recent years, the volume of direct observations of yield variation from precision agriculture technology gives a better picture of within-field variation, in which the effect of soil variation is dominant, is often extremely large, accounting for up to 3 or 4-fold yield variations – far greater than effect of the treatments. Experience suggests that even farmers are surprised by the scale of this variation. Micro-scale effects of soil variation are therefore highly significant.

Agriculture is seen as less and less important to the economy and life-style of people in the developed world. In the latter half of the 20th century, most soil maps in the developed world were produced for agricultural ministries, where possible, changing in the 1970's onwards to address problems of environmental management. Since most soil maps had been designed with the aim of supporting the former goals, this change was of mixed success. However, the global significance of agriculture, and the demands placed on it for soil information are greater now than before. Agriculture remains the mainstay of livelihoods in the developing world. Agriculture is the major driver of socio-economic development in most developing countries and accounts for 30–60% of GDP. Nash (2005) reported that 63% of global population (and 73% of poor, approximately 900 million) live and work in rural areas. Soil information can assist development by (a) enabling farmers to meet the threats posed by global climate change and increasing water scarcity and pressure of land degradation, and (b) identifying a pathway out of poverty through emerging opportunities to tap into markets.

The practical significance of meso-scale soil variation can be illustrated in relation to global climate change. Many consider that global climate change to be the greatest threat facing sustainable agriculture. The impacts seem destined to be distributed unequally such that impacts are likely to be most severe in sub-Saharan Africa (IPCC, 2001; Jones and Thornton, 2003) which, with almost 40% of people under-nourished already faces enormous problems of food insecurity (Pretty, 1999). The reality of climate change is likely to be felt most keenly at a local scale, where people who are considered to be amongst the most vulnerable in the world must strive to adapt to adverse change. It is now understood that adaptive change is the key to survival for such people, yet adaptation in ignorance of fundamental changes of risks to cropping, relating to interactions with soil water and nutrition – increases the risks of an already difficult existence. While endogenous information, generated through experience of adaptation locally, is a more powerful source of understanding, it seems clear that exogenous information is essential to accelerate its development.

An example of the type of information required is of drought risk, which is a major constraint to development in Sub-Saharan Africa and is cited by farmers as the principal hazard (Dercon, 2002). While drought risk is understood well by farmers, it is difficult to assess intuitively. Even the mere threat of drought risk slows down development, by encouraging alternative risk avoidance strategies that reduce productivity below the potential. Drought risk is influenced strongly by soil variation, yet the information is lacking on which to assess covariate risk within an area, and against which to improve predictive modelling. The uncertainty related to soil variation is highly practical.

3.2.2 Is the Information Novel?

The information must offering sufficient new insight from that which is currently available. At a micro-scale, a common obstacle to acquisition of information is that while soil variation is significant, soil maps fail to offer more information than ‘farmers already know’. At meso-scale, we perceive that soil maps are taken to answer all questions, even though such maps are often absent. At a macro-scale, digital soil mapping must offer substantial new insight to help understand soil-related problems such as carbon budgeting, management of scarce water resources or constraints to agricultural productivity.

The simplest illustration of this is provided at a micro-scale by experience of precision agriculture. Literally thousands of highly detailed maps have been produced of yield variation from North America, Europe and Australia, which in many cases, show significant variation that was not understood and of unexpected degree to experienced farmers. At a meso scale, it is easier to ensure that soil maps provide novel information where – as in the majority of areas – no such maps pre-exist. Certainly, in the developing world, virtually all soil information that is provided at this scale is novel where the best alternative is based on mapping at scales of 1:1 million.

At all scales, digital soil mapping provides novel information if it explains additional variation of soil attributes that cannot be adequately explained using more conventional information. This does not seem very hard with respect to specific soil variables, where conventional maps rely on soil classification.

3.2.3 Is the Information Actionable?

We use the term ‘actionable’ to distinguish information that is linked to specific decisions, such as a decision to invest in a particular area. The test of ‘actionability’ is perhaps the hardest to satisfy, because it relies on many other conditions that can influence the readiness to decide. The tests of significance and novelty specify the *potential* importance of digital soil information. While some soil maps may justify investment to satisfy a purely educative function, the predominant expectation is that information will ultimately promote specific actions. Sometimes these need spelling out.

In the context of developing agriculture, information can be acted upon in three ways: targeting of investment or aid; policy design or to direct action such as planting. In all cases, the decision to act is the result of interpreted soil information, rather than the raw information. For example, suitability maps directed soil information, with other information, towards a specific cultivation decision. Similarly, the World Food Program or USAID could use soil information, with other data, to help target activities to assist people in areas that are either drought stricken, or lower risk (hence more suitable targets for investment). An advantage of digital soil mapping is that information is not lost through soil classification, hence more easily re-interpreted with specific applications in mind. It is also easier to update provided the spatial infrastructure allows this. The problem seems to be that in making the information specifically actionable, there is a risk in over-specialisation, thereby restricting the range of potential users who will seek the information.

3.2.4 Can the Information be Delivered to Stakeholders?

Having demonstrated the *potential* demand of digital soil mapping, the final test is to consider the practicalities of delivering information to the user. There is increasing recognition of the importance of providing free access to information to a very wide range of potential users, from policy-makers to farmer representatives. The need to transmit actionable digital soil information to users presents major operational challenges of design. In the developing world, operational problems ensue as a result of the so-called digital divide, leaving many areas without access to information delivery. While access to Information Communication Technologies (ICTs) is growing in some regions (e.g. Latin America and South East Asia) through the increasing use of internet cafes and cellular phones, for many parts of the developing world, regular access to such information does not exist beyond regional cities.

A second aspect of deliverability is the 'self-financing' character of information. Experience in development with the adoption of tele-communications, microfinance and micro-insurance (all information-rich instruments) suggests that if the instrument is robust and of evident value to users, delivery occurs with remarkably little promotion – people at all levels work out how to use the instrument.

3.3 Capability Improved

We now mention some technological developments that increase the potential of digital soil mapping to contribute substantially to agricultural improvement. These comprise new data; new processing and delivery capability and new understanding of decision support needs.

3.3.1 New Data: Topography, Climate and Vegetation

New opportunities for digital soil mapping originate from a data revolution which is providing more data on environmental variables at higher resolutions (spatial and temporal), for the entire globe. The three principle advances are for higher resolution topography, climate and vegetation data. These include:

- SRTM: High resolution terrain model (90 m, spatial resolution – improving to 30 m). Processed and downloadable from http://srtm.csi.cgiar.org (Jarvis et al., 2004).
- WorldClim: 1 km spatial resolution climate data. Processed and downloadable from http://www.worldclim.org (Hijmans et al., 2005).
- MODIS: high temporal resolution thermal and spectral imagery providing global images of vegetation every 16 days, with a spatial resolution of 250 m.

There are numerous other types of data that have become available over the past decade and many are reviewed in Chapter 2.

3.3.2 New Processing and Delivery Capability: Web-Based Delivery of Very Large Data-Sets

IDIS (Marchand, 2006) is a web-based system that delivers large spatial datasets from several major river basins around the world for use by researchers, policy-makers and others. The system delivers a large variety of geo-referenced data and is envisaged as a medium for discussion and development of methods to further interpret the mass of data that is delivered by collaborators. Similar methods could be deployed to exploit the information coming from digital soil mapping, and to encourage a transparent development of interpretations from a broad constituency of users.

3.3.3 New Understanding of Decision-Support Needs

The third advance we note is the improvement in understanding of the nature of change in agriculture, from which we could expect a fuller appreciation of the potential roles for information. While some soil maps have doubtless proved extremely valuable to specific instances, there are probably an equal or greater number of instances when information has lain unused in filing cabinets, or that users felt they were not provided with the information required. Difficulties of communication between providers and users of soil information can reflect a mis-comprehension that change in agriculture is a linear process, whereas it is now viewed as a more complex process of adaptive management (Douthwaite, 2002). This is good news for providers of digital soil mapping which has flexibility to provide soil information suitable to be accommodated in a dynamic learning process. Since all observations during such a process are influenced, to some degree, by site conditions, the opportunity exists to use soil information to help explain variation of observed change and to accelerate further change towards ‘preferred sites’.

3.4 Case Study Using New Data

Pracilio et al., 2003 illustrate the use of digital soil information, coupled to crop simulation modelling, to represent spatial variations in soil water balance in an annual cropping system over a catchment in Western Australia. The catchment extent was about 500 km^2 and the mapping process could have been repeated over similar areas within the region for which input data was available. In this case, input data comprised a terrain model, pre-existing (low resolution) soil map, a geology map and partial coverage of airborne geophysical data. Several features distinguish the spatial information provided by the digital soil mapping from a conventional soil map, should it have been available.

The first feature was that the data was presented as a grid of higher spatial resolution than can be provided by normal soil maps. Effectively, terrain and geophysical data greatly improved the spatial resolution of soil information. This proved valuable to aid visual interpretation of patterns of variation in catchment hydrology and helped farmer groups, for whom the information was produced, understand the hydrologic consequences of changes in cropping patterns.

The second feature was that it was possible to accommodate the uncertainty of information about continuous variation, by using a probabilistic formulation, in ways that are difficult in conventional soil maps. A range of potential simulation model outcomes was designated for each grid cell, according to the strength of supporting evidence. This produced spatial information of outputs that accurately captured hydrologic variation. Should it have been required, uncertainty of input data could have been traced through the modelling process to identify error propagation.

The third feature was that the more transparent and flexible management of spatial soil data enabled Pracilio et al. (2003) to work ‘backwards’ from the demands of simulation modelling to determine what soil information was valuable. This contrasts to the conventional use soil maps which starts with soil map units and interprets forwards. The question that was asked was as follows: ‘Given a set of hydrologic behaviours that are associated with a known set of soil conditions, determine where these conditions are likely to be distributed over the catchment, hence the likely hydrologic behaviour’.

3.5 Conclusions

We draw the above observations together with consideration of a proposal to provide high resolution digital soil information for Africa, and show how digital soil mapping could respond to some major challenges facing agricultural development in Africa.

1. What significant problems would digital soil mapping help address?

Digital soil mapping could significantly reduce uncertainty to help address a range of major problems such as drought, adaptation to global climate change and improvement of production systems through improved nutrient management. For most parts of Africa, soil information is available at reconnaissance scale only, and then based on broadly based soil classifications that are of general, rather than specific application. digital soil mapping could provide information at more detailed spatial scale required to support local participatory initiatives that are seen as key to change. digital soil mapping could provide soil information in a more flexible and dynamic interpretative format that could help address the specific questions of groups of stakeholders.

Given the dearth of detailed soil information for most of the continent, the test of novelty (see Section 3.2.2) is easy to satisfy. Digital soil mapping would provide a huge lift of novel insight into sub-regional and local variation of agricultural performance relating to soil variation.

2. What specific actions could be supported by this information?

The range of actions supportable by digital soil mapping spread from broad support for policy design, consistent with best available information of risks and opportunities for agricultural change as they are likely to be expressed on the ground. Digital soil mapping could be used to improve targeting investment in specific agricultural technologies, starting with effective fertilizer use where the lack has constrained improvements in crop productivity. Finally, digital soil mapping could be used to vary the design of financial instruments to help manage production risks of drought and erosion such as the drought protection offered by site (and soil) specific insurance, whereby premiums could accommodate a range of risk profiles from most droughty to most retentive soils.

3. How will information be delivered?

This is perhaps the major practical challenge facing digital soil mapping because – despite the potential value of such information – it is difficult to envision national institutions having the financial or intellectual capacity to provide this information, nor the political will to invest in programs of mapping, hence development of capacity. Information would need to be coupled to specific demands for information to generate the political support and revenue necessary to initiate and sustain a program of digital soil mapping, while at the same time, a broadly-based program of capacity-building would be needed to address the major problems such as adapting to Global Climate Change.

Several options exist to encourage development:

- Development of high resolution data with global coverage, likely to be of value for digital soil mapping. Examples include SRTM, Worldclim data and coverages of soil maps such as the FAO Digital soil map of the World. Derivatives of this data are likely to be more valuable than the raw data itself.
- Case studies of digital soil mapping, linked to specific applications that are likely to be of broad significance. Examples might include the use of digital soil mapping to development of targeted adaptation to global climate change funded in their own right.
- Development of specific instruments, or derivates, that convert digital soil mappings into directly utilizable information to support decisions. An example is the incorporation of soil information into site-specific drought insurance premiums (Diaz-Nieto et al., 2006).

References

Cook, S.E., Corner, R.J., Grealish, G. and Chartres, C.J., 1996. A Rule-based system to map soil properties. Soil Science Society of America Journal 60, 1893–1900.

Dercon, S., 2002. Income risk, coping strategies and safety nets. The World Bank Research Observer 17(2), 141–166.

Diaz-Nieto, J., Cook, S., Lundy, M., Fisher, M. Sanchez, D. and Guevara, E. 2006. A system of drought insurance for poverty alleviation in rural areas. Final Report for BMZ project SLB-50. Bonn.

Douthwaite, B., 2002. Enabling Innovation: A practical guide to understanding and fostering technical change. Zed Books.

FAO, 2000. Digital Soil Map Of The World And Derived Soil Properties on CD-ROM. FAO, Rome. http://www.fao.org/AG/agl/agll/dsmw.htm.

Henderson, B., Bui, E., Moran, C., Simon, D. and Carlile, P., 2001. ASRIS: Continental-scale soil property predictions from point data. CSIRO Land and Water Technical Report 28/01. CSIRO Land and Water, Canberra.

Hijmans, R.J., Cameron, S., Parra, J., Jones, P.G. and Jarvis, A., 2005. WorldClim: Very high resolution global terrestrial climate surfaces for monthly temperature and precipitation. International Journal of Climatology 25(15), 1965–1978.

Hudson, B.D., 1992, The soil survey as paradigm-based science. Soil Science Society of America Journal 56, 836–841.

IPCC, 2001. Climate Change 2001: Impacts, Adaptation, and Vulnerability. Report of Working Group II of the Intergovernmental Panel on Climate Change (6th Session, Geneva, Feb. 2001).

Jarvis, A., Rubiano, J., Nelson, A., Farrow, A. and Mulligan, M., 2004. Practical use of SRTM data in the tropics – Comparisons with digital elevation models generated from cartographics data. Working Document no. 198, 32 pp. CIAT, Cali, Colombia, available from: http://srtm.csi.cgiar.org/PDF/Jarvis4.pdf.

Jones, P. G. and Thornton, P.K., 2003. The potential impacts of climate change on maize production in Africa and Latin America in 2055. Global Environmental Change 13, 51–59.

Marchand, P., 2006. IDIS. Integrated Data Information System. URL: http://dw.iwmi.org/dataplatform/Home.aspx.

Nash, J., 2005. Sustainable Development Strategies in Agriculture and Rural Development a paper presented during the WTO Symposium on Trade and Sustainable Development held on 10–11 October, 2005.

Pracilio, G., Asseng, S., Cook, S.E., Hodgson, G., Wong, M.T.F., Adams, M.L. and Hatton, T.J., 2003. Estimating spatially variable deep drainage across a central-eastern wheatbelt catchment, Western Australia. Australian Journal of Agricultural Research 54, 789.

Pretty, J., 1999. Can Sustainable Agriculture Feed Africa? New Evidence On Progress, Processes And Impacts. Environment, Development and Sustainability 1, 253–274.

Rowe, W.D., 1994. Understanding Uncertainty. Risk Analysis 14(5), 743–750.

Valentine, K.W.G., 1983. Guest Editorial: Another way of doing things. Soil Survey and Land Evaluation 3, 29–30.

Webster, R., 1968. Fundamental objections to the 7th approximation. Journal of Soil Science 19, 354–366.

Chapter 4
Development and Application of Digital Soil Mapping Within Traditional Soil Survey: What will it Grow Into?

D. Howell, Y.G. Kim and C.A. Haydu-Houdeshell

Abstract In this Chapter we describe our use of digital soil mapping estimates as input to traditional field soil survey in California, U.S.A. We also describe the development of these raster soil property models as stand-alone products, and practical implications of their use. This Chapter deals with application of existing digital soil mapping tools in active soil surveys, rather than research of new methods.

The soil survey program in the United States is nearing completion of "once-over" coverage of the nation. Many potential soil survey users in the remaining unmapped areas expect to use traditional polygon-based soil maps.

Soil-landscape models based on field point data have been developed in support of selected soil survey projects. We expand on our previous models in a test area that has existing point data and polygon soil mapping. New soil-forming factor covariates (IFSAR elevation data and ASTER satellite images) are used to derive the models. Minor improvements in the model estimates were obtained. Then significant variables from these models are used to test the feasibility of the creation of field soil survey office tools.

We feel that raster soil-landscape models are a developmental product of soil survey. They are just becoming useful as pre-mapping estimates of the spatial distribution of some individual soil properties. The explicit estimation of all significant soil properties based on a suite of individual models is still to be developed. This is required before informed land management decisions can be based on digital soil mapping.

Since natural resource management methods and regulations are coordinated locally, regionally, and nationally, standards for the creation and implementation of these models are required for consistent and coordinated outputs within a nation.

D. Howell
USDA Natural Resources Conservation Service, Arcata, CA, USA 95521
e-mail: David.Howell@ca.usda.gov

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

4.1 Introduction

Currently we are developing and evaluating digital soil-landscape modelling (digital soil mapping) as a pre-mapping aid in traditional soil survey activities in California, U.S.A. The goal is to help soil scientists plan their field work and augment their understanding of soil-landscape relationships.

Our past, initial work (Howell et al., 2006) focused on showing some of the existing methods (McKenzie and Ryan, 1999; McBratney et al., 2003) and how they could be useful to traditional soil survey in the western United States. The purpose of this Chapter is to extend that work with new, higher-resolution elevation data and satellite imagery with more shortwave and thermal infrared bands representing the soil-forming factor covariates. The models are evaluated against point data in a completed project area. The resulting models will be used in an adjacent, unmapped area later to produce estimates of the distribution of key soil properties prior to field work.

The soil survey program in California, and throughout the United States, still has active soil survey projects with field soil scientists conducting traditional field work in order to complete the mapping of the nation and to update previously completed work.

In addition to the statistical model development, we evaluated several standard geospatial tools for field soil scientists to help them put the information produced by these calculations to use.

4.2 Material and Methods

The study site is located in the western Mojave Desert approximately 160 km north-east of Los Angeles, California, U.S.A. The study site receives 76–127 mm of rain per year with the majority falling between November and March. This is the same study site as our previous work, although a portion was removed because satellite image coverage was lacking (Howell et al., 2006).

The size of the study area is 68 075 ha and the resolution of the raster modelling is 5 m, except for one model which was 90 m.

Point data were available from randomly located soil profile descriptions and from purposively located points from traditional soil survey activities (Haydu-Houdeshell, 2003). Models were derived from a combined set of the points ($n = 285$), and were then evaluated against a set of points (that were not used to derive the models) that represented 15% ($n = 49$) of the total number of the original dataset. These evaluation points were selected randomly from the full combined set.

Elevation data provided variables for slope steepness, slope curvatures, topographic ruggedness index (average elevation change between any point on a grid and its surrounding area) (Riley et al., 1999) and other derivatives. Minimum curvature and maximum curvature were calculated using the methods of Schmidt and Hewitt (2004). We used Interferometric Synthetic Aperture Radar (IFSAR) data

(Intermap Technologies Incorporated, 2005) with a resolution of 5 m. IFSAR data represented an improvement over 30 m DEMs and worked well in this sparsely vegetated, arid region.

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data (Yamaguchi et al., 1998) were used to provide surface reflectance and thermal emission information. The six shortwave infrared bands and the five thermal infrared bands theoretically would provide increased discrimination (when used in band-ratio combinations) over information available from Landsat 5 imagery used in our previous work (as described by Boettinger et al. in Section 16.2.2). ASTER data were resampled to 5 m (from 15 m for VNIR, 30 m for SWIR, and 90 m for TIR) to coincide with elevation data resolution.

Elevation derivatives and ASTER bands and band ratios produced 46 covariates. Available vegetation and precipitation data were reviewed and determined to be too general or not vary significantly within the study area. Available geomorphic and geologic data were also general and captured at a small scale (coarse resolution) so that precision of landform delineation was low and would negatively impact model precision. After screening and reduction of covariates because of an excessive amount of multicollinearity (SAS, 2003) we worked with a total set of 19 variables, and in most cases the selected models used far fewer than this total set (See Table 4.1).

The soil attributes we modelled were soil genetic features such as: presence (or absence) of argillic horizon, secondary carbonates, calcic horizon, durinodes, duripan, and separate (continuous) models estimating the depth to the occurrence of these features. We also estimated particle-size class (Soil Survey Staff, 1999).

Point data were used to calculate generalized linear models estimating the depth in the soil profile at which individual soil genetic features occurred. Presence or

Table 4.1 Covariates considered for modelling

Elevation	IFSAR elevation (metres above sea level)
Aspect	IFSAR derivative (ESRI, 2004)
Curvature	IFSAR derivative (ESRI, 2004)
Slope	IFSAR derivative (ESRI, 2004)
Plan Curvature	IFSAR derivative (ESRI, 2004)
Maximum Curvature	IFSAR derivative (Schmidt and Hewitt, 2004)
Minimum Curvature	IFSAR derivative (Schmidt and Hewitt, 2004)
ASTER Band Ratio T1/S1*	ASTER band 10 / band 4
ASTER Band Ratio T1/V3	ASTER band 10 / band 3
ASTER Band Ratio T4/V3	ASTER band 13 / band 3
ASTER Thermal IR Band 1	ASTER band 10
ASTER Thermal IR Band 2	ASTER band 11
ASTER Thermal IR Band 3	ASTER band 12
ASTER Thermal IR Band 4	ASTER band 13
ASTER Thermal IR Band 5	ASTER band 14
ASTER Band Ratio V2/V1	ASTER band 2 / band 1
ASTER Band Ratio V3/V1	ASTER band 3 / band 1
ASTER Visible Near IR Band 1	ASTER band 1
ASTER Shortwave IR Band 1	ASTER band 4

*T = thermal infrared, S = shortwave infrared, V = visible near infrared absence of these features was also estimated using logistic regression. The models developed for these estimations also provided a list of significant variables that were combined in a raster stack and then further analyzed using principal component analysis. Principal component analysis was evaluated using tools included in the software currently available in the field soil survey offices (Spatial Analyst, ESRI ArcGIS).

Models were implemented spatially using GIS raster software (ESRI, 2004) and evaluated by extracting (ESRI, 1998) model estimate values at the geographic locations of the test data set (n = 49) and comparing to those actual field observed properties.

4.3 Results and Discussion

4.3.1 Results

In Tables 4.2 and 4.3 comparisons are made of the models produced in this effort (IFSAR/ASTER) to the models from the previous work (DEM 30 m/Landsat 5) that used standard 30 m digital elevation model data produced by the United States Geological Survey and Landsat 5 band ratios, along with several other covariates (see additional considerations in Section 10.7.2). In general, there are some slight improvements in the estimated properties with the new models. We continue to assert that overall these estimates are useful as pre-mapping estimates of the distribution of soil genetic features.

However, fewer models met our significance test ($p < 0.05$). Direct comparisons are not perfect because different covariates were used and a different set of test points were used. We are still evaluating models and trying to improve the fit using these data.

A raster stack of the significant variables determined in these models was developed. Then standardized tools (ESRI, 2004) for calculating principal components were evaluated. Principal component analysis of 19 covariates produced poor model estimates based on eight principal components. This still seems like a worthwhile method to develop and we will work on this in the future.

Table 4.2 Comparison of model estimates of particle-size class to actual measured soil properties

		Number of classes model estimate was away from correct class (ten total classes)					
	n	Correct Class	1	2	3	4	5
		%	%	%	%	%	%
DEM 30 m/Landsat 5	97	24	49	21	3	2	1
IFSAR/ASTER (1)	49	35	24	29	12	0	0
IFSAR/ASTER (2)	49	41	31	20	6	2	0

Table 4.3 Comparison of model estimates of depth of soil genetic features to actual measured soil properties

	n	% Estimates within 0–10 cm	% Estimates within 10–20 cm	% Estimates within 20–30 cm	% Estimates within >30
Argillic					
DEM 30 m/Landsat 5	46	24	11	39	26
IFSAR/ASTER (1)	27	19	37	11	33
IFSAR/ASTER (2)	27	33	19	19	29
Calcic					
DEM 30 m/Landsat 5	39	18	39	26	18
IFSAR/ASTER	19	26	5	32	32
Carbonates					
DEM 30 m/Landsat 5	97	48	23	17	12
IFSAR/ASTER (1)	46	43	30	4	21
IFSAR/ASTER (2)	46	48	26	4	22

When more than one model was estimated they are indicated by (1) or (2).

Some models are more useful than others, certainly. Models that estimated presence/absence of the genetic features using logistic regression performed poorly with these data, as they did in the previous work. Possible reasons for this are an absence of reliable relationships between available covariates and subsurface features, inadequate spatial accuracy of data points, soil property variations at a scale finer than the covariate data, or other undiscovered causes.

The IFSAR elevation data did not produce the magnitude of improvements that we expected in the model estimates. But these elevation data are still an improvement for soil survey work because they have higher resolution, lack artifacts from fitting to hypsography, are more sensitive to small areas of varying elevation, and are directly measured. We also feel that further evaluations of band-ratio combinations of the ASTER data will produce enhanced images that correlate to variations in soil surface properties. We will continue to work with these.

Most importantly the models provided estimates within 20 cm of actual depths of secondary carbonates for about 73% of the test points; and estimates of the particle-size class that were correct or within one class of the actual class for about 72% of the test points. This is useful information to know prior to field mapping. (See Fig. 4.1.)

4.3.2 Discussion

The best test of these model estimates will come during the 2007 field mapping season when model outputs will be produced for an unmapped area near this project area. We will develop the models further prior to that application. Since we consider these as pre-mapping estimates, the field soil scientists will provide us feedback

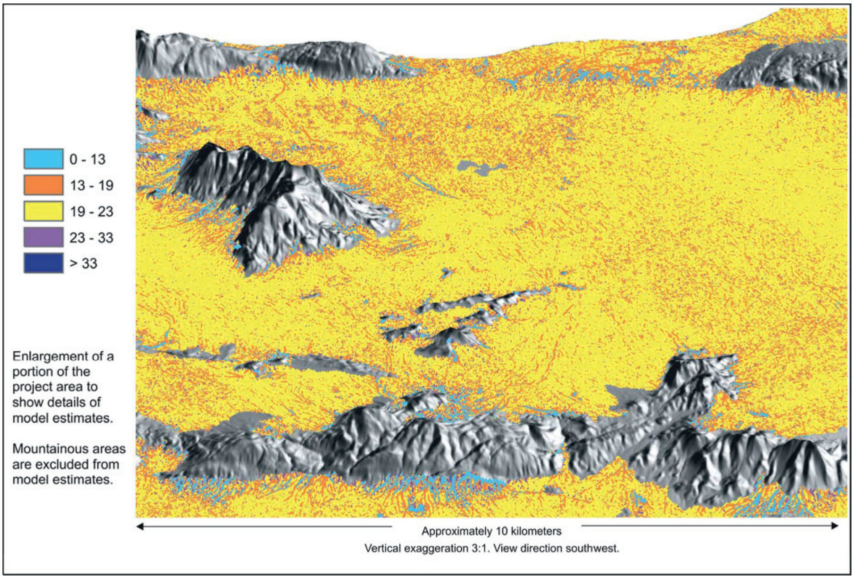


Fig. 4.1 Estimated depth (cm) to top of zone of accumulation of secondary carbonates (See also Plate 1 in the Colour Plate Section)

about how useful they were in guiding their fieldwork and building soil-landscape relationship concepts. At this time our use of models developed in one geographic area and then applied to another unmapped location is guided by proximity, judgment, and work priorities. No formal analysis of reference area or applicable extent for the models has been undertaken. We will evaluate the reference area concept (Lagacherie et al., 1995, 2001) for future work and it should be considered for development in the digital tools provided to field soil scientists.

In order for raster digital soil mapping to be implemented in California it needs to be standardized within geographic areas and implemented at the field soil survey office. Tools need to be developed within the standard software available to soil scientists. The landscape analysis and modelling methods need to be put into the hands of the field soil scientist, rather than being completed by a specialist who is not involved in the day-to-day field work of the project. Training needs to be developed and given so that field workers have the tools that they need and the knowledge of how to use them. In short, digital soil survey needs to be applied at the field soil survey office level. This will require development of scripts, geoprocessing models, and toolbars for the standardized GIS software used in soil survey offices. Soil scientists who are also GIS and remote-sensing specialists will support the development of these methods. They will help other soil scientists understand the geographic and predictive limitations of these tools. This is what we feel digital soil mapping will grow into, initially. The next generation of field soil mappers will use these tools in their daily work and innovate applications of new remote sensing and spatial statistics methods to produce useful explicit estimations of soil information.

The soil survey program in the United States is nearing completion of "once-over" coverage of the nation. Many potential soil survey users in the remaining unmapped areas expect to use traditional polygon-based soil maps. Natural-resource planning and management methods have been developed using this type of soil information. The soil map unit polygons serve as indications for management units. These are made available for nationwide online soil mapping at the Web Soil Survey (see also Section 24.2). While these traditional soil survey projects are being completed to meet these expectations, raster-based soil-landscape models are being developed and evaluated as input to the mapping and as stand-alone products for use in other models.

We feel that raster soil-landscape models are still a developmental product of soil survey. They are just becoming useful as pre-mapping estimates of the spatial distribution of some individual soil properties. The explicit estimation of all significant soil properties based on a suite of individual models is yet to be developed. For example, to use raster soil property estimates to make an interpretation of the soil limitation for septic tank installations a separate raster estimate is needed for each soil property used in the rating criteria, i.e., surface water ponding, depth to bedrock, depth to cemented pan, permeability, slope, flooding, and rock fragment content. This is necessary before informed land management decisions can be based on soil property raster estimates.

Since natural resource management methods and regulations are coordinated locally, regionally, and nationally, standards for the creation and implementation of these models are required for consistent and coordinated outputs within a nation. In addition, the model outputs present interpretation challenges to the natural resource planner and manager. How do they interpret or use a raster estimate or even a stack of raster estimates to decide on placement of facilities or practices? How would a regulatory agency such as a regional planning department review how the models were used by conflicting groups to make decisions in order to meet planning regulations? For example each regulatory agency would need to be able to objectively evaluate the source of covariate data, pre-processing methods, implementation of the model mathematically, and grouping or filtering of the model outputs. The adoption of raster soil property estimates in the United States as stand-alone products requires the development of these standards and practical interpretation methods. We are just beginning to discuss these topics as the potential uses of these digital raster data are demonstrated and accepted.

4.4 Conclusions

In the United States medium resolution satellite imagery (15–60 m), elevation data (10–30 m), vegetation mapping, other soil-forming factor data, and soil class maps are commonly available in the public domain. Geologic data are more limited and are often generalized to formations rather than rock types. Radiometric, hyperspectral, and high-resolution elevation data (e.g., Lidar, IFSAR) are still expensive and very limited. These are usually available only by contract and are proprietary. In some project areas, such as extensive desert ecosystems, the variation of the vegetation, precipitation, and geologic data may be small. This leads to models that focus on elevation derivatives and satellite images as the primary covariates, although all available covariates should be evaluated.

These data, along with field soil profile descriptions and laboratory data, provide inputs for explicit, soil-landscape model estimates. These model estimates are useful pre-mapping products. They can help guide field sampling and provide support for extrapolation to unvisited field locations.

We found that increasing the spatial resolution (changing the elevation data resolution from 30 m in our previous work to 5 m in this project) and increasing the attribute resolution (using ASTER data with 14 bands instead of Landsat data with 7 bands, i.e., more narrowly defined bands) did not increase the performance of the models in a dramatic way. Perhaps the significant elevation variance was portrayed by the 30 m DEM. The surface reflectance captured by the ASTER sensors may not have a significant correlation to subsurface features although in many areas burrowing animals and other disturbances have brought subsurface materials to the surface.

Tools for analyzing soil-landscape relationships need to be developed for easy application by field soil scientists using standard soil survey office software. We will focus on principal component analysis, unsupervised classification, and development of rule-based models. We will work with field soil scientists to develop these methods.

The biggest infrastructure challenge to implementation of explicit raster estimates of soil properties or classes in the United States is the adoption and communication of standards for model development and application. Another challenge is the development of methods for interpretation of raster soil estimates. Until conservation planners know how to use these raster estimates to make decisions or make recommendations for land use management, these maps will not replace traditional polygon soil class maps.

References

Environmental Systems Research Institute (ESRI). 2004. ArcInfo v9.0. Redlands, CA.

Environmental Systems Research Institute (ESRI). 1998. GridSpot70. [Online] Available: http://arcscripts.esri.com/details.asp?dbid=11037 [May 30, 2002].

Howell, D., Kim, Y., Haydu-Houdeshell, C., Clemmer, P., Almaraz, R., Ballmer, M. 2006. Chapter 34. Fitting soil property spatial distribution models in the Mojave Desert for digital soil mapping. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, An Introductory Perspective. Developments in Soil Science, Vol. 31. Elsevier, Amsterdam, pp. 465–475.

Haydu-Houdeshell, C. 2003. Soil profile descriptions Johnson Valley Off Highway Vehicle Area Soil Survey Project. Personal Communication. USDA Natural Resources Conservation Service.

Intermap Technologies Incorporated. 2005. Digital Terrain Model. Interferometric Synthetic Aperture Radar.

Lagacherie, P., Legros, J.P., Burrough, P.A. 1995. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. Geoderma 65, 283–301.

Lagacherie, P., Robbez-Masson, J.M., Nguyen-The, N., Barthès, J.P., 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma 101, 105–118.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A terrain ruggedness index that quantifies topographic heterogeneity, Intermountain Journal of Sciences 5, 1–4.

SAS Statistical Software Release 9.1. 2003. SAS Institute Inc. Cary, NC, USA.

Soil Survey Staff. 1999. Soil Taxonomy, Second Edition, Agriculture Handbook No. 436. United States Department of Agriculture, Natural Resources Conservation Service. Washington, D.C.

Schmidt, J., Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. Geoderma 121, 243–256.

Yamaguchi, Y., Kahle, A.B., Tsu, H., Kawakami, T., Pniel, M., 1998. Overview of Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). IEEE Transactions on Geoscience and Remote Sensing 36, 1062–1071.

Chapter 5
Soil Map Density and a Nation's Wealth and Income

Alfred E. Hartemink

Abstract Little effort has been made to link soil mapping and soil data density to a nation's welfare. Soil map density in 31 European countries and 44 low and middle income countries is linked to Gross Domestic Product (GDP) per capita and the number of soil scientists per country. National coverage of exploratory soil maps (>1:250 000) is generally higher in the poorest countries and decreases with increasing GDP per capita, whereas the national coverage of detailed soil maps (<1:50 000) tends to increase with increasing GDP. GDP is larger in countries with more soil scientists per unit area, likewise, the number of soil scientists increases with increasing GDP. More soil scientists per ha of agricultural land was found to be related to higher crop yields. Obviously, there are many confounding and interacting factors but this analysis illustrates how proxies for soil map density can be used; it is suggested that appropriate indicators should also be developed for spatial data infrastructures and digital soil maps to demonstrate their effectiveness for society and human welfare.

5.1 Introduction

Some countries are poor, some are rich, and there a lot of countries in between. Explaining the differences is not easy and related to a whole series of factors. Wealth and income of countries is driven by macro-economics but also by, for example, geography and the richness of natural resources: e.g. soil, climate and mineral wealth (Sachs, 2005). It is hard to unravel the influence of each developmental factor – many of which are interacting and are also greatly affected by humans. If the wealth of a nation can be viewed as its accrued assets and inherent property, the income is the yearly money that is derived from that wealth. The soil is an obvious factor in the wealth and income of a nation and may have a clear relation to a nation's wealth

A.E. Hartemink
ISRIC – World Soil Information, PO Box 353, 6700 AJ Wageningen, The Netherlands
e-mail: alfred.hartemink@wur.nl

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

and income. It is assumed that such relation not only holds for the wealth of soil resources itself, but also the wealth of information about those soil resources.

Little effort has been made to link soil information to developmental indicators or to quantify the effects of soil research on the wealth and income of nations. That is not surprising as the benefits of soil research have been poorly quantified (Greenland, 1991) as they are hard to measure and may be masked by other factors. Since soil science emerged in the mid 1800s, an enormous amount of information has been collected and insight has been gained in both the intrinsic properties of soils and the spatial soil distribution in different parts of the world. Relating intrinsic soil properties to crop productivity ($\approx$ yield, income) is relatively easy. For example, the economics of fertilizer applications or large-scale drainage scheme have been studied in most parts of the world and have shown to be essential for generating income and wealth. Currently, such efforts have reached a new stage with the rapid developments in soil sensors and precision agriculture (see Section 2.2 and 2.3 for an overview of new hardware and software). Given the quantitative nature of these studies and their associated uncertainty it should now be possible to accurately estimate the economic (and ecologic) benefits of soil management strategies. This mostly applies to the farm level although the variation and uncertainty in the information (McBratney, 1992) will affect the outcome of economic evaluations. Very few economic benefit studies are available at higher levels of aggregation (e.g. nations, continents) on which we mostly rely on old maps and old data. These maps were produced using traditional techniques and are generic and multipurpose so that is difficult to assess the economic benefits of the soil maps (see also Section 24.6).

There have been many claims, mainly by soil surveyors that soil surveys and mapping are economically beneficial. A problem with assessing the cost-benefit ratios of soil mapping is, however, that it is not possible to make precise generalisations about the costs of producing soil maps (Bie and Beckett, 1970). What is known is that the cost of soil survey (per unit area) rises sharply with the purity or uniformity to be achieved (Bie et al., 1973). Klingebiel (1966) reviewed a series of soil surveys and estimated that the benefit-costs ratios are larger than 50 for the USA, whereas Dent and Young (1981) also mentioned that these ratios for soil surveys are usually very high. Although only few studies have assessed the benefits of soil mapping and research (Giasson et al., 2000), there are several examples of projects that have failed because of a lack of soil information in all parts of the world (Bie and Beckett, 1970).

Globally, about two-thirds of the countries have been mapped at a 1:1 million scale or larger, but over two thirds of the total land area has yet to be mapped even at a 1:1 million scale (Nachtergaele and Van Ranst, 2003). That resulted from soil surveys conducted after World War II and up to the 1980s. At present, few traditional soil surveys are being carried out and many soil survey centres in the world have closed. There are great differences between countries in the status of mapped areas (extent, scale) but also in the status of digitising old information and combining it with other data layers to produce digital soil maps (McBratney et al., 2003) – see also Section 3.1.1. Fairly accurate data exist on the coverage of soil maps at different scales for most countries. In this chapter, soil map density at different scales is linked to GDP per capita and the number of soil scientists of 31 European countries and 44 low and middle income countries. Soil map density is used here as a proxy for soil data density. First, I shall look at the number of soil scientists per country because: no soil maps without soil scientists.

5.2 Soil Scientists per Country

The amount and quality of soil research is dependent on the number of soil scientists and their resources. It is possible to estimate the research resources of individual departments and centres, but quantifying the total money available and earmarked for soil science in a nation is hardly possible. Data on the number of soil scientists, however, can be obtained from national soil science societies and the International Union of Soil Sciences (IUSS). Van Baren et al. (2000) linked the number of IUSS members to total inhabitants and the agricultural land area for different countries. This information has been updated with recent figures from the national soil science societies (Table 5.1).

According to the IUSS membership data, the USA has the largest number of soil scientists (approximately 4000), followed by Germany (2311) and India (1846). Clearly, in all these countries there may be a few more soil scientists as not all of them will be members of the national societies, and not all members of these societies are active soil scientists. Some of the numbers are very small and probably wrong (e.g. underestimates for Brazil and China) Switzerland has the highest number of soil scientists per capita; roughly one in twenty thousand Swiss is a member of their national soil science society. The lowest number per capita is found in Brazil, India, Mexico, South Africa, and Turkey where less than 2 in 1 million inhabitants are member of their national soil science societies. A high number of soil scientists per ha agricultural land is found in Germany, Japan, the Netherlands, South Korea and Switzerland. The lowest number of soil scientists per ha agricultural land is found in Australia, Brazil, China, Mexico, Russia, South Africa, and Turkey. Clearly, there are a lot of "chicken and egg" type of relationships in this table. There is a fairly direct relation between the share of GDP spent on research and development and the average grain yield; countries that spent more on research have higher yields. The relation between the share of GDP spent on research and development and GDP per capita is strong, richer countries spend more money on research and vice versa. Also, GDP per capita relates very well to the number of soil scientists in a country. Richer countries have more soil scientists per capita. The total number of members of a national soil science society is well-correlated ($R^2 = 0.7^{***}$) with the number of inhabitants in a country. Also, members and the total area under agriculture are fairly well-correlated ($R^2 = 0.5^{**}$); countries with large areas under agriculture often have more soil scientists.

Table 5.1 Total members for each national soil science society, per unit land area, GDP per capita, research and development spending as % of GDP, and average grain yield per ha, for 30 countries

Country	Total no of members	Number of members per million inhabitants	Number of members per 1000 km^2 total area	Number of members per 1000 km^2 agricultural land	GDP per capita US$	R&D as % of GDP	Grain yield Mg ha^{-1}
Australia	496	24.7	0.1	0.1	32 000	1.70	1.9
Austria	183	22.4	2.2	5.4	32 900	2.33	5.4
Belgium	250	24.1	8.2	3.0	31 900	1.90	8.0
Brazil	200	1.1	0.0	0.1	8 400	0.98	2.4
Canada	255	7.8	0.0	0.4	32 900	1.93	2.8
China	1000	0.8	0.1	0.2	6 300	1.44	4.0
Denmark	70	12.9	1.6	2.6	33 400	2.63	6.2
Finland	203	38.9	0.6	9.0	30 600	3.46	3.5
France	413	6.8	0.7	1.4	30 000	2.16	7.2
Germany	2311	28.0	6.5	13.6	29 800	2.49	6.5
Hungary	200	20.0	2.1	3.4	16 100	0.88	3.6
India	1846	1.7	0.6	1.0	3 400	0.85	1.9
Israel	50	8.0	2.4	8.8	22 300	4.46	2.4
Italy	223	3.8	0.7	1.5	28 400	1.14	4.9
Japan	800	6.3	2.1	15.5	30 700	3.15	4.3
Mexico	50	0.5	0.0	0.0	10 100	0.40	2.7
Netherlands	409	24.9	10.0	21.2	30 600	1.85	7.9
New Zealand	63	15.6	0.2	0.4	24 200	1.16	6.3
Norway	50	10.9	0.2	4.8	42 400	1.75	3.9
Poland	200	5.2	0.6	1.2	12 700	0.58	2.5

Table 5.1 (continued)

Country	Total no of members	Number of members per million inhabitants	Number of members per 1000 km^2 total area	Number of members per 1000 km^2 agricultural land	GDP per capita US$	R&D as % of GDP	Grain yield Mg ha^{-1}
Portugal	237	22.4	2.6	6.3	18 600	0.78	2.7
Russia	360	2.5	0.0	0.2	10 700	1.17	1.6
South Africa	69	1.6	0.1	0.1	12 100	0.76	2.9
South Korea	520	10.7	5.3	17.6	20 400	2.64	4.4
Spain	408	10.1	0.8	1.4	25 200	1.11	3.6
Switzerland	354	47.3	8.6	23.2	35 300	2.57	6.6
Thailand	300	4.7	0.6	1.6	8 300	0.26	2.0
Turkey	50	0.7	0.1	0.1	7 900	0.66	2.3
UK	815	13.5	3.3	4.8	30 900	1.89	7.2
USA	4000	13.5	0.4	1.0	42 000	2.68	5.8

Member data from IUSS (R. Harris) 2004 and 2005; Agricultural land use 2003 from FAO FAOSTAT; Population data from 2005 www.census.gov; GDP data from 2003–2004 www.cia.gov; Research and Development data from World Bank and UNESCO 2001–2004; Grain yield equivalent data from 2000, FAOSTAT, FAO

5.3 Soil Maps – Europe

The first soil maps in Europe were made in the 1800s. They were mostly produced for agricultural purposes or the taxation of rural lands and emphasised surficial geology and the degree of weathering of the regolith (Stremme, 1997). The first task of the International Society of Soil Science (ISSS but since 1998: International Union of Soil Sciences, IUSS) established in Rome in 1924 was to produce a Soil Map of Europe. This was necessary to overcome language problems and differences in mapping approaches. Countries in Eastern Europe followed the Russian (= V.V. Dokuchaev and N.M. Sibirtsev) approach of mapping soils as natural bodies, whereas those in Western Europe – where systematic mapping started later – followed a more geological approach.

The first European soil map was published in 1928 at a scale of 1:10 million. It has 27 map units and was based on the geological map at a scale of 1:5 million. The map was presented at the first World Congress of Soil Science in Washington D.C. (USA) in 1927 where it was agreed to produce a more detailed map at a scale of 1:2.5 million. This map was published in 1937 and has 43 map units grouped in seven sets.

The next Soil Map of Europe was produced 30 years later by the Food and Agricultural Organization of the UN and the EEC (FAO, 1965). Systems of classification used in the different countries varied in approach but for the 1965 map a uniform legend was presented. The legend consists of soil associations composed of soil units. Many countries only started systematic soil surveys after the Second World War, and this map contains the best soil distribution information available at that time. The earlier maps of the 1920s and 1930s were not used in the 1965 European soil maps or in successive efforts.

The next European soil maps were produced in the framework of the 1:5 million Soil Map of the World for which preparation began in 1961 as a joint project of FAO and UNESCO following a recommendation of the ISSS. The complete set of the Soil Map of the World was presented at the 10th World Congress of Soil Science in Moscow in 1974, and publication of all 19 map sheets was achieved by 1981. The European volume was the last sheet that was published. Most of the European region was covered by systematic soil surveys but only Iceland, the northern parts of Finland and the USSR and Turkey in Asia were mapped at the reconnaissance level. On the 1:5 million map, units are associations of soil units (e.g. Arenosols, Vertisols) which were assigned texture and topography (slope class) of the dominant soil. Phases (e.g. stony, phreatic) are superimposed on the map units. At last, in 1985 a 1:1 million soil map of Europe was published (Commission of the European Communities, 1985). The map has 20 soil orders (major soil groups) like Gleysols or Luvisols and more than 60 great groups or soil units (e.g. Chromic Cambisols). The legend of the map shows 312 different map units which consist of associations of soil units occurring within the limits of a mappable physiographic entity.

The completion of the Soil Map of the World by FAO-UNESCO has been one of the main contributions of the ISSS (Van Baren et al., 2000) and has since its completion found wide applications, like for example: assessment of desertification, delineation of major agro-ecological zones, evaluation of global land degradation, calculation of population supporting capacity, creation of a World Reference Base for Soil Resources, and the creation of a digital global Soils and Terrain Database (SOTER) (Oldeman and van Engelen, 1993).

5.3.1 Three Generations of Soil Maps

Table 5.2 summarises the available soil maps for Europe. The first generation maps of the 1920s and 1930s have a strong agro-geological base and were based on limited soil survey work. These soil maps stimulated soil survey and research in most European countries of which the fruits were harvested for the second generation of European soil maps (1965–1985). These developed in the heydays of soil survey and were based on hundreds of detailed national and regional maps. The second generation is now being replaced by a third generation of maps – digital soil maps in which full use is made of existing soil and other information with advancements in GIS, remote sensing and quick and accurate soil observations using a range of sensors (McBratney et al., 2003).

When comparing the 1965 soil map of Europe to the 1981 and 1985 maps there is much more detail reflected in the number of mapping units and scale of the map. All three soil maps summarize soil survey activities in each country and soil survey was at its zenith. Then the mapping was more or less over as most governments withdrew their support for multi-purpose and generic soil surveys. As a result, little traditional soil mapping (auger, spade, stereoscope) took place since the 1980s.

The coverage of detailed (1:50 000) and exploratory (1:250 000) maps was linked to the size of 31 countries in Europe. It seems that smaller countries have better coverage of both exploratory and detailed soil maps (Fig. 5.1). About 45% of the countries have complete coverage with detailed soil maps and 9 countries in Europe have less than 20% of their total area mapped at 1:50 000 and these include France, Spain and Sweden. More than 60% of the countries have 100% coverage with exploratory soil maps.

Table 5.2 Soil maps of Europe, their scale, number of legend units and map sheets (Hartemink, 2006b)

Year of publication	Map scale	Number of map units	Number of map sheets	Reference
1928	1:10 million	27	1	Stremme (1928)
1937	1:2.5 million	43	12	Stremme (1937)
1965	1:2.5 million	34	6	FAO (1965)
1981	1:5 million	> 700	2	FAO-Unesco (1981)
1985	1:1 million	312	7	Commission of the European Communities (1985)
2005	1:1–1:6.5 million	163	17	European Soil Bureau Network of the European Commission (2005)

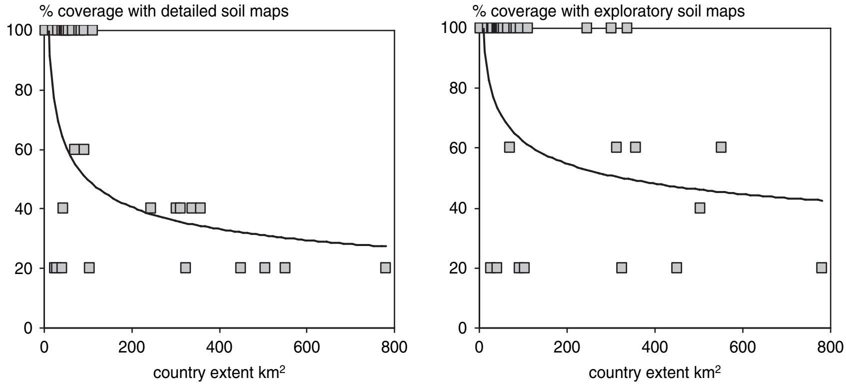


Fig. 5.1 Relation between size of 31 EU countries and the coverage with detailed soil maps (<1:50 000) and exploratory soil maps (>1:250 000)

Table 5.3 Correlation (R^2) between coverage of soil maps at a scale of 1:50 000 or 1:250 000 and country size, total population and population density of 31 European countries. Data extracted from: European Soil Bureau Network of the European Commission (2005)

	Size of the country	Total population	Population density
1:50 00 soil maps	0.364*	0.358*	0.743***
1:250 000 soil maps	0.472**	0.492**	0.795***

*,**,*** indicates significance at $P < 0.05$, $P < 0.01$ and $P < 0.001$, resp.

Correlation between a country's population density and the availability of soil maps is fairly strong and highly significant (Table 5.3). Small, highly-populated countries in Europe have the most detailed soil information; large, less densely populated countries like France, UK and Germany generally have less detailed soil maps. Correlation between number of soil scientists in 1998 or 2005 and the coverage of soil maps in 2005 is poor. However, the coverage of soil maps in 2005 is related to the number of soil scientists in 1974 (Fig. 5.2). The larger the number of soil scientists per unit area of agricultural land in 1974, the greater the coverage of soil maps, particularly exploratory soil maps in 2005.

5.4 Soil Maps – Low and Middle Income Countries

Coverage of soil maps in low and middle income countries is shown in Table 5.4. The Gambia, Jamaica and Trinidad & Tobago are covered with detailed soil maps (scale >1:25 000). About one-third of the countries have soil maps at a scale of 1:100 000–1:500 000 but these countries have hardly any maps on a larger scale, that is 1:50 000. Some countries like Congo and Algeria have very limited soil maps at any scale.

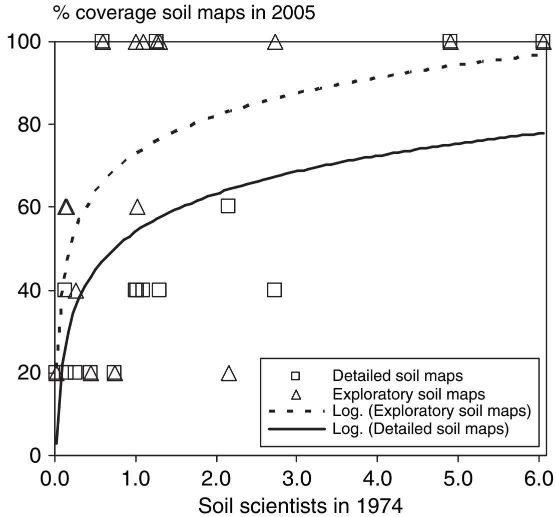


Fig. 5.2 The number of soil scientists per 1000 km^2 agricultural land in 1974 and the coverage with detailed soil maps (<1:50 000) and exploratory soil maps (>1:250 000) for 16 European countries in 2005

5.5 Soil Maps and GDP

Gross Domestic Product (GDP) per capita is often used as an indicator for a country's welfare. GDP is defined as the market value of all goods and services produced within a country in a given period of time; all other things being equal, standard of living tends to increase when GDP per capita increases. Economic data from UNDP was combined with data on the status of soil mapping in different countries (Nachtergaele and Van Ranst, 2003; Zinck, 1995). National coverage of soil maps is linked to GDP per capita (2001 data) for 44 countries (Fig. 5.3). Although the data are scattered, regression suggests that national coverage of exploratory soil maps is generally greater in the poorest countries and decreases with increasing GDP per capita; the national coverage of detailed soil maps tends to increase with increasing GDP. However, total coverage is very low in most of these countries (<20%).

GDP is larger in countries with increasing number of soil scientists (Fig. 5.4) – of course, the other way around is reasonable as well: the number of soil scientists increases with increasing GDP. More soil scientists per ha agricultural land often lead to higher yields (Fig. 5.5). Correlation between soil map density and grain yield equivalents was very low.

5.6 Discussion

The soil science community has not clearly demonstrated the benefits of soil science for society (Greenland, 1991; Hartemink, 2006a). If everyone were convinced that

Table 5.4 Coverage of soil surveys in 44 low and middle income countries. Adapated from Nachtergaele and Van Ranst (2003) and Zinck (1995)

	Small scale 1:500 000–±100 000 (%)	Medium scale 1:100 000–±50 000 (%)	Large scale ≤1:25 000 (%)
Algeria	0	5	5
Bangladesh	95	0	0
Benin	100	10	2
Botswana	40	5	0
Brazil	35	5	5
Burkina Faso	100	25	0
Burundi	100	0	0
Cameroon	30	5	1
China	100	100	0
Colombia	85	5	5
Congo	10	5	0
Costa Rica	100	20	5
Egypt	100	10	10
Gabon	30	0	0
Gambia	100	0	100
Ghana	95	0	0
India	80	0	0
Indonesia	40	10	0
Iran	0	10	10
Jamaica	0	100	100
Kenya	100	25	0
Malaysia	100	10	0
Mali	50	0	0
Mexico	75	40	0
Morocco	0	40	20
Myanmar (Burma)	100	20	2
Nigeria	70	35	0
Pakistan	85	3	0
Panama	50	0	0
Papua-New Guinea	5	10	0
Peru	50	0	0
Philippines	100	10	0
Rwanda	100	100	0
South Africa	70	0	0
Sri Lanka	100	10	2
Swaziland	100	10	5
Tanzania	50	0	0
Thailand	0	100	20
Togo	80	20	0
Trinidad-Tobago	0	0	100
Uganda	100	0	0
Uruguay	20	20	0
Venezuela	90	5	2
Vietnam	0	40	30

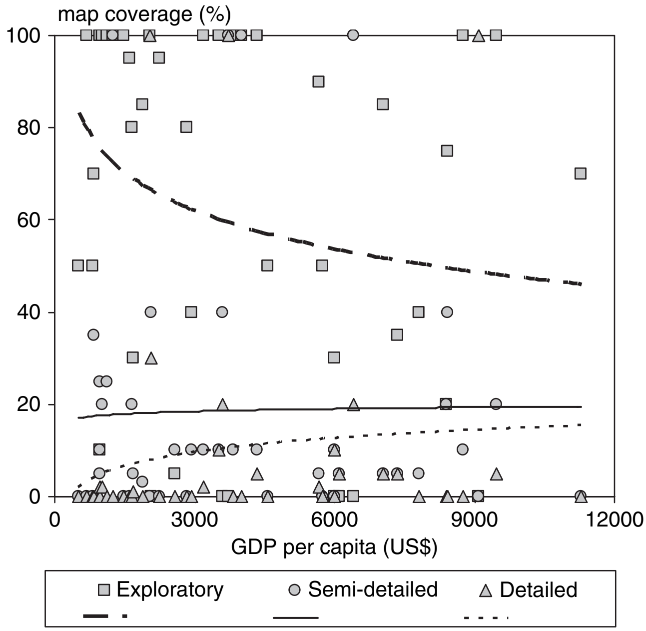


Fig. 5.3 Relation between GDP per capita (2001 data) and national coverage of exploratory soil maps (≈ 1:100 000–1:500 000), semi-detailed soil maps (≈ 1:100 000–1:50 000) and detailed soil maps (≈<1:25 000) of 44 low and middle income countries (<US$ 12 000 GDP per capita in 2001 – UNDP data)

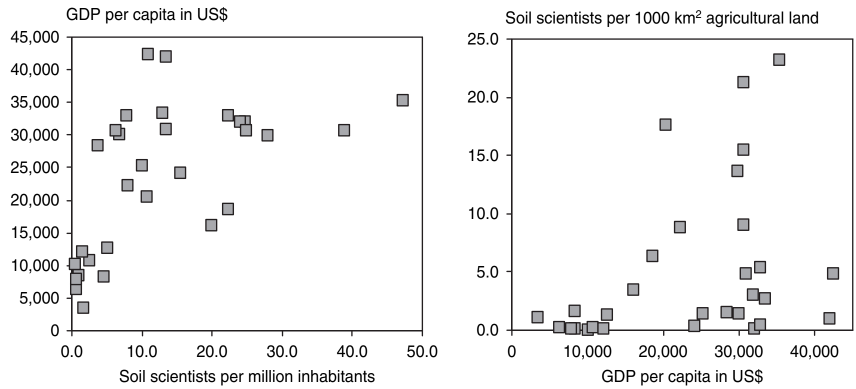


Fig. 5.4 Relationship between soil scientists per million inhabitants and GDP per capita; and between GDP per capita and the number of soil scientists per 1000 km^2 agricultural land

soil science is essential for human welfare perhaps this demonstration would not be needed (see also Chapter 3), but I fear that is not the case. Decreasing funds for soil research, and the inability of the soil science community to effectively show the benefits has resulted in fewer soil scientists and far fewer students in many universities across the globe but in particular in the USA and Canada (Baveye et al., 2006).

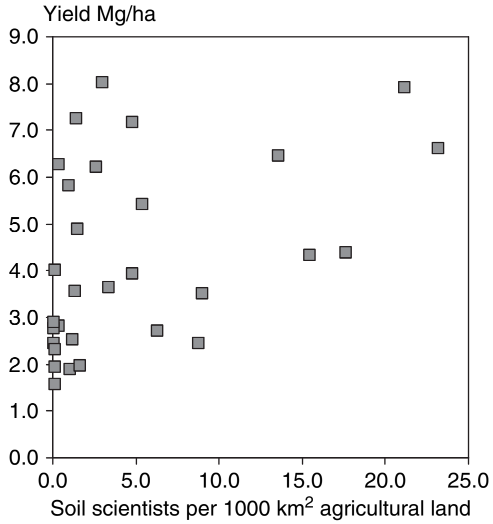


Fig. 5.5 Relation between soil scientists per 1000 km^2 agricultural land and average grain yields (FAOSTAT data)

Soil science has distinctly different foci in the developed compared to developing countries (Hartemink, 2002) whereas Chapters 22–34 show that there are many similarities in approach and problems that are to be tackled. This chapter has shown that there are large differences in these regions in terms of soil data density. Some poor countries have very good data and maps (for example, Rwanda); some rich countries are poor in data. For both groups it is imperative that the usefulness of soil information for development is illustrated. The development of digital soil maps takes places in both regions (Lagacherie et al., 2006) and it is important that appropriate indicators are sought to illustrate the effectiveness of digital soil maps. The methodologies (Chapters 13–21) exist and are continuously being developed but the extent of digital soil maps needs further increasing (Section 1.7.1).

This chapter has show a link between soil science information (maps) and GDP and some other variables. Although there are many confounding factors, these relations warrant further investigation. Clearly, few people would deny the use and relevance of soil information for agricultural project development or urban city planning but quantifying the economic benefits remains a large task (Giasson et al., 2000). Previous studies (e.g. Klingebiel, 1966; Dent and Young, 1981) have shown high benefit-cost ratios for soil surveys but these studies were based on traditional survey methods. Bui (2007) gives some cost estimates for traditional soil surveys in Australia and compared these ratios for producing digital soil maps. Costs for traditional surveys were AU\$12–28 per km^2 whereas the digital approaches were costing AU\$3–9 per km^2. Most of the reduction in costs was achieved by fewer person years to map the same area. These costs excluded infrastructure or the computer network and the costs for training a new generation of digital soil surveyors. She concluded that in a country with an aged workforce the uptake of digital soil mapping will be slow (Bui, 2007) – see also the Foreword of this book. This applies to many countries reviewed in this chapter. The real challenge for digital soil mapping is not the aging workforce but the training of a fresh generation of soil scientists that will widely use and advance new techniques (Section 6.4). Given the benefits of soils and soil information for humankind and a nation's wealth and income that new generation has a bright future ahead.

Acknowledgments I thank Dr. F. Nachtergaele of the Food and Agricultural Organization of the UN for some of the data presented in this paper, and Dr. D. Dent and ir N. Batjes of ISRIC – World Soil Information in Wageningen for comments on the draft.

References

Baveye, P., Jacobson, A.R., Allaire, S.E., Tandarich, J.P. and Bryant, R.B., 2006. Whither goes soil science in the United States and Canada? Soil Science, 171: 501–518.

Bie, S.W. and Beckett, P.H.T., 1970. The costs of soil survey. Soils and Fertilizers, 33: 203–217.

Bie, S.W., Ulph, A. and Beckett, P.H.T., 1973. Calculating the economic benefits of soil survey. Journal of Soil Science, 24: 429–435.

Bui, E.N., 2007. A review of digital soil mapping in Australia. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital soil mapping – an introductory perspective. Developments in Soil Science, Vol. 31. Elsevier Science, Amsterdam, pp. 25–37.

Commission of the European Communities, 1985. Soil map of the European Communities 1: 1000000. Directorate-General for Agriculture Coordination of Agricultural Research. EEC, Luxembourg.

Dent, D. and Young, A., 1981. Soil survey and land evaluation. E & FN Spon, London, Melbourne, 278p.

European Soil Bureau Network of the European Commission, 2005. Soil atlas of Europe. Office for Official Publications of the European Communities, Luxembourg.

FAO-Unesco, 1981. Soil map of the world, volume V Europe, 1: 5 000 000, Rome.

FAO, 1965. Soil map of Europe, Carte des Sols de L'Europe, Mapa de Suelos de Europa. 6 Map sheets, explanatory text by R. Dudal, R. Tavernier and D. Osmond, Rome.

Giasson, E., van Es, C., van Wambeke, A. and Bryant, R.B., 2000. Assessing the economic value of soil information using decision analysis techniques. Soil Science 165: 971–978.

Greenland, D.J., 1991. The contributions of soil science to society – past, present, and future. Soil Science 151: 19–23.

Hartemink, A.E., 2002. Soil science in tropical and temperate regions – some differences and similarities. Advances in Agronomy 77: 269–292.

Hartemink, A.E. (Editor), 2006a. The future of soil science. IUSS, Wageningen, 165 pp.

Hartemink, A.E., 2006b. Soil maps of Europe. Book review of: Soil Atlas of Europe, by European Soil Bureau Network of the European Commission, 2005. Journal of Environmental Quality 35: 952–955.

Klingebiel, A.A., 1966. Costs and returns of soil surveys. Soil Conservation 32: 3–6.

Lagacherie, P., McBratney, A.B. and Voltz, M. (Editors), 2006. Digital soil mapping – an introductory perspective. Developments in Soil Science, Vol. 31. Elsevier Science, Amsterdam.

McBratney, A.B., 1992. On variation, uncertainty and informatics in environmental soil management. Australian Journal of Soil Research 30: 913–935.

McBratney, A.B., Santos, M.L.M. and Minasny, B., 2003. On digital soil mapping. Geoderma 117: 3–52.

Nachtergaele, F.O. and Van Ranst, E., 2003. Qualitative and quantitative aspects of soil databases in tropical countries. In: G. Stoops (Editor), Evolution of tropical soil science: Past and future. Koninklijke Academie voor Overzeese Wetenschappen, Brussels, pp. 107–126.

Oldeman, L.R. and van Engelen, V.W.P., 1993. A world soils and terrain digital database (SOTER) – an improved assessment of land resources. Geoderma 60: 309–325.

Sachs, J.D., 2005. The end of poverty – economic possibilities for our time. The Penguin Press, New York.

Stremme, H., 1928. General map of the soils of Europe (Ogolna Mapa Gleb Europy). International Society of Soil Science, Warszawa.

Stremme, H., 1937. International soil map of Europe, 1:2 500 000. Gea Verlag, Berlin.

Stremme, H.E., 1997. Preparation of the collaborative soil maps of Europe, 1927 and 1937. In: D.H. Yaalon and S. Berkowicz (Eds.), History of Soil Science: International Perspectives. Advances in Geoecology. Catena Verlag, Armelgasse 11/35447 Reiskirchen/Germany, pp. 145–158.

van Baren, H., Hartemink, A.E. and Tinker, P.B., 2000. 75 years the International Society of Soil Science. Geoderma 96: 1–18.

Zinck, J.A., 1995. Introduction. In: J.A. Zinck (Ed.), Soil Survey: Perspectives and Strategies for the 21st Century. ITC, Enschede and FAO, Rome, pp. 1–6.

Part II
Dealing with Limited Spatial Data Infrastructures

Chapter 6
Digital Soil Mapping as a Component of Data Renewal for Areas with Sparse Soil Data Infrastructures

D.G. Rossiter

Abstract This chapter introduces the concepts of data *rescue* of legacy soil surveys, here defined as a simple conversion to archival format by scanning or direct entry into a database, and data *renewal*, here defined as the process of bringing these surveys up to modern standards by taking advantage of technological and conceptual advances in geoinformation technology. This is especially important in areas with sparse soil data infrastructure, as it is both more likely that the data will be lost and less likely that a new survey can be commissioned. Digital Soil Mapping (DSM) techniques, although designed for new surveys, can play an important role in data rescue and renewal, in particular as geodetic control for a GIS coverage, as a medium-resolution elevation model (DEM) and derived terrain parameters to adjust terrain-related boundaries, and synoptic satellite imagery to adjust vegetation or landuse-related boundaries. The semantic issues raised by soil-landscape modelling within DSM are especially important for data renewal and integration with supplementary surveys. As with DSM in general, a data renewal exercise may require cultural and institutional change in traditional soil survey organization.

6.1 Introduction

The World is full of unused, even forgotten, soil geographic information in the form of soil surveys (also called “soil resource inventories”). Some are sitting forlornly on a library shelf; some are turning to dust in a storage cabinet; some are in the private collections of deceased or retired soil scientists; and some are even in digital form and yet unused, perhaps on obsolete or decaying media. These are part of the legacy of previous generations of soil surveyors. Disasters – natural, man-made, and political – or simply inattention can destroy these forever. This is especially unfortunate in areas with sparse soil data infrastructure (“data-poor areas”), as it is both more likely that the data will be lost and more unlikely that a new survey can

D.G. Rossiter
International Institute for Geo-information Science & Earth Observation (ITC), Postbus 6, 7500 AA Enschede, The Netherlands
e-mail: rossiter@itc.nl

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

be commissioned. Even if new data is to be collected, it is wasteful and scientifically irresponsible to throw away existing data. Indeed, standard practice when beginning a new survey is to incorporate previous work as much as possible. Legacy data can also be used to provide a baseline for longitudinal studies of soil change (e.g. Bellamy et al., 2005).

Similarly under used are "card catalogue" data bases of point observations with no or only descriptive geo-reference, such as laboratory reports and profile descriptions. Their rescue and renewal (e.g. Batjes, 1995; Batjes et al., 2007) is not facilitated by DSM techniques; on the contrary, their rescue facilitates DSM. They will not be further discussed here; examples are given in Chapters 23 and 27.

Soil surveys in data-poor areas are often of good quality, in the sense that the surveyors understood soil-landscape relations, made field observations, often in difficult conditions, and synthesized these into map units. Many former colonies were surveyed by the colonial power; more recently, development projects often included soil surveys. These surveys differed in their objectives, standards, and concepts; but still provide valuable, often irreplacable, information. However, very little soil geographic information on these areas is available in digital form (Rossiter, 2004). The author keeps up-to-date with on-line data by regularly searching the Web to update the *Compendium of On-Line Soil Survey Information* (Rossiter, 2007), and this situation has not appreciably changed in the past six years with the notable exception of the data rescue efforts of ISRIC and EUSB (see below).

During recent years there has been a movement among soil geographers towards so-called *Digital Soil Mapping* (DSM), defined as "soil resource assessment using geographic information systems (GIS), i.e. the production of digital soil property and class maps with the constraint of limited relatively expensive fieldwork and subsequent laboratory analysis" (McBratney et al., 2003). This concept has also been termed *predictive soil mapping* (PSM): "the development of a numerical or statistical model of the relationship among environmental variables and soil properties, which is then applied to a geographic data base to create a predictive map" (Scull et al., 2003, p. 171); this narrow definition can be expanded to non-statistical models (e.g. expert judgement) and to soil classes or even behaviour.

Whether called "digital" or "predictive", the movement has resulted in a large number of powerful geo-processing, statistical, and conceptual methods for describing the distribution of soils on landscapes; this was after all the aim of those who made what we now call "legacy" soil surveys. This chapter presents a conceptual framework for using DSM techniques to support data renewal, with emphasis on areas with sparse soil data infrastructures, and soil maps following the discrete model of spatial variation (DSMV) (Heuvelink and Webster, 2001), i.e. area-class ("polygon") maps, since these are the great majority of legacy soil maps. Although most DSM products are based on the continuous model (CMSV), Ibañez et al. (2005) have recently argued that the DSMV, which corresponds to natural soil bodies (or their associations) with relatively sharp transitions in the landscape, is often an efficient information carrier and as such is an efficient stratification for DSM sampling or selection of covariates.

6.2 Data Renewal

The effort to locate and catalog historical records has been termed *data archaeology* and the effort to preserve them *data rescue*. An example is the project to find and scan paper meteorological[1] and oceanographic[2] records by NOAA. For soils data, ISRIC-World Soil Information has a long-standing policy of collecting and cataloging all soil surveys, published or not. The recent intiative of ISRIC and the JRC of the EC to scan ISRIC's soil map collection on to DVD (Selvaradjou et al., 2005a, b, c) is an outstanding and most welcome example of data rescue.

The step after data rescue is *data renewal*: existing information is not only saved from extinction, but is also put in modern form and made useful. This has been termed *resurrection* of legacy soil surveys by Dent and Ahmed (1995) in their work in the Gambia; we have chosen a more neutral term.

We propose that a *renewed* legacy soil survey be defined by the following elements:

1. The area-class delineations as a GIS coverage: geo-referenced and geodetically-correct to some specified accuracy, commensurate with the original mapping scale;
2. (If available) Sample points as a GIS coverage: geo-referenced and geodetically-correct as above;
3. Linked area-class and (if applicable) point attribute databases; in a sound relational structure;
4. A synoptic medium-resolution (10–30 m horizontal) multi-spectral image ("TM-type") as background; see the list in Section 2.2.1;
5. A medium-resolution (30–90 m horizontal) elevation model (DEM) and derived terrain parameters (slope gradient and aspect, curvatures, wetness index etc.) (Oksanen and Sarjakoski, 2005; Wilson, 2000) as background and supplemental terrain properties;
6. Metadata explaining the semantics of all terms, either internally or by reference to external standards such as soil classification systems and laboratory procedures;
7. A users' guide for soil specialists and any interpretations for other uses from the original survey;
8. Licensing and usage restrictions, if any;
9. Can be downloaded via internet and/or requested on optical disk;
10. Integrated into the relevant national or international Geospatial Data Infrastructure (GSDI) (Groot and McLaughlin, 2000), if such exists.

Thus the legacy soil survey becomes a modern digital product, directly accessible and useful for a wide variety of uses, and with improved and assured quality. The

[1] http://docs.lib.noaa.gov/rescue/data_rescue_home.html; accessed 17-October-2007

[2] http://www.nodc.noaa.gov/General/NODC-dataexch/NODC-godar.html; accessed 17-October-2007

closest examples to this ideal are provided Canada (Coote and MacDonald, 2000) and the USA (Soil Survey Staff, 2007), although in these cases the terrain and imagery information are the responsibility of other agencies.

The renewal exercise would ideally include some new fieldwork; if extensive enough to check DSM products, this would supersede the legacy survey. However, in resource-poor organizations this will likely be impossible in the short term; so we concentrate on renewal as just one step past rescue, and perhaps a step on the way to a new or updated survey.

Many recent technological and conceptual developments which facilitate renewal are within the financial reach of most soil survey organisations and projects; some of these are specifically DSM-related and will be discussed in detail in the next section.

1. Increasingly-powerful geo-information technology (computer systems and programs) is available to facilitate the work in many ways; some programs are open-source or otherwise free (e.g. GRASS, SAGA, R, ILWIS) while others are low-cost commercial (e.g. IDRISI);
2. Major conceptual advances in data structuring, data modelling, and metadata standards; improved tools to facilitate these;
3. Medium-resolution satellite images are easily available at either free or very low cost and are easily imported to many programs;
4. Medium-resolution terrain models are also available at either free or very low cost, most notably from the Shuttle Radar Topography Mission (SRTM) (Rabus et al., 2003); import and processing is moving from the development to the routine stage, especially now that corrected products are available (Gorokhovich and Voustianiouk, 2006);
5. Other wide-area medium-resolution digital coverages relevant to soil geography may be available at either free or very low cost (i.e. other themes have often been more quickly made digital than the soils theme); examples are geology, land cover and climate;
6. GPS receivers for geodetic control, field checking, and supplementary sampling are cheap; even consumer-grade receivers are accurate enough for medium-scale soil maps;
7. Field computers running mobile GIS applications and integrated with GPS for field checking.

6.3 DSM to Assist in Data Renewal

DSM methods can contribute directly to three facets of data renewal:(1) geodetic control for a GIS coverage; (2) a medium-resolution elevation model (DEM) and derived terrain parameters to adjust terrain-related boundaries; (3) the semantic issues inherent in conceptual soil-landscape modelling.

Geodetic control Almost all legacy soil surveys have major problems with geodetic control. For example, in the USA it proved impossible to rectify surveys published on semi-controlled corrected photomosaics (D'Avelo and McLeese, 1998).

A geometrically-correct topographic base with no georeference can be georeferenced by a small set of field GPS points at well-defined cultural features (e.g. road junctions); with attention to changes since the original publication date. The same base may be available as a digital coverage, since national base mapping is usually more advanced than thematic mapping. However, the topographic base may be poorly-reproduced, in which case the map unit boundaries must be re-compiled as explained below.

Single unrectified airphotos with soil map unit boundaries can be converted to orthophotos and then mosaicked, at the same time rectifying the boundaries (Rossiter and Hengl, 2002); this depends on a geodetically-correct and georeferenced topographic base or else a fairly dense set of GPS points which can be located on the photograph; and (in hilly terrain) a coarse-resolution DEM.

Maps on un- or semi-controlled photomosaic bases are impossible to correct. Here the soil boundaries must be re-compiled by eye, by a compiler with expert knowledge of the soil-landscape relations, onto a geometrically-correct and georeferenced topographic base; this was the situation in the USA when building the SSURGO database (D'Avelo and McLeese, 1998). Ideally, the original surveyor is available to communicate the expert knowledge; second-best is a soil survey report that documents these relations for the compiler. The recompiled lines should follow the original lines within map accuracy standards (e.g. Davis et al., 1981, pp. 556–560); however the original lines may well represent imprecise boundaries either of location or concept (Lagacherie et al., 1996), in which case being moved somewhat (depending on the scale of the soil landscape and map) is not so serious. At 1:50 000 scale, a 0.5 mm line covers 25 m; this implicit boundary width is double the maximum location accuracy (Forbes et al., 1982).

Using terrain models to adjust boundaries Do the lines, even if rectified, optimally separate contrasting soil bodies? In many surveys it was common practice for draftsmen (not soil surveyors) to transfer soil boundaries by eye (perhaps with assistance from a non-precision optical device) from field sheets to base map. Without the surveyor's expert eye, soil boundaries that followed obvious landscape features were not reproduced correctly. The use of low-cost medium resolution DEM (SRTM, ASTER or SPOT) can help an expert compiler to manually adjust the lines to ensure this. The DEM resolution matches well with the mapping scale of most legacy data. Further, if the original map was based on soil-landscape relations, the DEM is being used at the most obvious landscape transitions, where uncertainties in derivatives are least. The procedure is as follows:

1. Describe the probable association of soil-landscape units from the original survey with terrain parameters (e.g. slope gradient and aspect, curvatures, wetness indices); these should be described in the original survey report, e.g. by block diagrams or in the map unit descriptions;
2. Compute relevant parameters from the DEM;
3. Classify the terrain according to the pre-defined cluster centres;
4. Adjust lines by hand.

This process is knowledge-intensive and interactive; however still much cheaper than original survey. This is related to DSM methods for identifying land elements which may be then used to delineate soil boundaries (MacMillan et al., 2000; Schmidt and Hewitt, 2004). However, here we assume that the original lines are thematically correct, albeit geodetically suboptimal; the aim is not to identify them but to adjust them. Although terrain parameters derivated from medium-resolution DEM may not be precise enough for DSM (Thompson et al., 2001), here we are only interested in adjusting boundaries of medium-scale polygons maps. The adjustment is typically manual (see for example Section 29.2); however an automatic approach may be successful, as shown in Fig. 18.4.

Heterogeneous soil-landscape units present additional difficulties. An automatic landform classifier will usually be able to delineate more homogeneous elements smaller than the minimum legible delineation (MLD) of the published survey (Hengl and Rossiter, 2003); the compiler may choose to delineate these if it is clear to which member of the association each applies; this has been termed "de-convolution of the soil-landscape paradigm elaborated during a soil survey" (Bui and Moran, 2001).

Using imagery to adjust boundaries In some environments landcover boundaries, either of natural vegetation or of land uses, match soil map unit boundaries in the conceptual model of the original survey. Thus a geometrically-correct landcover map can be used to adjust these boundaries, in the same way as terrain parameters are used to adjust terrain-related boundaries. The procedure is as follows:

1. Describe the probable association of soil map units from the original survey with natural land cover (e.g. coastal mangroves, dune vegetation) or landuse (e.g. irrigation scheme); these should be described in the original survey report;
2. Identify these landcover classes on imagery (typically synoptic medium-resolution multi-spectral) by conventional landcover classification techniques;
3. Adjust soil map unit lines by hand.

Imagery products other than landcover classifications may also be useful, for example a vegetation index to find high-and low-moisture areas, locally related to slope position (Fig. 16.1) and therefore soil depth and development, in strongly-sloping terrain. In (semi-)arid environments it may be possible to directly map contrasting parent materials (Fig. 16.2) or salt-affected soils (Metternicht and Zinck, 2003); these boundaries should divide soil map units, so can also be used to adjust legacy lines.

Semantics Both the original soil survey and DSM projects can be viewed as "knowledge systems" of soil-landscape relations (Bui, 2004). Thus semantics are central to re-interpreting the knowledge presented (sometimes implicitly) in legacy surveys. One component of DSM is multi-source, multi-scale data integration of legacy soil surveys and ancillary data (Krol et al., 2006); this has led to active research in semantics and ontology, i.e. the meaning of terms in each data set. Even when dealing with just one data set (the legacy soil survey) there are often difficult questions as to the meaning of legend categories, both the type such as "association" and the classes themselves. The meaning must be extracted from a close reading of the soil survey report and supporting documents, and structuring this into a database design.

6.4 Cultural and Institutional Challenges

The conceptual and technical aspects of data renewal discussed above are carried out within a cultural and institutional context where such an exercise usually implies a major challenge. Especial difficulties are establishing a work flow, quality assurance procedures, and training. Once data are renewed, inter-institutional issues arise, most notably data sharing, data and metadata standards, and responsibilities within a geospatial data infrastructure (de Man, 2006). This is new territory for most organisations responsible for soil information. In some countries where the traditional soil survey organization is not modernizing their approach to soil mapping, another organization from the same government (e.g. a space agency or a planning ministry) has stepped in; however in this case there is usually little appreciation for, or skill in interpreting, legacy data.

6.5 A Small Example

Perhaps the region of the world with the greatest need for data renewal, and at the same time a high potential for this approach, is sub-Saharan Africa. The recently-issued "European Digital Archive of Soil Maps (EuDASM) – Soil Maps of Africa" (Selvaradjou et al., 2005a) shows how much map information there is: ISRIC's 40-year collection, which was opportunistic rather than systematic, consists of over 2000 maps, about half of them of soil maps, the rest being of related themes such as geology, geomorphology, land use, and agro-ecological zoning. Although many of these maps are at reconaissance scales, important regions are covered at medium and even detailed scales. More such maps can be rescued from the successors of former colonial soil survey organizations, development projects, and land investment schemes.

Figure 6.1A shows a small portion of a representative soil map from the colonial period (Kenya Department of Agriculture, 1961) included in the EuDASM project. It was published by the Department of Agriculture, Kenya (then still a British colony and protectorate) but financed in large part and supported technically by the International Co-operation Administration of the USA.

This 1:50 000 map has a fairly high level of cartographic detail, matched with medium categorical detail. The geodetic control is good: both geographic and grid coördinates are printed in the margin, although no intersections are marked; the map projection is not given explicitly, rather the topographic base map is named, from which some detective work (Mugnier, 2003) reveals that this was developed in the East African War System transverse Mercator projection, belt I on the Arc

(a)

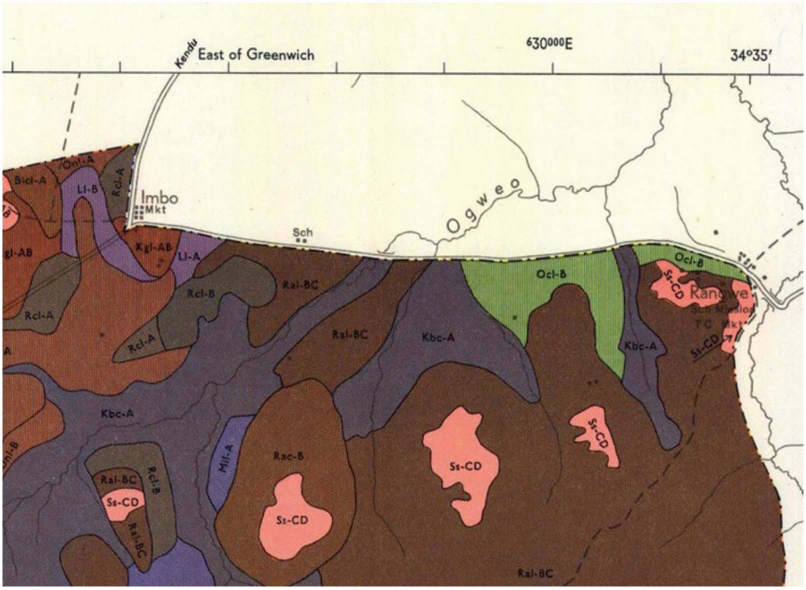


Fig. 6.1A Portion of East Konyango (Kenya) soil map (See also Plate 2 in the Colour Plate Section)

(b)

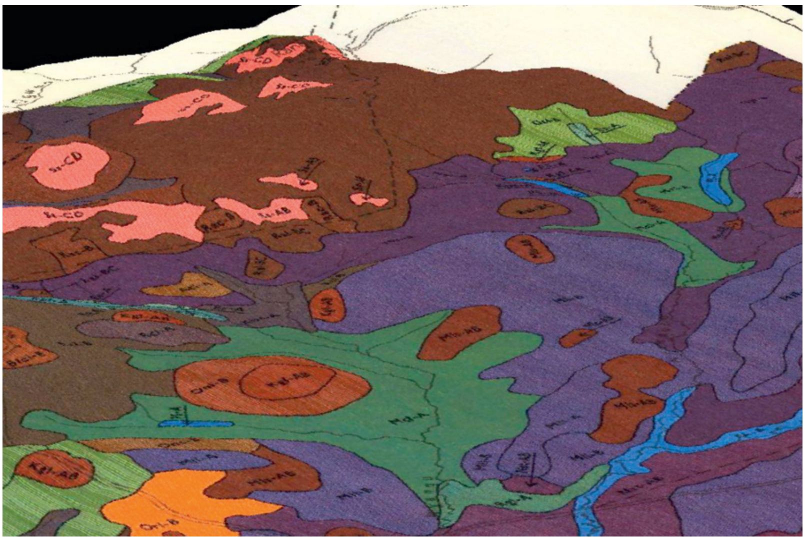


Fig. 6.1B Perspective view, soil map draped on SRTM elevation model (See also Plate 3 in the Colour Plate Section)

1950 datum. Major roads and rivers are shown, so that GPS control points could be established if necessary; however in this case there is sufficient information on the map projection to allow exact transformation. The map units are slope phases of 28 soil types, which are topsoil texture variants of locally-defined soil series. These series are defined by horizonation and profile development. A 66-page appendix gives extensive analytical data on one georeferenced (by map grid) soil profile per soil type. The survey has some elements in its origin that suggest it may be of high quality. The survey team was made up of senior soil surveyors with earlier experience in Kenya, and also included the top American soil classifier of the time (Thorp) and a well-known analytical chemist (Mehlich). The report is well-written and includes keen observations of the landscape and probable soil genesis, as well as current and potential land uses.

This survey was available for the compilation of the Kenya 1:1 000 000 SOTER database (ISRIC, 1995); however a renewal of this survey would keep the original scale and categorical detail. To renew this survey we must overcome some typical limitations:

- All people with direct knowledge of this survey are retired or deceased;
- The soil series are classified in an American system (Thorp and Smith, 1949) that was obsolete even at the time it was used and which, as the authors acknowledge, was not well-suited to tropical soils;
- Only one profile per soil type was analyzed, and these with (today) outdated methods;
- Profiles were located subjectively to be the "most representative"; we thus rely on the surveyor's expert reading of the landscape;
- The base map series was state-of-the-art for its time but uses an obsolete datum, projection and grid;
- Soil boundaries were inferred from transects spaced from approximately 800–1600 m apart, and were transferred from unrectified air photos to the topographic sheet by an optical transfer scope.

DSM techniques can directly address some of these limitations, and assist with others. The most obvious link to DSM is terrain modelling. Figure 6.1B shows a portion of the map geo-referenced and draped on a 90 m horizontal-resolution SRTM elevation model; this is the highest resolution currently available for much of the world. Many of the map unit boundaries appear to follow major landscape features which can be identified on a terrain model. For example, unit *Mcl-A* (Marinde clay loam, 0–3% slopes) falls nicely in the concave colluvial slopes bordering the Olunga river (flowing towards the viewer in the figure); the units of *Ss-CD* (Stony land from siliceous rocks, 8–20% slopes) are on the highest sideslopes within the extensive (and easily-cultivated) Rangwe sandy loams, 3–13% slopes of map unit *Rsl-BC*. Both *Ss-CD* and *Rsl-BC* were formed on coarse-textured residuum from the Nyanzan rhyolite; these units thus occur in the resistant hills of the highest elevations. Clearly, the surveyors had a sound landscape interpretation and transferred it with sufficient geodetic accuracy to the published map. The report also emphasizes landscape relations and includes two block diagrams showing these. Because of the strong map unit – landform relations, DSM techniques such as drainage network extraction and supervised landform classification should be successful in refining map unit boundaries; the latter may even be successful in disaggregating compound map units.

This analysis is obviously preliminary; many other DSM aspects remain to be explored, in particular semantic matching of the soil units as described with soil properties, and the use of the profile observations.

6.6 Conclusion

It is clear that DSM can benefit from legacy soil surveys which provide raw data (in the form of point observations) and reality checks for validation (in the form of interpreted polygons), as well as the surveyor's concept of soil geography as revealed in the soil survey report; Fig. 2.1 shows "existing soil maps" as a primary input layer to the *scorpan* approach (McBratney et al., 2003) to DSM. This chapter has shown that data renewal exercises can also benefit from DSM techniques. Soil survey organizations may choose to renew legacy surveys as a step towards a fully-digital mapping exercise, thereby gaining experience in some DSM technologies. They may also choose to renew legacy surveys as an end product (the area-class map with linked attribute database) that is familiar to their clients. In either case, legacy soil surveys are too information-rich to be left mouldering on the shelf.

References

Batjes, N. H. (Ed.), 1995. A homogenized soil data file for global environmental research: a subset of FAO, ISRIC and NRCS profiles (Version 1.0). Working Paper and Preprint 95/10b. International Soil Reference and Information Centre (ISRIC), Wageningen (NL).

Batjes, N. H., Al-Adamat, R., Bhattacharyya, T., Bernoux, M., Cerri, C. E. P., Gicheru, P., Kamoni, P., Milne, E., Pal, D. K., Rawajfih, Z., 2007. Preparation of consistent soil data sets for modelling purposes: Secondary SOTER data for four case study areas. Agriculture, Ecosystems & Environment 122, 26–34.

Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M., Kirk, G. J. D., 2005. Carbon losses from all soils across England and Wales 1978–2003. Nature 437, 245–248.

Bui, E. N., 2004. Soil survey as a knowledge system. Geoderma 120, 17–26.

Bui, E. N., Moran, C. J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103, 79–94.

Coote, D., MacDonald, K., 2000. The Canadian soil database. In: Sumner, M. (Ed.), Handbook of Soil Science. CRC Press, Boca Raton, FL, pp. H-41–H-51.

D'Avelo, T. P., McLeese, R. L., 1998. Why are those lines placed where they are?: An investigation of soil map recompilation methods. Soil Survey Horizons 39, 119–126.

Davis, R. E., Foote, F. S., Anderson, J. M., Mikhail, E. M., 1981. Surveying: Theory and Practice, 6th Edition. McGraw-Hill, NewYork.

de Man, W. H. E., 2006. Understanding SDI: complexity and institutionalization. International Journal of Geographical Information Science 20, 329–343.

Dent, D. L., Ahmed, F. B., 1995. Resurrection of soil surveys: a case study of the acid sulphate soils of The Gambia. I. Data validation, taxonomic and mapping units. Soil Use and Management 11, 69–76.

Forbes, T., Rossiter, D., Van Wambeke, A., 1982. Guidelines for evaluating the adequacy of soil resource inventories. SMSS Technical Monograph 4. Cornell University Department of Agronomy, Ithaca, New York.

Gorokhovich, Y., Voustianiouk, A., 2006. Accuracy assessment of the processed SRTM-based elevation data by CGIAR using field data from USA and Thailand and its relation to the terrain characteristics. Remote Sensing of Environment 104, 409–415.

Groot, R., McLaughlin, J. (Eds.), 2000. Geospatial Data Infrastructure: Concepts, Cases, and Good Practice. Oxford University Press, Oxford.

Hengl,T., Rossiter, D., 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Science Society of America Journal 67, 1810–1822.

Heuvelink, G. B. M., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Ibañez, J. J., Ruiz Ramos, M., Zinck, J. A., Bru, A., 2005. Classical pedology questioned and defended. Eurasian Soil Science 38(Supplement 1), 75–80.

ISRIC, 1995. SOTER database of Kenya. Kenya Soil Survey, ISRIC – World Soil Information, Wageningen (NL), Nairobi.

Kenya Departmentof Agriculture, 1961. Soil survey of the East Konyango area: parts of survey of Kenya, G.S.G.S., 4786 series, sheets 129/II, 129/IV, 130/I & 120/III. Map & Report, Government Printer, Nairobi.

Krol, B., Rossiter, D. G., Siderius,W., 2006. Ontology-based multi-source data integration for predictive soil mapping. In: Lagacherie, P., McBratney, A., Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science 31. Elsevier, Amsterdam, 119–133.

Lagacherie, P., Andrieux, P., Bouzigues, R., 1996. Fuzziness and uncertainty of soil boundaries: from reality to coding in GIS. In: Burrough, P. A., Frank, A. U. (Eds.), Geographic objects with indeterminate boundaries. GISDATA 2. Taylor & Francis, London, pp. 275–286.

MacMillan, R. A., Pettapiece, W. W., Nolan, S. C., Goddard, T. W., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Sets and Systems 113, 81–109.

McBratney, A. B., Mendonça Santos, M. L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Metternicht, G. I., Zinck, J. A., 2003. Remote sensing of soil salinity: potentials and constraints. Remote Sensing of Environment 85, 1–20.

Mugnier, C., 2003. Grids and datums: Republic of Kenya. Photogrammetric Engineering & Remote Sensing 69, 593–597.

Oksanen, J., Sarjakoski, T., 2005. Error propagation of DEM-based surface derivatives. Computers & Geosciences 31, 1015–1027.

Rabus, B., Eineder, M., Roth, A., Bamler, R., 2003. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. ISPRS Journal of Photogrammetry and Remote Sensing 57, 241–262.

Rossiter, D. G., 2004. Digital soil resource inventories: status and prospects. Soil Use and Management 20, 296–301.

Rossiter, D. G., 2007. A compendium of on-line soil survey information. On-line document, `http://www.itc.nl/personal/rossiter/research/rsrch_ss.html`; Accessed: 17-October-2007.

Rossiter, D. G., Hengl, T., 2002. Creating geometrically-correct photo-interpretations, photomosaics, and base maps for a project GIS; 3rd revised version. Technical note, International Institute for Geo-information Science and Earth Observation (ITC), `http://www.itc.nl/personal/rossiter/teach/sis/TN_Georef_wFigs_Screen_V3.pdf`; Accessed: 16-October-2007.

Schmidt, J., Hewitt, A., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. Geoderma 121, 243–256.

Scull, P., Franklin, J., Chadwick, O., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography 27, 171–197.

Selvaradjou, S.-K., Montanarella, L., Spaargaren, O., Dent, D., 2005a. European Digital Archive of Soil Maps (EuDASM) – Soil Maps of Africa. EUR 21657EN. Office for Official Publications of the European Communities, Luxembourg.

Selvaradjou, S.-K., Montanarella, L., Spaargaren, O., Dent, D., 2005b. European Digital Archive of Soil Maps (EuDASM) – Soil Maps of Asia. EUR 21823 EN. Office for Official Publications of the European Communities, Luxembourg.

Selvaradjou, S.-K., Montanarella, L., Spaargaren, O., Dent, D., 2005c. European Digital Archive of Soil Maps (EuDASM) – Soil Maps of Latin America and Carribean Islands. EUR 21822 EN. Office for Official Publications of the European Communities, Luxembourg.

Soil Survey Staff, 2007. Soil Survey Geographic (SSURGO) Database. `http://soildatamart.nrcs.usda.gov;` Accessed: 16-October-2007.

Thompson, J. A., Bell, J. C., Butler, C. A., 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. Geoderma 100, 67–89.

Thorp, J., Smith, G., 1949. Higher categories of soil classification. Soil Science 67, 117–126.

Wilson, J. P., Gallant, J. (Eds.), 2000. Terrain analysis: principles and applications. Wiley & Sons, New York.

Chapter 7
Challenges to Digital Soil Mapping

**J.W. Hempel, R.D. Hammer, A.C. Moore, J.C. Bell,
J.A. Thompson and M.L. Golden,**

Abstract Digital Soil Mapping to capture or determine categorical or property information has undergone a tremendous increase in capability and application during the past decade. Many successful technologies have been developed through research activities worldwide, including generalized linear models, classification and regression trees, neural networks, fuzzy systems, expert systems, and geostatistical methods and applications. These technologies have matured beyond a research activity and have potential for use by soil scientists to more accurately, consistently, and efficiently define soil categories and soil properties based on digital proxies to soil-forming factors. These applications for producing soil maps are now poised to become production tools to either update older soil survey information or to produce soil information on previously unmapped areas.

As these technologies move into the mainstream for producing soil survey information, there are challenges that must be overcome. The community of soil scientists and soil classifiers engaged in producing soil information must become familiar with the technologies and their potential uses and limitations. More importantly, the users of soil survey information must be convinced of the relevance and applicability of maps and data that appear different from the "traditional" products with which they have become familiar. New challenges include developing acceptable standards and procedures for the production and quality control and interpretation of the information that relates to agricultural, engineering, forestry and other soil-landscape uses.

7.1 Introduction

Increases in computational power, informational technology, Geographic Information System (GIS) techniques and the ever-increasing quantities and sophistication of geographic data, have led to extraordinary advances in producing digital soil

J.W. Hempel
USDA Natural Resources Conservation Service, Morgantown, WV 26506
e-mail: jon.hempel@wv.usda.gov

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

survey information. These methodologies can revolutionize the approach for producing soil survey information as well as the ways interpretations are made and information is applied for human uses (see Chapters 1, 10). Soil information can now be modeled to define the natural continuum of soil properties and can potentially illustrate gradational boundaries of soil caterogorical maps. Soil maps produced with digital techniques have the potential to spatially define complex landscape soil components that are now represented by a more generalized, artificially homogeneous polygon structure.

Digital techniques utilize quantifiable methodologies that produce more consistent outputs and are potentially reproducible. (see Sections 25.2, 34.3, 34.4). There is growing documentation indicating that digital soil mapping techniques offer the potential to greatly accelerate the rate of map production. (see Chapter 9) The potential for these production gains as well as the potential for defining soil information at levels of detail that many would have never dreamed possible is creating a great deal of excitement within the soil science community. The combination of enhanced detail and faster production offers increases in efficiency and precision that will offset the inevitable reduction in numbers of field soil scientists as the National Cooperative Soil Survey (NCSS) moves into the Major Land Resource Area (MLRA) based maintenance and update phase. The ability to produce digital soil survey products at a variety of scales also offers the potential to produce user and site-specific soil attribute maps for individual customers and land uses. This flexibility has important applications for Homeland Security, urban land use planning, and other recently emerging soil survey applications.

Although digital soil mapping techniques potentially offer extraordinary improvements and efficiencies in producing more precise soil survey information, traditional soil survey persists as the most popular form of soil mapping and inventory, and in many cases is the only manner in which the highly variable nature of the soil landscape is catalogued (Scull, et al. 2005). The "traditional" methodology produces soil maps whose classes are spatially homogenous with crisp boundaries separating adjacent polygons (mapping units).

The traditional soil survey map simplifies the complex continuous distribution of soil types across a landscape by depicting the individual soil map units as discrete polygons with definite boundaries (Zhu et al., 2004). The simplification (homogenization) was mandated by the scale at which surveys were conducted, time constraints of mapping, and being restricted to displaying the soil pattern as two-dimensional lines on paper. The frustration for most field soil scientists was that they *knew*, beyond all doubt, that the soil-landscape was more complex than they were able to depict.

Hudson (1992) precisely categorized what the soil scientist knew, but could not map, as "tacit knowledge," and pointed out that capturing this information had important implications. Early attempts to address the "unmappable" heterogeneity included: (1) creation of block diagrams to show soil patterns at finer scales; (2) creating a "range of characteristics" for each soil series; (3) describing inclusions, both limiting and non-limiting; and (4) listing "geographically associated soils" in the soil survey manuscript. These acknowledgements that "more is out there" were not in sufficient detail, nor were they in formats suitable for use, by individuals who lacked field mapping or field research experience. Zhu (1999) recently has begun efforts to develop software whose purpose is to allow soil scientists to apply tacit knowledge to new soil survey products.

When comparing the two processes for producing soil survey information, digitally produced soil maps appear to have a distinct advantage over the traditionally produced soil survey maps with respect to the potential accuracy and amount and kinds of detail that can be portrayed.

7.1.1 History and Background of Traditionally Produced Soil Maps

Applications of Soil Survey information derived from the traditional data have a long standing history in assisting with land use decisions related to agriculture, forestry and engineering. The production of polygon based maps that define soils and their spatial extents are the most common means by which this complex natural resource information has been portrayed to the traditional soil survey user community.

These soil surveys make readily available, important chemical and physical property information such as depths, permeabilities, available water-holding capacity, soil reaction (pH), and shrink-swell potential in a layer by layer format. Other important soil and water properties such as flooding and high water table depths, depths to bedrock, potential frost action and corrosivity risks to concrete and steel are easily accessed. Engineering properties important to many applications are indexed, including: textures, sieve sizes, fragment size, liquid limits and plasticity indices. From these properties many interpretations can be inferred. The most common interpretations have included crop yields, woodland management and productivity, suitabilities for environmental plantings, recreational development, wildlife habitat, building site development, sanitary facilities, construction materials and water management. Producing customized interpretations to meet the evolving needs of an ever-widening customer base is a goal of the U.S. National Cooperative Soil Survey Program.

Soil survey information in its current format has a long history in the United States. "The authorization for documenting, cataloguing and presenting soil survey information began in the United States by the US Department of Agriculture Appropriations Act for fiscal year 1896. This act provided funding for an investigation of the relations of soils to climate and organic life and to measuring textures and compositions of soils in the field and laboratory" (Soil Survey Division Staff, 1993, p. 11). Reports on field investigations and soil mapping were developed by the US Department of Agriculture as early as 1899. Early soil surveys investigated the potentials of soils primarily for agricultural and forestry. The evolution of soil survey included recognition of the soil as a natural body whose chemical and physical properties were both records of the historical development of soil profiles, and the keys to classifying and identifying soils with common properties and responses to human impacts (Arnold, 1983). An important key to the long-term success of soil survey has been the incorporation into field, laboratory and reporting methods of new techniques and technologies that enhance soil scientists' abilities to map and interpret soils and to deliver information to soil survey users.

7.2 Current Progress and Application

"Soil surveys published between 1920 and 1930 reveal a marked transition from earlier concepts to give emphasis to soil profiles and soils as independent bodies" (Soil Survey Division Staff, 1993, p. 11). During the mid-1930's soil surveys began to use aerial photographs as base maps. This greatly increased the accuracy of plotting soil boundaries. This process of mapping over the next 70 years evolved into what is described as the "modern soil survey." This collection of published soil surveys is nearly complete in the continental United States. Most soil survey information uses county boundaries as the "Soil Survey Area". The distribution and status of completed soil surveys in the US and its territories is portrayed in Fig. 7.1.

The basic soil unit, which is the foundation for information products, is the "soil map unit." The soil map is delineated on the orthophoto basemap (Fig. 7.2) (Soil Survey of Barbour County Map Manuscript, Alabama, 1995). The number inside the map unit delineation on the orthophoto is the map symbol, which identifies the

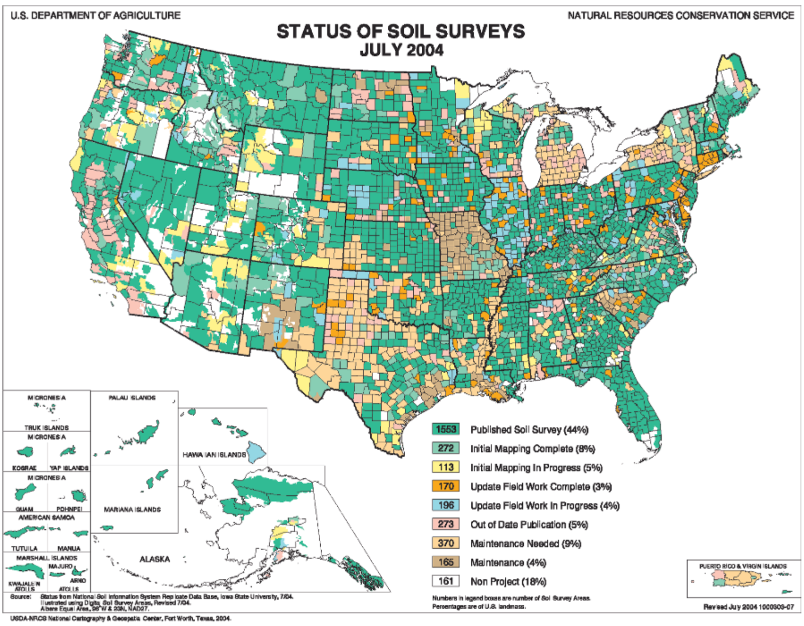


Fig. 7.1 Status of soil surveys in the United States. (approximately 1:2,000,000) (See also Plate 4 in the Colour Plate Section)

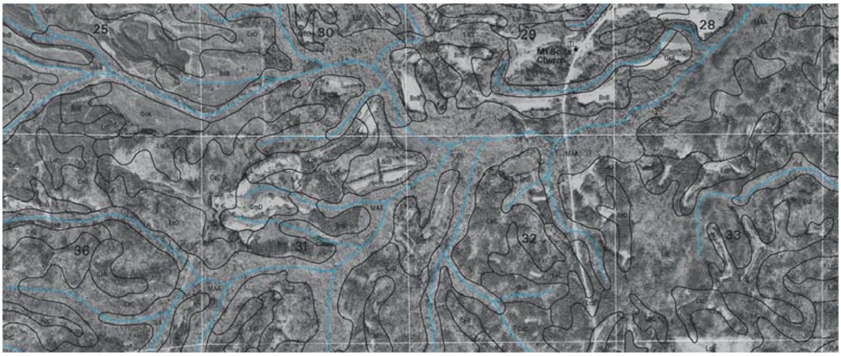


Fig. 7.2 Individual soil mapping units as portrayed in the Soil Survey of Barbour County, Alabama (Trayvick, 1995) (See also Plate 5 in the Colour Plate Section)

mapped soil in the soil survey legend. The mapping unit may be an individual soil series, a phase of a soil series or a combination of soils identified at the series or other taxonomic level. The map unit is a collection of polygons that represents soils as similar as possible at the mapping scale.

This paper "hard copy" of the modern soil survey is being replaced by a digital product known as SSURGO (Soil Survey Geographic Overlay). This electronic soil survey data layer is being digitized from paper manuscript soil surveys in the United States and is available either by downloading from the Soil Data Mart http://soildatamart.nrcs.usda.gov/ or the Geospatial Data Gateway http://datagateway.nrcs.usda.gov/NextPage.asp for use in geographic information systems. The SSURGO information is also in an online format from the Web Soil Survey http://websoilsurvey.nrcs.usda.gov/app/. All completed soil surveys digitized to the SSURGO standards and specifications are available through this website. Fig. 7.3 is an example of a digitized soil survey from the Web Soil Survey.

Although costs for generating SSURGO are not fully documented at this time, digitizing to produce SSURGO certified databases will cost in excess of $100M

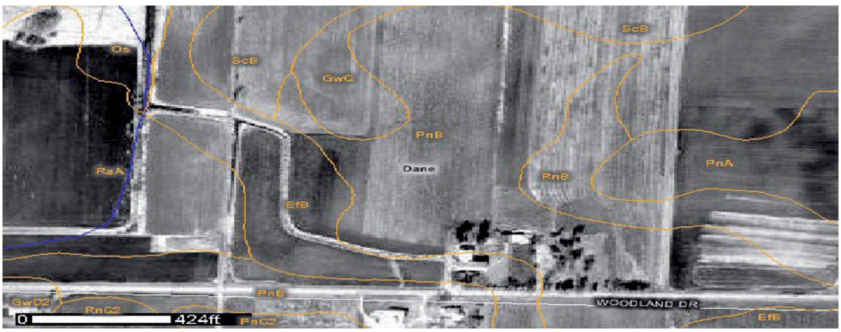


Fig. 7.3 Portion of the Dane County, Wisconsin soil survey (Glocker and Patzer, 1978) produced from the Web Soil Survey (See also Plate 6 in the Colour Plate Section)

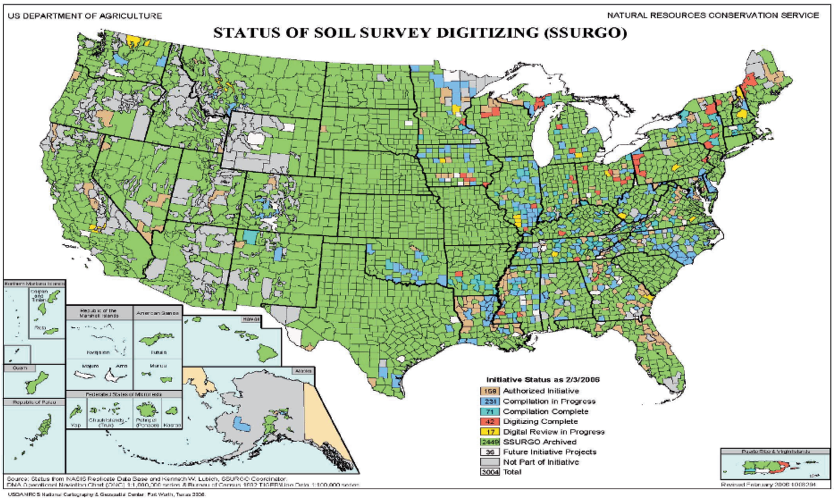


Fig. 7.4 Status of SSURGO digitizing in the United States. (approximately 1:2,000,000) (See also Plate 7 in the Colour Plate Section)

to fully complete all counties in the United States that have modern soil surveys. Figure 7.4 is a map of the status of SSURGO digitizing projects. Counties colored green are completed.

Information from the previously discussed examples of the modern soil survey database has been applied, used and become institutionalized as the foundation for many county, state, federal and private programs. The generated data and information is being used for environmental regulations, guidelines, laws and soil interpretations.

The Farm Security and Rural Investment Act of 2002 is landmark legislation for conservation funding and for focusing on environmental issues (Farm Bill Conservation Provisions, 2006). The United States Department of Agriculture Farm Bill programs depend directly on soil survey information for eligibility and administration of a variety of programs including the specific programs listed in Table 7.1.

The newest of these farm bill programs, the Conservation Security Program (CSP), has been in effect since fiscal year 2003. Through fiscal year 2005, CSP has funded $181M of conservation payments. The US Congress will cap total expenditures for CSP at $6.037 billion (between FY-2005 and FY-2014) (Why a Watershed Approach Is Being Used, 2006).

Soil survey information from traditional soil survey products has been used as background data for tax assessment and evaluation. The state of Iowa uses soil survey information for agricultural land assessment and valuation. In counties or townships in which field work on a modern soil survey has been completed since January 1, 1949, the assessor places emphasis upon the results of the survey in spreading the valuation among individual parcels of such agricultural property (Code of Iowa).

The state of Iowa also uses soil survey or derived data for other assessments. Corn Suitability Ratings (CSR) is an index procedure developed in Iowa to rate

Table 7.1 Federal programs that rely on or utilize soil survey information for planning and resource allocation

Program	Costs	Contracts	Acres
Agricultural Management Assistance (AMA)-provides cost-share and incentive payments to agricultural producers to voluntarily address issues, such as water management, water quality and erosion control by incorporating conservation practices into their farming operations	$23.4M	1899	392,000
The Conservation Security Program (CSP)-voluntary conservation program that supports ongoing stewardship of private agricultural lands by providing payments for maintaining and enhancing natural resources.	$181M	14975	12.1M
The Environmental Quality Incentives Program (EQIP)-voluntary program that provides assistance to farmers and ranchers who face threats to soil, water, air and related natural resources on their land	$1.08B	117,625	51.5M
The Farm and Ranch Lands Protection Program (FRPP)-voluntary program that helps farmers and ranchers keep their land in agriculture	$321M	2080	367,510
The Conservation Reserve Program (CRP)-voluntary program that grants annual rental payments for set-aside highly erodible land	$1.66B	34.7M	665,101M
The Grassland Reserve Program (GRP)-voluntary program that helps landowners and operators restore and protect grassland, including rangeland, pastureland, shrubland and other certain lands, while maintaining areas as grazing lands.	$111M	908,400	3.036M
The Wetlands Reserve Program (WRP)-voluntary program that provides technical and financial assistance to eligible landowners to address wetland, wildlife habitat, soil, water, and related natural resource concerns on private lands in an environmentally beneficial and cost effective manner.	$759M	544860	3.013M

each soil for its potential row-crop productivity. An annual conservation and land preservation tax is imposed on each acre of agricultural land that is converted to a commercial, industrial, or residential use on or after the effective date of this law. The tax rate is based on the CSR of each acre of the converted agricultural land. If the CSR of the acre is less than fifty, no tax is imposed on that acre. If the CSR of the acre is fifty or higher, the tax is fifty dollars plus one dollar for each whole rating unit in excess of fifty. Soil survey information for this and other interpretations, including smart growth, land evaluations, site assessments and regional planning, is an indispensable tool for environmental planning in many communities.

7.3 Developments of Standards and Procedures

The US National Soil Survey Handbook (U.S. Department of Agriculture, Natural Resources Conservation Service, 2005) identifies standards and specifications for producing traditional soil survey information and provides specific background for conducting all phases of a soil survey. These standards and specifications are revised regularly, and since the early 1950's have been the accepted protocol for producing soil surveys by the US National Cooperative Soil Survey.

Different modeling methodologies documented for producing digital soil survey maps include geostatistical and statistical methods, decision tree analyses, and expert systems (see Chapters 2, 13, 18, 19). No accepted procedures, standards or protocols have been established for producing digital soil survey products in the US. Leaders in the NCSS will be reluctant to accept digital mapping without standards in place. A historical strength of the NCSS has been the continuous application of standards and quality assurance.

In addition, application of digital soil mapping in some areas of the US at scales required by USDA programs is hampered by inadequate digital elevation model (DEM) resolution, inherent errors in DEMs, lack of complete understanding of appropriate DEM resolution and neighborhood analysis for accurate calculation of landscape derivatives for various landforms and landscapes. Additionally inadequate relationships between the distribution of soil classes or properties and digital proxies for soil forming factors (e.g. glacial till plains), areas where terrain is not a primary control for spatial distribution of soils, (e.g. floodplains, glaciofluvial deposits), terrain-based predictive models probably will have limited use. The lack of accepted standard methods within the US NCSS for production of digital soil maps and the lack of accepted procedures within the US NCSS for verifying the accuracy of digitally produced soil maps along with dearth of readily available and easily accessible methods for viewing, analyzing, and interpreting raster soil survey data will also hinder the acceptance of digital soil mapping methodologies.

Currently, there is an aggressive research agenda to investigate the potentials of digital soil mapping and how it best fits into the National Cooperative Soil Survey Program in the US. Applied digital soil mapping research in the areas of quantitative analysis of terrain, spectral analysis (see Chapter 16) and other data using GIS, remote sensing and statistical methods is on-going at a number of universities in the US to better understand how these processes can be used to create or update soil survey information. In addition, NRCS soil scientists in many locations are applying DSM techniques to produce either predictive soil maps or intermediate data layers used to support the polygon based soil mapping process.

7.4 Conclusion

Great potential exists for digital soil mapping techniques to improve the consistency, accuracy, detail and speed at which soil survey information is produced. These techniques can be applied to updating existing soil survey information and creating information in unmapped areas. This raster information can be aggregated into map units (polygons) and converted to vectors for inclusion into the national SSURGO dataset. The underlying raster information can be used to spatially locate inclusions in the map units, an important component lacking in the current database model.

Currently soil survey information in the USDA NRCS program is based on categories or soil classes. Digital Soil Mapping techniques offer the potential for producing soil property information that could be used for a wide variety of interpretive information not currently available to the user community. Soil property information can be very helpful in addressing climatic and environmental degradation issues.

Important issues must be resolved before these potentials can be realized. Research and clarification of important technical questions relating to input data must be resolved. Applications and standards are required before digital soil survey products can be used with the same confidence and acceptance enjoyed by the "historical" orthophoto based polygon maps. Currently, there is a lack of a clear direction within the NCSS for presenting, visualizing, and interpreting digitally produced soil data in its native format. The development of methodologies and standards for applications of digital soil survey for making and interpreting information is necessary for acceptance for users of soil survey information. Finally, the long history, investment in, institutionalization of and broad application of the current soil survey information will be obstacles for the acceptance of digital soil survey mapping techniques and the resulting data as viable options to traditional soil survey products and processes.

References

Farm Bill Conservation Provisions. US Department of Agriculture-Natural Resources Conservation Service. 19Mar. 2006 <http://www.nrcs.usda.gov/programs/farmbill/2002/>.

Why a Watershed Approach Is Being Used. US Department of Agriculture-Natural Resources Conservation Service. 19 Mar. 2006 <http://www.nrcs.usda.gov/programs/csp/watershed_approach_reasons.html>.

Arnold, R.W. 1983. Concepts of soils and pedology. In: L.P. Wilding, N.E. Smeck and G.F. Hall (eds.), Pedogenesis and Soil Taxonomy I. Concepts and Interactions. Elsevier, NY, pp. 1–21.

Code of Iowa, 441.21 Actual, assessed and taxable value. <http://www.legis.state.ia.us/IACODE/2003SUPPLEMENT/441/21.html.

Glocker, C.L. and R.A. Patzer. 1978. Soil Survey of Dane County, Wisconsin. U.S. Dept. Agric., Soil Cons. Serv. in coop. with College of Agric. and Life Sci., Univ. of Wisconsin.

Hudson, B.D. 1992. The soil survey as a paradigm-based science. Soil Sci. Soc. Am. J. 56: 346–841.

Trayvick, J.C. 1995. Soil Survey of Barbour County, Alabama. U.S. Dept. Agric., Nat. Res. Cons.Serv. in coop. with Alabama Agric. Exp. Stat. and Alabama Soil and Water Cons. Dist.

Scull, P., J. Franklin, O.a. Chadwick, and D. McArthur. 2003. Predictive soil mapping: A review. Progress in Physical Geography 27: 171–197.

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

Soil Survey of Barbour County Map Manuscript, Alabama, 1995. Clayton North Quadrangle, Sheet 11 of 27 Alabama.

U.S. Department of Agriculture, Natural Resources Conservation Service, 2005. National Soil Survey Handbook, title 430-VI. [Online] Available: http://soils.usda.gov/technical/handbook/.

Zhu, A.X. 1999. A personal construct-based knowledge acquisition process for natural resource mapping. Int. J. Geog. Inform. Sci. 13: 119–141.

Zhu, A.X., A.C. Moore, M.P. Smith, J. Liu, J.E. Burt, F. Qi, D. Simonson, J. Hempel and K. Lubich, 2004. Advances in information technology for soil surveys: The SoLIM effort. In: H. Eswaran, P Vijarnsorn, T. Vearasilp, and E. Padmanabhan (eds.), Innovative Techniques in Soil Survey: Developing the Foundation for a New Generation of Soil Resource Inventories and Their Utilization. Land Development Department, Chattuchak, Bangkok, Thailand, pp. 25–42.

Chapter 8
Mapping Potassium Availability from Limited Soil Profile Data in Brazil

R.B. Prado, V.M. Benites, P.L.O.A. Machado, J.C. Polidoro, R.O. Dart and A. Naumov

Abstract Brazilian soils are generally acidic with low base saturation and low plant available potassium (K). Potassium fertilizers play important role in production costs and farmers receive no governmental subsidies. Strategies are needed to improve potassium fertilizer delivery to different regions in Brazil and to establish affordable prices and balanced potassium consumption. For such strategy, it is necessary to take into account the different soil classes with its varying K levels. The purpose of this study was to map soil K in Brazil considering the different biomes and applying techniques to reduce problems caused by limited soil profile data. A soil profile data set was constructed from the soil archives of Embrapa Soils, Rio de Janeiro, Brazil. Descriptive statistics was performed on K levels in different soil classes and biomes. The different soil K levels were grouped in intervals and mapped using ArcGIS 9.1 tools from ESRI. Brazil's soil map and biome map at 1:5,000,000 scale were used in the geoprocessing. Our results showed that mapping soil K levels based on soil survey reports at the regional scale is difficult because of limitations in georeferencing and spatial distribution of soil profiles. However, this mapping would help fertilizer distribution planning in Brazil.

8.1 Introduction

In 2005, Embrapa Soils and the International Potash Institute (IPI) started to organize soil data for mapping the plant potassium availability of several soil surveys of the National Soil Archives of Embrapa Soils. There was a need for optimum regional distribution of K fertilizers in Brazil, which are mostly imported. Optimized fertilizer distribution would help fertilizer delivery to farmers at lower costs. Presently, up to 40% of the total production cost of grain crops is due to inorganic fertilizers

R.B. Prado
Embrapa Soils, Rua Jardim Botanico, 1024, Jardim Botanico, Rio de Janeiro, RJ, Brazil, CEP 22460-000
e-mail: rachel@cnps.embrapa.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

(Bernardi et al., 2002). Brazilian farmers receive no governmental subsidies and the prices of inorganic fertilizers inhibit them to apply adequate and balanced amounts of fertilizer.

Mapping soil K availability is relevant because potassium is the second largest plant nutrient taken up by the main crop plants grown in Brazil such as soybean, coffee, common beans, cotton (Bernardi et al., 2002). Thus, balanced potassium fertilization of soils is essential to avoid both plant and soil degradation. In Brazil, crop plants are mostly grown in the Mata Atlantica and Cerrado biomes, whose soils are originally K deficient.

Since 1990s, compared to other plant nutrients the consumption of potassium fertilizer has been the largest in Brazil, and the recent growth of agribusiness have promoted potassium fertilizer use, particularly in the Cerrado region (Mascarenhas et al., 2004). Also, Lopes (2005) has estimated an increase of K_2O consumption in Brazil from 3.65 million t in 2003 to 5.2 million t in 2010 leading to US$ 6.08 million for importation expenditures in 7 years.

One of the main reasons for the change in agricultural production is the change in the consumer patterns combined with environmentally friendly technologies because of social and environmental concerns (Poulisse, 2003). There is a clear need for environmental sound techniques and food security with low environmental impact (Sanchez, 1997).

The use of geotechnologies in scientific studies may help to transform agriculture. Reliable data acquisition, organization in a georeferenced data base, and mapping are efficient tools. Plant nutrient mapping has been used at the farm level as part of precision agriculture (Bernardi et al., 2002). However, a regional approach is often needed for both governmental and private business purposes.

Available soil fertility data are sparse and collected at different scales. Interpolation is often difficult because of unreliable georeferencing. Problems related to sparse data infrastructures is further discussed in Chapter 2. The purpose of our study was to map K availability of Brazilian soils using soil survey reports, considering different biomes and applying techniques to reduce problems caused by limited soil profile data.

8.2 Material and Methods

A soil profile data set was constructed from the National Soil Archives of Embrapa Soils, Rio de Janeiro, Brazil. Exchangeable potassium data set were collected from 2600 soil profile (8500 soil horizons) surveyed in different biomes between 1958 and 2001.

8.2.1 Selection of Soil Profiles and Horizons

The predominant soil classes in Brazil are Acrisols, Luvisols, Ferralsols and Arenosols and soil K data were grouped for different depths (0–10, 0–20, 0–30, and 0–40 cm). On average accumulated soil K varies (>20%) with depth, particularly at 0–10 and 0–20 cm. Soil profiles selected were with soil horizons to 30 cm. Those soil horizons over 30 cm depth were considered up to 40 cm (i.e. 25–35 cm, 25–40 cm) and in soil profiles where the final depth was labeled 30 (i.e. 25–40 cm was considered 25–30 cm). Soil horizons deeper than 40 cm were excluded (e.g. 25–45 cm, 25–80 cm). Soil K concentration was restricted to a depth of 30 cm because most of plant root system are up to this depth.

8.2.2 Calculation K Level for Each Selected Soil Profile Data

For soil profiles showing two or more horizons the calculation of K level was as follows:

$$\mathrm{Kp} = ((e_1^{*}K_1) + (e_n^{*}K_n)/30)$$

where:

Kp = soil K in the profile (mg kg^{-1})
e = depth of soil horizon (cm)
K = soil K in the horizon (mg kg^{-1})
n = number of soil horizons until 30 cm depth

8.2.3 Verification and Data Exploratory Analysis

Prior to the exploratory analysis all data was verified and outliers were deleted. Values outside the limit of $X \pm 3^{*}SD$ (standard deviation) were considered an outlier. After the elimination of outliers a correlation analysis between K content and chemical and textural attributes was performed. A forward stepwise multiple correlation analysis was also performed to identify the attributes that showed strong influence over K content. Statistica 7.0 software was used in all analysis.

8.2.4 Map Units of Biomes and Soils

The single-part polygon was used as map unit and the polygon was generated from the intersection between soil class and biome type. It was assumed that soil K may vary as a function of soil class and biome. The Brazilian soil map (IBGE, 2001a) and Brazilian biome map (IBGE-MMA, 2005), both at the 1:5,000,000 scale, were intersected using the geoprocessing tool of ESRI ArcGIS 9.1. The area of the biome-soil unit was calculated using Albers Equal Area Conic projection and SAD69 datum of ESRI ArcGIS 9.1. Finally, two methods were applied to construct the soil K availability map.

The total number of map units from biomes and soil classes was 3332. Figure 8.1 shows the location of the 1764 K soil profiles. Most profiles are located in the

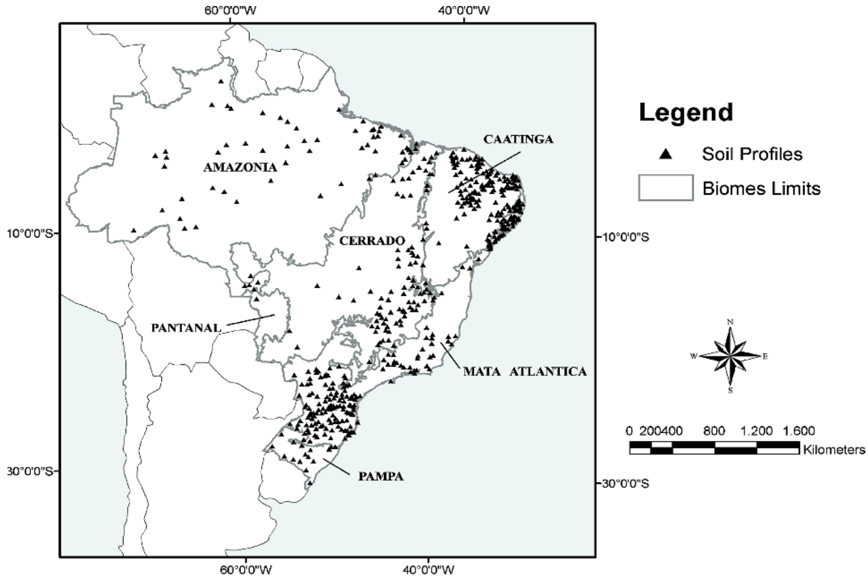


Fig. 8.1 Brazil's biome map at 1:5,000,000 and soil profiles distribution

northeastern, southeastern and southern Brazil and in the northern and central regions there were very few soil profiles available.

8.2.5 Mapping K Availability from Calculated K Soil Profiles with Associated Spatial Information

Most soil profiles of soil surveys were not fully georeferenced. Thus, the municipality central coordinates from the municipality grid map (IBGE, 2001b) at 1:250,000 scale were combined with the calculated K soil profile data, using the join tabular tool of ESRI ArcGIS 9.1. See also Chapter 29 for methodology used to convert printed into digital soil maps from the Amazon Region.

Soil profiles selected from the National Soil Archive were classified to the major soil group level (e.g. Ferralsol) whereas soil classification used in the 1:5,000,000 soil mapping was a third-order classification (e.g. Rhodic Ferralsols – Latossolo Vermelho distroferrico). Soil profile without representation at the 1:5,000,000 soil mapping were eliminated. This resulted in 482 calculated K soil profiles to be extrapolated to soil-biome units leading to 177 units with soil K data. Some units contained more than one soil profile and in these cases mean values were calculated for each soil unit.

Extrapolation of soil K from 177 soil-biome units to other units that were not mapped was done by the summarize tool of ESRI ArcGIS 9.1. A mean value of soil K was generated for each soil-biome unit group and this mean value was associated to other units not mapped using join tabular. In this step, the soil K of 1992 soil-biome units were estimated and used to map the K availability from soil profile data.

8.2.6 Mapping K Availability in the Third-Order of Acrisols, Luvisols and Ferralsols

At first, the percentage of Brazilian soil classes in different biomes was calculated. Only the K data of the Acrisols, Luvisols and Ferralsols were used because these major soil classes are dominant in Brazil (Acrisols and Luvisols 24% and Ferralsols 32% of Brazil area) and its agricultural suitability is generally high. The 296 (119 Acrisols and Luvisols, and 177 Ferralsols) profiles classified in the first-order were classified to third-order taking into account the humid colour of the first B horizon (10R-7.5R-5R-2.5YR = red; 5YR-7.5YR = yellow-red; 10YR-2.5Y-5 Y-7.5Y-10Y = yellow) and the base saturation percentage (<50% = dystric and >50% = eutrophic). Descriptive statistics for Acrisols, Luvisols and Ferralsols (third-order) K level and biome classes was calculated using Statistica. The results were associated by join tabular tool of ESRI ArcGIS 9.1 to the biome-soil units. In addition, 5 classes of interpretation of soil K were used to both mapping K availability, as suggested by Van Raij (1985). One more class was associated to no data related to biome-soil units with no associated soil K:

No data – biomes-soils units with no associated soil K
0–30 mg kg^{-1} soil – very low
30–60 mg kg^{-1} soil – low
60–120 mg kg^{-1} soil – medium
120–240 mg kg^{-1} soil – high
> 240 mg kg^{-1} soil – very high

8.3 Results and Discussion

Compared to Mata Atlantica, Caatinga, and Cerrado the soils of the Amazonia biome showed the lowest levels of available K. Results for both Pantanal and Pampa should be considered with care as the number of observations was low (Fig. 8.2).

All biomes and soil classes were considered in the map unit definition as climate and soil class strongly influence plant K availability. The soils from Caatinga soils have a high fertility and low annual rainfall (approx. 250 mm). Soils of the Amazonia biome have the lowest amount of available K because of high annual rainfall and intense leaching. In the Cerrado biome, the levels of plant available K ranged from low and medium but soil use of the extensive pasture systems may promote K depletion. In the Mata Atlantica and Caatinga there many smallholder farmers.

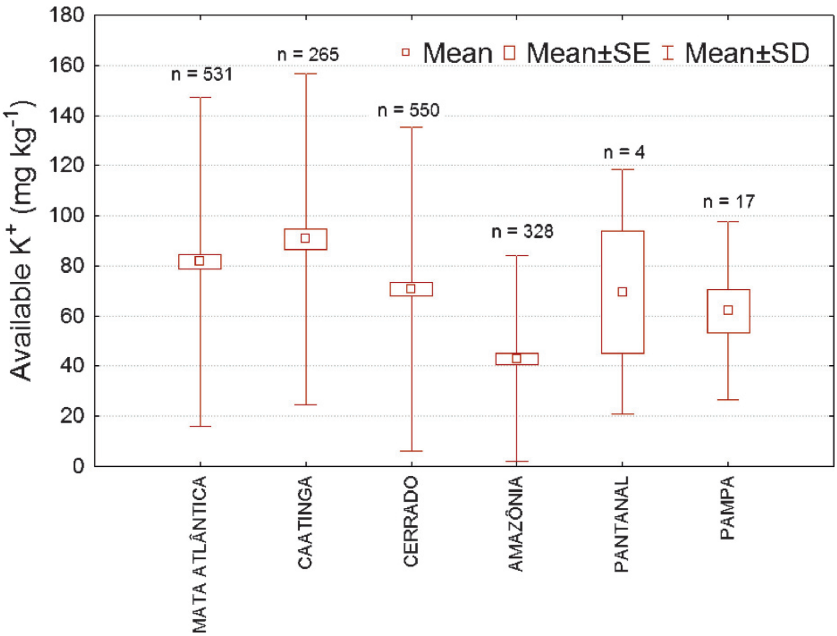


Fig. 8.2 The box and whisker plots of available K in Brazilian biomes. SD – Standard deviation, SE – Standard Error

8.3.1 Descriptive Statistics of K Data and Correlation with Other Soil Properties

Plant available K ranged from 3.90 to 456.3 mg kg^{-1} (mean = 79.0 mg kg^{-1}; SD = 79.6 mg kg^{-1}) showing a larger variability than the other soil properties such as Ca, Mg, Na, Al, and P (Table 8.1). This is probably due to larger and more frequent addition of mineral K fertilizers. Medium to low values of plant available K were also found by Silva et al. (2000) in a study on K forms in Ferralsols from the Cerrado biome. The authors reported that low K values are typical of highly weathered Ferralsols.

Sand content is negatively correlated with K, whereas silt values are positively correlated with K (Table 8.2). Sandy soils generally contain low K due to low nutrient retention capacity and high leaching. The silt fraction contains K-rich minerals, especially 2:1 minerals (Table 8.1).

As expected for acid tropical soils, plant available K is well-correlated with soil pH (both in water and KCl solution). Also, acid soils ($pH < 5.0$) rarely have high K levels (Fig. 8.3).

Calcium and magnesium levels were related to K levels (Fig. 8.4A), whereas soil Al is negatively correlated with K (Fig. 8.4A). This is due to soil liming (e.g. $Ca.MgCO_3$) which combined with K-fertilization increases the association. There was little relation between K and organic C levels (Fig. 8.4B).

Table 8.1 Descriptive statistics for soil properties at the surface depths ($n = 1976$)

Soil properties	Unit	Valid N	Mean	Minimum	Maximum	Std. Dev.	CV
sand		2496	463	0	980	264	57%
silt	$g\ kg^{-1}$	2530	195	0	650	137	70%
clay		2530	341	10	950	215	63%
silt/clay		2530	0.8	0.0	6.6	0.8	103%
water pH		2530	5.1	3.3	7.7	0.8	16%
KCl pH		2494	4.3	2.9	7.4	0.7	16%
Δ pH		2494	−0.8	−2.8	0.8	0.3	46%
Ca+Mg	$cmol_c\ kg^{-1}$	2530	3.5	0.0	24.6	4.6	131%
available K	$mg\ kg^{-1}$	2530	79.0	3.9	456.3	79.6	101%
Na	$cmol_c\ kg^{-1}$	2523	0.1	0.0	8.1	0.3	411%
Sum of bases	$cmol_c\ kg^{-1}$	2530	3.8	0.0	25.3	4.8	126%
Al	$cmol_c\ kg^{-1}$	2511	1.1	0.0	6.6	1.3	121%
H+Al	$cmol_c\ kg^{-1}$	2485	5.7	0.1	35.0	4.0	72%
P	$mg\ kg^{-1}$	1976	4.7	0.0	329.0	15.7	330%
Org.Carbon	$g\ kg^{-1}$	2530	14.6	0.4	69.1	9.5	65%
N	$g\ kg^{-1}$	2358	1.4	0.0	6.9	0.8	60%

Table 8.2 Correlation between K and other soil properties ($n = 1976$)

	sand	silt	clay	Ca+Mg	pH	OC	Na+	N	Al	P
available K	−0.30	0.37	0.14	0.63	0.47	0.13	0.05	0.29	−0.24	−0.07

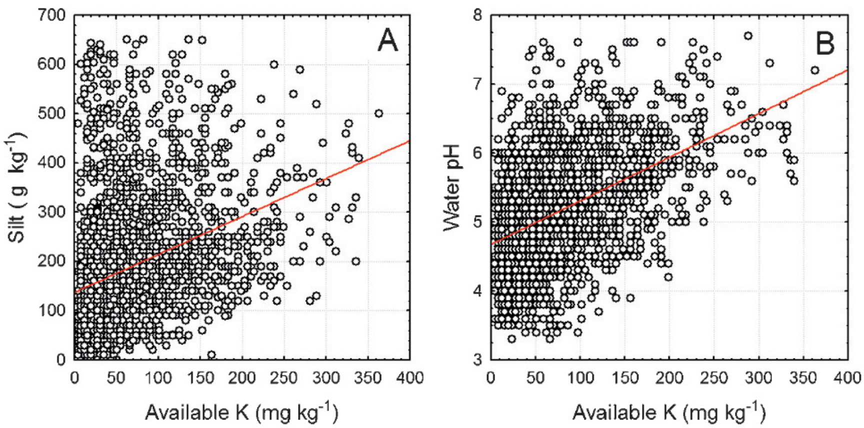


Fig. 8.3 Silt fraction (**A**) and soil water pH (**B**) in relation to available K

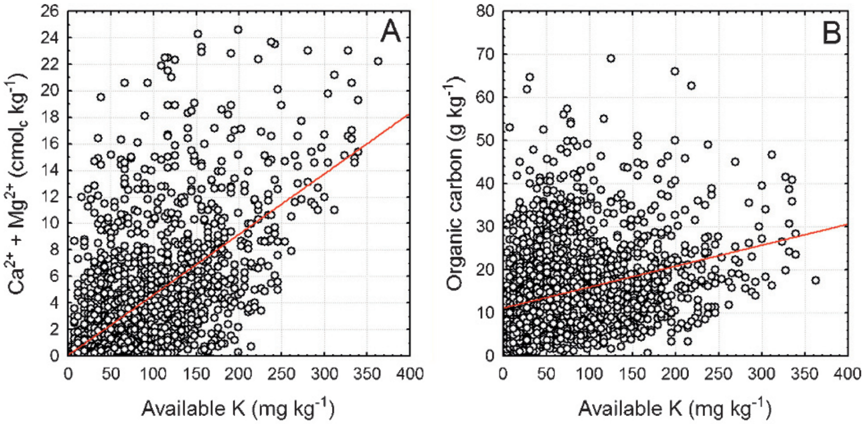


Fig. 8.4 Soil Ca+Mg (**A**) and organic carbon (**B**) in relation to available K

A multiple regression model for K estimation was performed based on Ca+Mg, water pH, organic carbon, and silt. The model explains about half of the K variation with a standard error of the estimate of 43.7 mg kg^{-1} (Fig. 8.5).

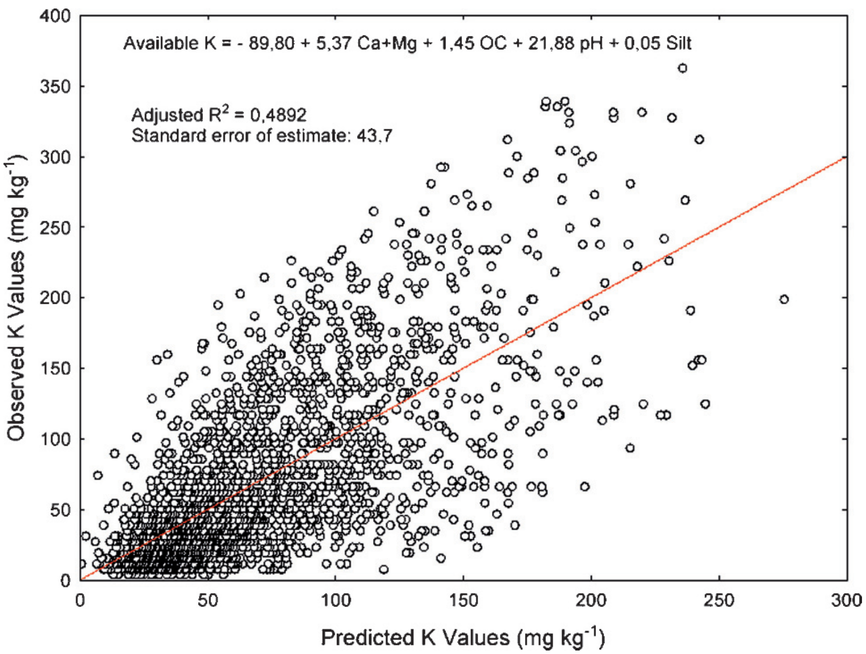


Fig. 8.5 Relationship between observed K and predicted K values using the model based on calcium, magnesium, organic carbon, water pH, and silt of Brazilian soils (n = 7072)

8.3.2 K Availability in Brazilian Biomes and Soils

Acrisols, Luvisols and Ferralsols are predominant in most biomes except for Pampa and Pantanal, in which the Planosols cover most of the areas (Table 8.3).

The results from mapping K availability from calculated K soil profiles with spatial information are shown in the Fig. 8.6. The 482 calculated soil K (mg kg^{-1}) were used to map 1992 soil-biome units. The legend presents the relative proportion of mapped area for each class.

The results from mapping the K availability in the third-order of Acrisols, Luvisols and Ferrasols are shown in Fig. 8.7. The legend presents the percentage area for each class. The most part of biome-soil units showed low amounts of K availability and the small biome-soil units were high. There was no biome-soil units associated with a very high K-levels. It seems that this results reflects better the

Table 8.3 Distribution of Brazilian soil classes in different biomes

Biome	Soil class	Percentage (%)
Amazônia	Acrisols and Luvisols	30.9
	Ferralsols	30.3
	Gleysols	8.0
	Others	30.5
Caatinga	Lithosols and Arenosols	28.8
	Ferralsols	21.0
	Acrisols and Luvisols	15.4
	Others	34.7
Cerrado	Ferralsols	40.7
	Lithosols and Arenosols	23.1
	Acrisols and Luvisols	12.0
	Others	24.1
Mata Atlântica	Ferralsols	35.5
	Acrisols and Luvisols	28.8
	Cambisols	15.5
	Others	20.0
Pampa	Planosols	26.0
	Lithosols and Arenosols	23.4
	Acrisols and Luvisols	22.3
	Others	28.2
Pantanal	Planosols	31.8
	Podzols	19.9
	Plinthic soils	18.7
	Others	29.4

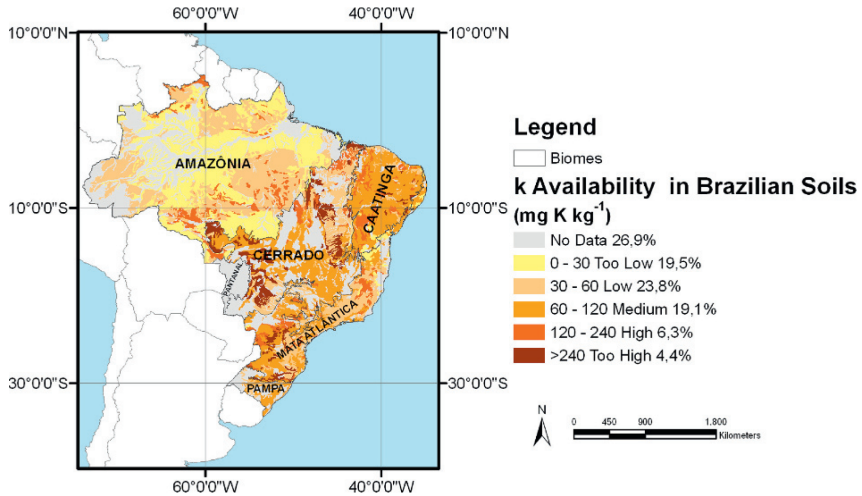


Fig. 8.6 K availability map from calculated K soil profiles with associated spatial information (See also Plate 8 in the Colour Plate Section)

reality because it was considered the soil classification in third-order. However, it is necessary to classify the other soil classes to third-order to map K under this method for all Brazil. The profiles ($n = 296$) used in this case were Xanthic Ferralsols in Amazonia, and Rhodic Ferralsols in the Cerrado and the Pampa and Pantanal biomes did not have K soil profile data.

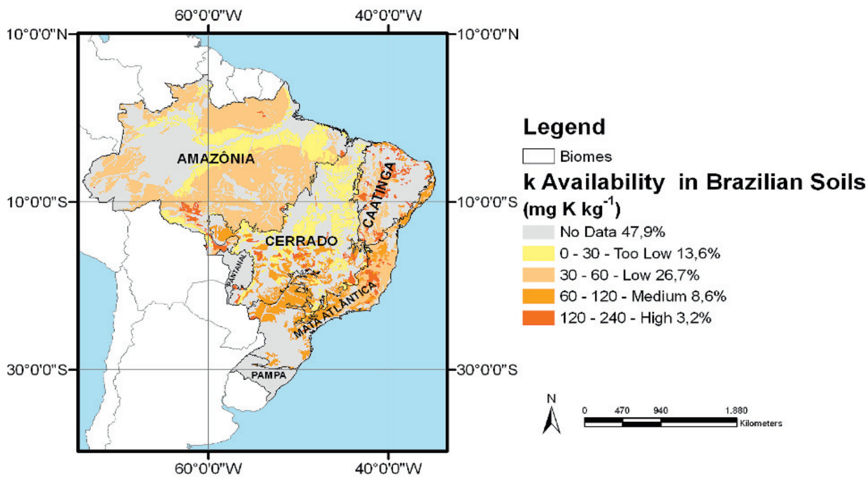


Fig. 8.7 K availability from descriptive statistics application to calculated K soil profiles in a third-order classification to Acrisols, Luvisols e Ferralsols (See also Plate 9 in the Colour Plate Section)

8.4 Conclusions

The biome and soil intersection used to obtain the mapping units of the soil properties may be more useful than municipality borders, because they have natural limits. Our results showed that the mapping of soil K levels based on soil survey reports, at a regional scale, is difficult because of limitations in georeferencing and spatial distribution of soil profiles. Than, it is necessary to test other methods to improve accuracy of K mapping using limited soil profile data in Brazil. For this, it is important consider other themes in the mapping like land use and terrain elevation. This kind of mapping would help fertilizer distribution planning but it is not suitable for fertilizer recommendation. Soil map scaling is also discussed in Chapter 17. For mapping soil K availability in a better scale, like 1:250,000, it is recommended to use georeferenced and more representative soil data set.

Acknowledgments This work was financially supported by the International Potash Institute (IPI). The Soil Database and substructure used was supplied by Embrapa Soils, Rio de Janeiro, Brazil.

References

Bernardi, A.C.C., Machado, P.L.O.A., Silva, C.A. 2002. Fertilidade do solo e demanda por nutriente no Brasil. In: Manzatto, C.V., Freitas Junior, E., Peres, J.R.R. (Eds.), Uso agrícola dos solos brasileiros. Rio de Janeiro: Embrapa Solos 61–77.

Environmental Systems Research Institute (ESRI). 2006. ArcGIS v.9.1. Redlands, CA.

Instituto Brasileiro de Geografia e Estatística – IBGE e Ministério do Meio Ambiente – MMA. 2005. Mapa de Biomas do Brasil. Escala 1:5.000.000. Rio de Janeiro: IBGE.

Instituto Brasileiro de Geografia e Estatística – IBGE, 2001a. Mapa de Solos do Brasil. Escala 1:5.000.000. Rio de Janeiro: IBGE.

Instituto Brasileiro de Geografia e Estatística – IBGE/DGC/DECAR. 2001b. Malha Municipal Digital do Brasil. Escala 1:2.500.000. Rio de Janeiro: IBGE. CD-ROM.

Lopes, A.S. 2005. Reservas de minerais potássicos e produção de fertilizantes potássicos no Brasil. In: Yamada, T., Roberts, T.L. (Eds.), Potássio na agricultura brasileira. Potafos: Piracicaba, Brasil 21–32.

Mascarenhas. H.A.A., Tanaka. R.T., Wutke. E.B., Braga. N.R., Miranda. M.A.C. 2004. Potássio para a soja. Informações Agronômicas. N° 105. Potafos: Piracicaba 1–2.

Poulisse, J. 2003. Issues of sustainable agriculture in developing countries. In: Johnston, A.E. (Ed.), Feed the soil to feed the people: The role of potash in sustainable agriculture. Invited papers. Basel. International Potash Institute 49–69.

Van Raij, B., da Silva, N.M., Bataglia, O.C., Quaggio, J.A., Hiroce, R., Cantarella, H., Bellinazzi Jr., R., Dechen, A.R., Trani, P.E. 1985. Recomendações de adubação e calagem para o Estado de São Paulo. Boletim Técnico. Campinas, Instituto Agronômico 107p.

Sanchez. P.A. 1997. Changing tropical soil fertility paradigms: from Brazil to Africa and back. In: Moniz, A.C., Furlani, A.M.C., Schaeffert, R.E., Fageria, N.K., Rosolem, C.A., Cantarella, H. (Eds.), Plant-soil interactions at low pH. Viçosa. MG. Brazilian Soil Sci. Soc. 19–28.

Silva, I.R., Furtini Neto, A.E., Fernandes, L.A., Curi, N., Vale, F.R. 2000. Formas, relação quantidade/intensidade e biodisponibilidade de potassio em diferentes Latossolos. Pesq. Agropec. Bras. 35, 2065–2073.

Chapter 9
GIS as a Support to Soil Mapping in Southern Brazil

E. Weber, H. Hasenack, C.A. Flores, R.O. Pötter and P.J. Fasolo

Abstract Traditional soil surveys follow a specific methodology to identify, characterize, and fit mapping units in a classification system and to spatialize them in order to produce soil maps. The need for observation and characterization on field, associated with the physical and chemical analyses, makes the surveys expensive and therefore scarce. The low number of surveys stimulated the development of models for digital soil mapping, whose results proved to be possible to predict and spatialize many soil characteristics. However, conventional soil surveys remain important as a basis for the development of digital soil mapping models, setting a reason to continue the development of methodologies to improve the conventional surveys. Technologies like GPS and GIS contribute to make field observation and soil sampling more objective and make the mapping process and the production of hardmaps easier and faster. The objective of this study was to develop methodologies to integrate cartographic base elements with field work, using GIS and GPS in an area corresponding to 20 topographic charts in scale 1:50,000 in the State of Rio Grande do Sul, Southern Brazil, to obtain soil mapping based on the Brazilian Soil Classification System. The result obtained was a georeferenced digitized soil map, continuous for the whole region, free of inconsistency among neighbor map sheets and with attributes associated with the mapping units. These characteristics allow the use and application of the soil map for many purposes like zoning, diagnosis, suitability analysis as well as serving as a basis to the development of models for digital soil mapping.

9.1 Introduction

According to Hudson (1992), soil survey is a scientific strategy based on the concepts of factors of soil formation coupled with soil–landscape relationships. Soil surveys follow a specific methodology that aims to identify, characterize, and fit

E. Weber
Universidade Federal do Rio Grande do Sul, Centro de Ecologia, Porto Alegre, RS, Brasil
e-mail: eweber@portoweb.com.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

mapping units in a classification system. Moreover, such methodology delimits these mapping units in order to obtain outputs such as charts or maps. Its procedure consists in studying the area and the main features of the soil categories, including morphological description, physical and chemical analyses and taxonomic classification, besides the delimitation of cartographic units.

The needs for characterization of soil categories in the field (auger sampling and observation of trenches and road embankments) and the laboratory analyses make the surveys expensive and slow. Time and costs involved are probably one of the reasons for the scarcity of more detailed and updated soil maps. With some exceptions, in larger countries the availability of soil data is poor (See Section 5.1). Brazil, for example, is uniformly covered by soil maps in scale not larger than 1:1,000,000, like the Soil Map of Brazil and the Agricultural Suitability Map of Brazil, exploratory soil maps by the RADAM/EMBRAPA Solos project and Agroecological Zoning (McBratney et al., 2003).

In opposition to the lack in soil surveys, there has been an increasing demand for territiorial information in the last few years. The arrival and spread of technologies related to collection and analysis of spatial data, such as the Geographic Information Systems (GIS), have made easier the integration of information from different origins and stimulated the development of countless studies and projects based on spatial data. In several cases, the soil maps are extremely important to obtain the expected results, as for the studies of watershed management, environmental assessment, land zoning and planning, among others.

The scarcity of spatial data about soil has stimulated the development of digital modeling techniques in order to predict and to spatialize soil classes and properties. The obtained results are different from those of the conventional survey, since they try to estimate soil parameters for specific purposes when pedological maps are not available or appropriate (Morris et al., 2000; Zhu et al., 2001; Brodsky et al., 2006). Although many advances have been obtained through soil digital modeling, conventional surveys are and will continue to be important because they gather very detailed information about mapping units, systhesizing the knowledge acquired by soil scientists throughout decades. Such analogical information has been partially transferred to digitized maps and its attributes stored in databases. This is essentially a conventional process which intends to make use of the benefits of GIS in structuring the information, additionally supporting the digital modeling itself (Carré and McBratney, 2006).

One of the recent challenges in soil mapping consists of improving the conventional surveys in order to make them less time consuming, less expensive and more rapidly available (See also Section 3.1 and 24.6). One of the typical characteritistics of the traditional surveys is the existence of a considerable delay between the field work for landscape observation and sample collection, the delimitation of mapping units and the publication of the results. There are many reasons: the time needed for the physical and chemical sample analyses, the availability of technicians to interpret the results and delimit the units and the effort necessary to write and to prepare the final reports and maps for publishing. If there is a lack of time between the conclusion of the survey and the publishing of the results, authors will have little opportunity to review and correct any inconsitency.

The search for improvements in different stages of conventional soil survey must necessarily take into account the use of new technologies, such as GPS (Global Positioning System), PDA (Personal Digital Assistants) and GIS to support the different activities involved. The use of such resources may contribute to a faster and more objective field data collection and to make the delimitation of mapping units more precise, making the availability of the final product easier and potentializing the use and the application of the results (Aronoff, 1991; Morris et al., 2000; Hempel et al., 2006).

The objective of the study here described was to integrate elements of cartographic bases with traditional soil surveys in the Serra Gauúcha region, State of Rio Grande do Sul, Brazil, aiming at the development of methodology supported by GIS and GPS to obtain continuous digitized soil maps. It should not be confused with Digital Soil Mapping (DSM), as defined by McBratney et al. (2003), since methodologies are based on knowledge of soil surveyors about soil-landscape relationships and does not use predictive models to estimate soil properties or to define soil classes. The procedure adopted can be understood as a step before, an approach to improve traditional surveys in order to generate soil maps in less time an to produce better and more useable results.

9.2 Material and Methods

The State of Rio Grande do Sul is located between latitudes 27°00′S and 33°45′S and longitudes ranging from 57°40′S to 49°35′W, and has borders with Argentina and Uruguay (See Figure 9.1). The study area is defined by a rectangle with a surface of 13,490 km^2, situated at the Serra Gaúcha region, NE portion of the State, between latitudes 28°30′S and 29°30′S and longitudes 50°45′W and 52°W (See Fig. 9.1). The area covers 86 municipalities, 44 of them totally covered and 42 partially covered (20 with more than 50% of their territory and 22 with a lower proportion).

The material used consisted of navigation GPS receivers, Cartalnix (Clarklabs©) vector editing software, Idrisi (Clarklabs©) GIS software, Corel Draw (Corel Corporation©) desktop publishing software, and a cartographic base of the study area.

The cartographic database for the rectangle corresponding to the study area is composed by a set of 20 topographical map sheets from the Brazilian Systematic Mapping in scale 1:50,000 in UTM Projection, Zone 22. They were generated by the Brazilian Army (*Diretoria de Serviço Geográfico – DSG*) and are in the largest scale available as continuous mapping of the whole region. Despite their small scale, they represent a material of great relevance, since there is not any other better base available in larger scale.

The mapping process was conducted in a way to integrate the information about a semi-detailed soil survey and elements of the cartographic base for the generation of a GIS-structured digitized soil map. The soil survey was based on the methodology established by EMBRAPA (2006) for semi-detailed soil survey according to the

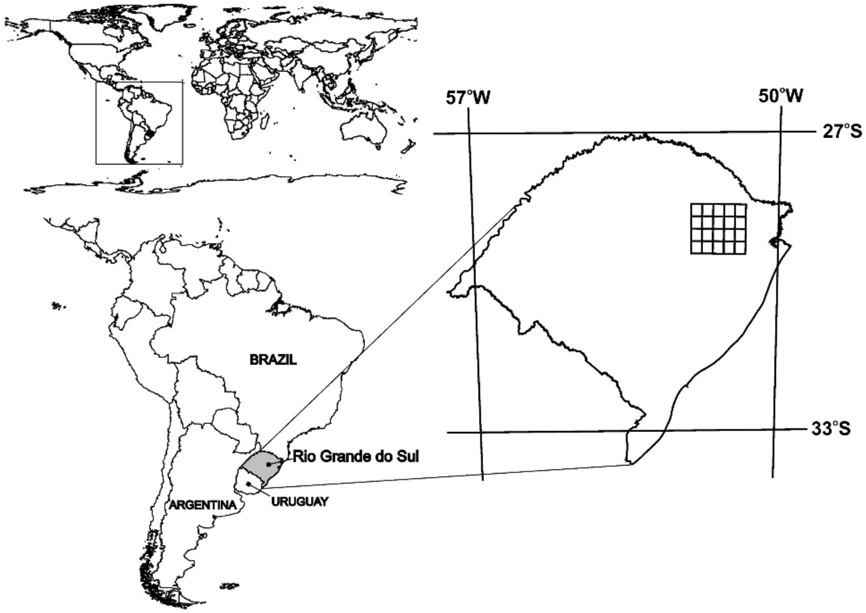


Fig. 9.1 Location of the State of Rio Grande do Sul and the 20 map sheets in scale 1:50,000 of the Wine Zoning Project in Rio Grande do Sul

Brazilian Soil Classification System. Field work was carried out in order to make a description of the soil profiles and soil samples were collected for physical and chemical analyses with the support of GPS receivers and 1:50,000 scale topographic map sheets, laminated with a plastic film.

The use of GPS aimed to facilitate the association of the visited places with their correspondent sites in the topographic sheets. The use of laminated map sheets made the delimitation of the mapping units easier, giving the spatial basis for drawing the limits and providing support to make spatial adjustments and corrections. Moreover, laminated map sheets are more resistant and prevent the loss of information due to humidity, rain, tears, and other damages that may easily take place during the field work.

During the mapping stage, the main physiographic units of the study area were firstly delimited on the topographic charts. Then these physiographic units were traversed in the field from the lower parts to the highest point of the terrain, in order to visualize the sequence of soil distribution in the landscape and to establish a preliminary legend for soil types.

After that, the necessary routes for data collection were established based on detailed examination of the 1:50,000 topographic sheets and smaller scale maps of geology and soil types of the region. The soil survey along these routes was done through auger sampling and observation of trenches and road embankments.

During the field work, all the routes performed, the soil profiles described and additional points of interest were registered with GPS. Also at this stage, the necessary adjustments and corrections to the preliminary legend of soil types were done in order to obtain a correct classification of the soil types found.

The distribution of the identified soil types, the knowledge about the relation soil-landscape acquired during the establishment of the preliminary legend and improved during the survey, the use of GPS and the equidistance of the contour lines in the topographic sheets made possible the identification of the points for observation and sample collection, as well as places of soil taxonomic class changing. For each taxonomic unit, a complete profile was described (Klamt et al., 2000) and, in some cases, a complementary one was done, based on Lemos and Santos (1996). This information was used to draw the limits of the mapping units on the topographic map sheets in scale 1:50,000.

The mapping process took into account the set of features that were potentially important in soil use. Among them, vegetation, relief, and the presence of gravels or rock outcrops were used to subdivide the units and used as indicators of water conditions, the susceptibility to erosion, and the possibility of mechanization. Some other elements used in separating the units were clay's activity, saturation with bases, saturation with aluminum, the type of horizon A, texture and, for the less developed soils, the rock substratum. In some cases it was not possible to individualize soil types, either for the fact that some classes did not present geographic extension enough or because their intricate occurrence did not allow individual delimitation in the desired scale. In such cases, the mapping was done as soil associations.

Parallel to the field work for soil survey, a cartographic base was structured in a GIS through the digitizing of map sheets in scale 1:50,000. Paper maps were scanned using a large-sized scanner and were georeferenced based on an UTM grid. The main information, like contour lines, hydrographic network, road system, and urban areas were then digitized on-screen. The respective layers were topologically structured and the objects were associated with a set of attributes in linked tables. Lastly, a Digital Elevation Model (DEM) of the whole region was generated through interpolation based on the digitized contour lines (See Fig. 9.2).

After finishing the field work and the GIS cartographic base, the soil mapping units resulting of the conventional survey were digitized and edited. Following the same steps used for the cartographic base, the laminated map sheets with the soil

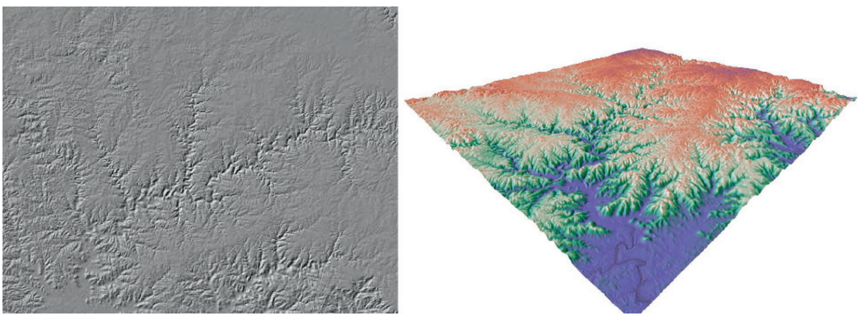


Fig. 9.2 DEM Hill shading of the 20 map sheets and 3D view of the region (SW to NE)

mapping units drawn on the field were scanned and georeferenced based on an UTM grid. Then the limits of the mapping units were digitized manually on-screen, using the georeferenced field map sheets as a basis, producing vector line features (See Fig. 9.3).

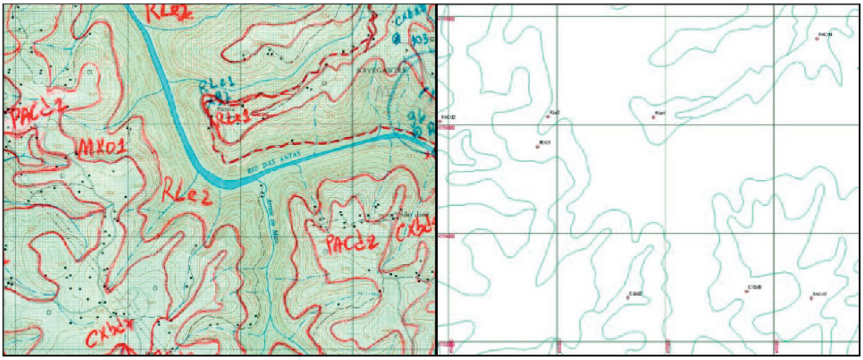


Fig. 9.3 The topographic map sheet resulting from the field work and the mapping units after digitized (See also Plate 10 in the Colour Plate Section)

Vector lines extraction was performed in a continuous way, alternating the backdrops but capturing the limits of the mapping units in a unique layer without the map sheet's divisions. After the topologic structuring of the limits of mapping units, soil polygons were built and associated with a set of attributes, such as their area (in hectares), order, suborder, group and subgroup, acronym of the mapping unit, among other features. The objective of this procedure was to guarantee the consistence of the attributes and the spatial contiguity of the polygons among contiguous sheets, in order to generate a vector polygon file of soil types with the continuous coverage of the 20 map sheets.

The last step involved the preparation of the material for printing. The area of each of the 20 map sheets was clipped and used to generate a printing layout with the desired layers, including soil layer, complementary layers (hydrographic network, road system, and urbanized areas) and ancillary information (legend, map grid, text layer, etc.). In order to keep a regional context, the complete legend of the whole region was used in each sheet legend, turning grey the ones absent at the respective map sheet. In this printing layout the soil mapping units were colored using the colors defined by EMBRAPA (2006). Finally, in order to include relief information with an easier perception and comprehension than with simple contour lines, the colored soil polygons were combined with a DEM analytical hill shading.

9.3 Results and Discussion

The semi-detailed soil mapping of the Serra Gaúcha region covered an area of $13,490\,\mathrm{km}^2$, and 61 soil mapping units were identified and delimited. These mapping units were represented by 1,626 polygons divided in 7 soil groups: 24 Cambissolos (Inceptisols); 1 Gleissolos (Entisols); 1 Latossolos (Oxisols); 7 Chernossolos (Molisols); 15 Nitossolos (Alfisols); 8 Argissolos (Ultisols), and 5 Neossolos (Entisols).

The use of topographic maps and technologies like GPS and GIS in this study facilitated soil mapping in many aspects: support to field work activities, georeferencing of sampling information, drawing of mapping units, soil data check-up and correction, maintenance of spatial consistency and storage of soil attributes, agility and uniformity at the generation of the printing layout.

The integration of field information with the data from the cartographic base in a GIS environment allowed the production of a continuous and georeferenced digitized soil map for the whole region, covering a surface that corresponds to 20 map sheets in scale 1:50,000 (See Fig. 9.4). The output is a vector polygon file that guarantees the consistence of the attributes and the spatial contiguity of the polygons among contiguous map sheets. Moreover, this file is associated with tables that contain the main attributes about the mapping units.

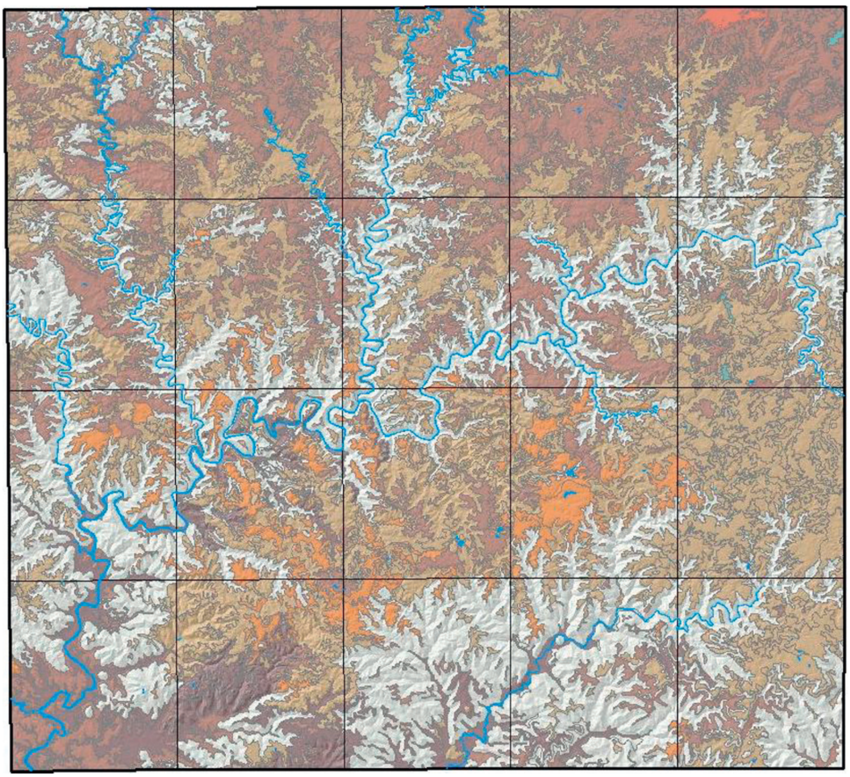


Fig. 9.4 Continuous georeferenced digitized soil map of the Serra Gaúcha region (See also Plate 11 in the Colour Plate Section)

Concerning the product for printing, the fusion of the soil polygon information with the hill shading derived from a DEM has a pleasant effect, showing the relief forms through the shading effect. The perception of the position of a given soil mapping unit on the relief is direct, making the interpretation and use of the map easier (See Fig. 9.5). The usual way of introducing topography into the maps is through contour lines, but in steep relief, however, contour lines may excessively congest information.

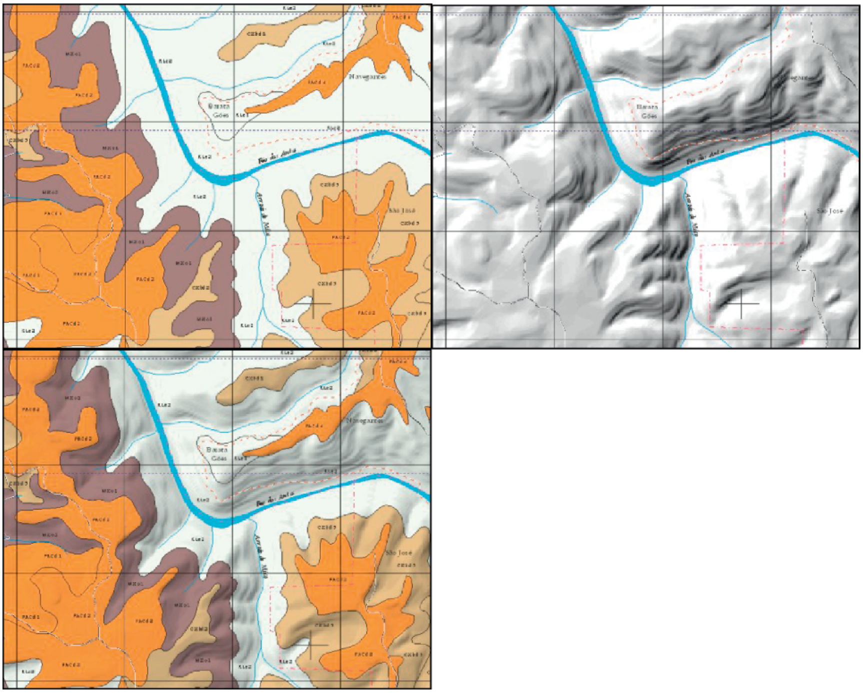


Fig. 9.5 Soil map with conventional soil information, analytical DEM hill shading and fusion of the hill shading with the conventional soil map (See also Plate 12 in the Colour Plate Section)

Printing of the final material was done in a fast and uniform way and presented excellent visualization. Moreover, it was done at the same time that the mapping process was concluded, which facilitated check-ups and corrections. The fusion of the soil map with the hill shading derived from a DEM resulted in a graphic material that highlights relief forms through the hill shading. The perception of the position of the mapping units in the landscape is direct, making the interpretation and use of the maps easier.

The final digitized soil map structured in GIS allows one to query physical and chemical characteristics of soils in a given place or to select places where soils have a set of desired characteristics. It also allows to quantify the surfaces and to cross soil information with other georeferenced data of the region. Therefore, the soil map presents a great potential of application for many purposes, like zoning, diagnosis, suitability evaluation, among others. For instance, the product can be used as reference or as ancillary information for future studies on digital soil mapping, giving support to assess soil parameters and to validate modeling results.

Soil can be predicted through several approaches, for areas previously mapped, or for new areas. Among others, prediction can be done from soil attributes at the same location or at neighbouring locations from itself, from other soil attributes and from environmental attributes. In other words, real soil observations with a good density will ever be essential to develop, to evaluate or to fit the models (See also Chapter 17). Although conventional soil maps are essentially static products, when structured in GIS they can provide useful information for analysis or interpretation aiming to soil prediction. This is true for soil polygons as well as for the field sample collection points, which data have higher confidence. So, we should not stop with traditional surveys, but improve methodologies attempting to optimize data collection procedures and to make results more consistent and easily available.

9.4 Conclusions

Besides the introduction of techniques to support and to improve conventional soil surveys, the methodology used in the semi-detailed soil mapping in the Serra Gaúcha region proved to be a useful way to organize survey efforts. The adoption of an existing systematic articulation made easier the organization and application of the mapping activities and the association of the soil map with ancillary information to produce the final map. Such procedure may facilitate the planning for mapping expansion, which favors the optimization of future efforts.

The generation of a continuous digitized polygon map permitted to avoid many problems that usually occur with adjacent map sheets. The cutting of the area for each sheet of the continuous map is easy, safe, and guarantees a perfect continuity among the elements of contiguous charts. The same does not happen when several contiguous sheets are produced individually.

The simultaneous execution of the several stages in the mapping process proved to be an adequate strategy that allowed the conclusion of the mapping almost at the same time the field work ended. Thus, it was possible by the soil surveyors to do all the check-ups and adjustments directly on the material under elaboration.

Regarding to the use of GIS, as referred by Burrough and McDonnell (1998), GIS is a tool for collating all kinds of spatial information, but in itself is incapable of soil mapping. Intellectual framework and accurate soil and environmental data are needed. Looking at a stage of digital soil mapping, these can be some of the major challenges to be faced in larger countries, specially developing countries. Constructing a database of existing soil profiles and sample points from traditional surveys could be a good strategy to facilitate access to soil data by soil modellers.

References

Aronoff S., 1991. *Geographic information systems: a management perspective*. WDL publications. Otawa. 294p.

Brodsky, L., Boruvka, L., Hajek, F., 2006. Obj*ect-oriented fuzzy analysis of remote sensing data for bare soil mapping*. In: Global Workshop on Digital Soil Mapping, 2. Proceedings. Rio de Janeiro, CD-ROM. 6 p.

Burrough, P.A., McDonnell, R.A., 1998. *Principles of Geographical Information Systems*. Oxford University Press, Oxford.

Carré, F., McBratney, A. B., 2006. Minasny, B. *Estimating and improving the quality of legacy soil samples for digital soil mapping*. In: Global Workshop on Digital Soil Mapping, 2. Proceedings. Rio de Janeiro, CD-ROM. 6 p.

EMBRAPA, 2006. Solos. *Sistema Brasileiro de Classificação de Solos*. 2.ed. Rio de Janeiro, 306p.

Hempel, J.W., Hammer, R.D., Moore, A.C., Bell, J.C., Thompson, J.A., Golden, M.L., 2006. *Challenges to digital soil mapping*. In: Global Workshop on Digital Soil Mapping, 2. Proceedings. Rio de Janeiro, CD-ROM. 10 p.

Hudson, B.D., 1992. *The soil survey as paradigm-based science*. Soil Science Society of America Journal 56, 836–841.

Klamt, E., Dalmolin, R.S.D., Gonçalves, C.N., Inda Junior, A.V., Almeida, J., Flores, A.C., 2000. *Proposta de normas e critérios para execução de levantamentos semidetalhados de solos e para avaliação da aptidão agrícola das terras*. Pelotas: NRS-SBCS, 44p. (NRS-SBCS. Boletim Técnico, 5).

Lemos, R.C., Santos, R.D. dos., 1996. *Manual de descrição e coleta de solo no campo*. 3.ed. Campinas: Sociedade Brasileira de Ciência do Solo, 83 p.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. *On Digital Soil Mapping*. Geoderma, 117, p. 3–52.

Morris, D.K., Stienhardt, G.C., Nielsen, R.L., Hostetter, W., Haley, S., Struben, G.R., 2000. *Using GPS, GIS, and Remote Sensing as a Soil Mapping Tool*. In Proc. 5th International Conference on Precision Agriculture, Bloomington, MN. 16–19 July, Center for Precision Agriculture, University of Minnesota, St. Paul, MN.

Zhu, A. X., Hudson, B., Burt, J., Lubich, K., Simonson, D. 2001. S*oil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic*. Soil Science Society of America Journal, Vol. 65, September–October.

Chapter 10
Experiences with Applied DSM: Protocol, Availability, Quality and Capacity Building

R.A. MacMillan

Abstract This chapter considers both opportunities and constraints to applied, operational digital soil mapping (DSM) from the points of view of a) availability of suitable input data layers, b) protocols available for DSM, c) quality of input data layers and resultant output maps and d) other efforts required to build predictive mapping capacity and apply it effectively.

Many potential DSM practitioners are discouraged by the real or perceived lack of availability of suitable input data layers to support DSM, particularly in regions with weakly developed spatial data infrastructures. Solutions to addressing problems of limited or sparse spatial data sets are identified for input layers derived from digital elevation models (DEM's), remotely sensed imagery and available secondary source maps.

A variety of protocols for producing predictive soil maps are discussed under the general headings of unsupervised, supervised and knowledge-based (or heuristic) approaches. These key protocol activities support the ability to make maps of the spatial distribution of soil classes or attributes by developing predictive relations between spatially distributed input variables or classes and the desired output classes. Different strategies are reviewed to acquire and formalize tacit knowledge embodied in soil-landform conceptual models and to capture this tacit knowledge as quantitative rules.

A considerable amount of resistance to DSM arises from real or perceived concerns about the quality of the resulting maps in comparison to existing maps produced using traditional mapping methods. Quality, defined as the ability of a map or product to correctly predict the characteristics of the landscape at particular points or within particular small areas, is discussed as are suitable approaches for evaluating and reporting it.

The capacity to apply DSM routinely and operationally requires additional support in the form of training, access to suitable tools and software and access to suitable input data. Approaches to developing support for DSM from decision makers and

R.A. MacMillan
LandMapper Environmental Solutions Inc., Edmonton, Alberta, Canada, T6G 1V4
e-mail: bobmacm@telusplanet.net

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

funding agencies in the face of institutional and discipline resistance to embracing new technologies are identified, specifically incremental projects with clearly defined goals and testable measures of success. Finally, it is noted that perhaps the biggest hurdle to building capacity is our own hesitancy to believe in ourselves and to dream big and try big. It is hoped that this chapter will encourage individuals with an interest in applying new predictive mapping techniques to embrace change and to try to create useful, operational maps for large areas in their own regions of interest.

10.1 Introduction

Digital soil mapping (DSM) (McBratney et al., 2003) has been an area of active research for more than a decade (see Zhu, 1994; Skidmore et al., 1991) but it has yet to achieve widespread adoption for operational mapping by main-stream mapping agencies (see similar comments in Sections 1.1, 2.1, 3.1 and 19.2). Most reported applications have involved investigating and evaluating the capabilities of DSM as a potential tool for replacing or enhancing soil maps prepared using conventional manual methods. The SOLIM approach of Zhu et al. (2001) may well represent the predictive mapping protocol that has advanced furthest along the path towards routine operational use. SOLIM has been used to complete several large scale pilot projects to evaluate its potential for use in routine, operational soil survey (see http://solim.geography.wisc.edu/projects/index.htm). Bui (2000) and Bui and Moran (2003) used predictive techniques to remap a large area of soils in the Murray-Darling Basin at a grid resolution of 250 m but most of the projects reported on by this group had more of a focus on applying and evaluating new predictive techniques than on applying them for routine operational use in soil survey (Bui et al., 1999; Bui and Moran, 2001; Moran and Bui, 2002; Bui, 2004). Similarly, projects reported on by Thomas et al., (1999), Lagacherie et al., (1995), Cole and Boettinger, 2004) and Hengl and Rossiter (2003) were also more concerned with evaluating example applications of new methods than with applying them for operational use. What then are the constraints that limit adoption of new predictive DSM methods for routine operational use and what opportunities exist for overcoming these constraints?

10.2 Protocols for DSM

Almost all efforts to develop and apply DSM techniques can be seen to follow approximately the same basic steps (Fig. 10.1) (see also Section 1.1). Differences in protocols arise from differences in the kinds of outputs that are to be predicted, the kinds of input data layers selected to support predictions and the kinds of equations or rules developed to make predictions. These steps are inter-related such that decisions on what to predict (individual soil properties or soil classes) influence both the selection of input variables and the development of predictive equations and vice versa.

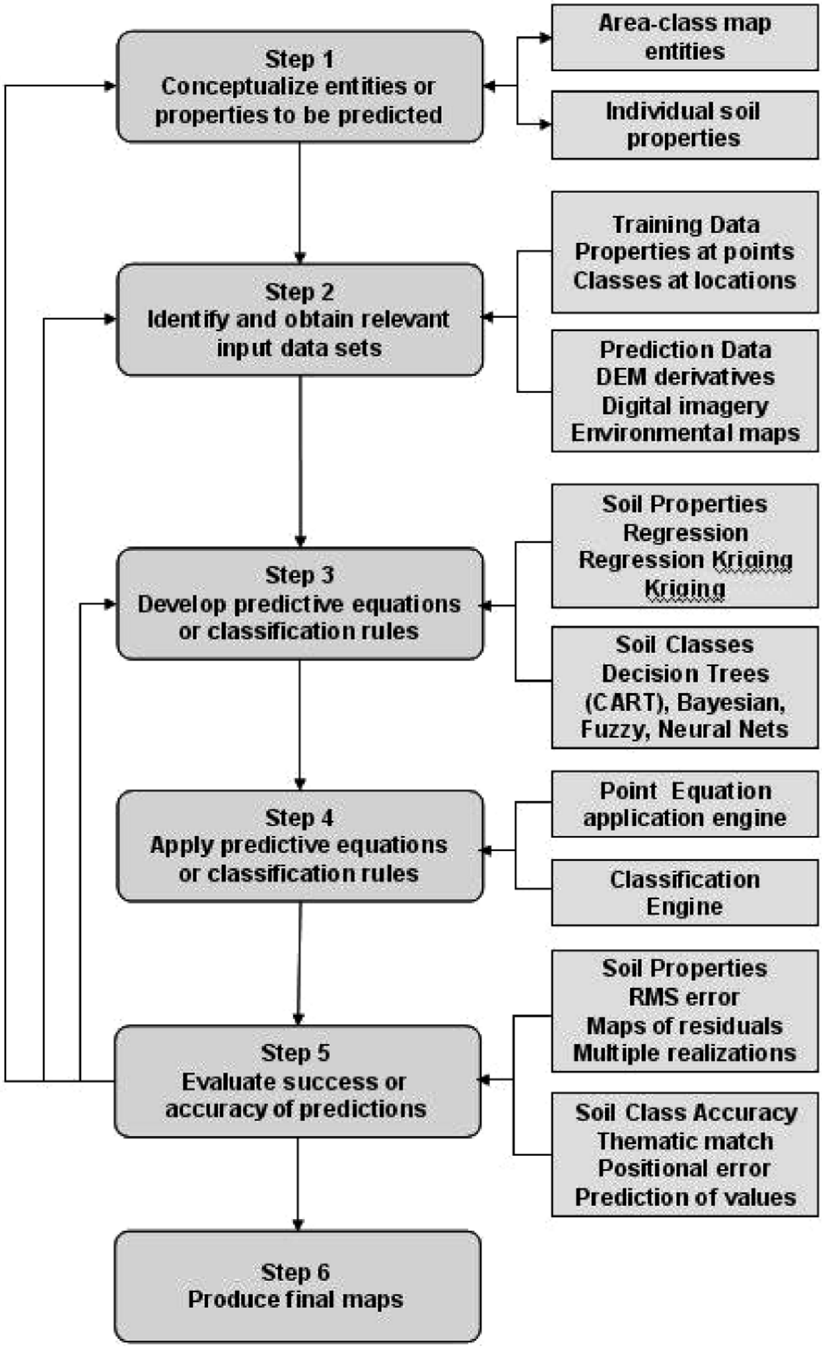


Fig. 10.1 Steps common to most DSM protocols

This chapter focuses mainly on protocols that aim to predict the spatial distribution of discrete classes of soil (area-class maps), rather than continuous soil variables. Protocols for predicting the continuous distribution of single soil properties (see Chapter 33) typically require a large number of spatially-registered point observations at which the variable of interest has been measured or sampled in order to develop locally-relevant predictive equations (also discussed in Sections 1.4 and 2.1). Such large volumes of point data are not common in regions of generally sparse data.

10.3 What to Map?

The size, scale and conceptualization of the entities that we try to map exercises a profound influence on the choices that must then be made about what data are required to support the predictions and what protocols or predictive methods will be most appropriate for making the predictions. Protocols for predicting the spatial distribution of individual continuous soil properties typically require different data inputs and predictive methods than ones focused on predicting spatial patterns of soil classes. If predicting soil classes, it is necessary to clearly identify whether the objective is to predict the exact spatial location of narrowly defined entities at the level of soil series or of more broadly defined collections of soils at the level of soil associations. If the target is individual soil series, then it is necessary to establish the shortest distances over which these soils typically change so that predictor data sets capable of capturing and characterizing variation over distances that are shorter than this range can be selected. It is also necessary to establish whether individual soil types exhibit a predictable variation in response to available input predictor layers. In particular, it is desirable to establish that the target soils vary systematically in response to topographic controls that can be approximated by one or more measures extracted from digital elevation data (DEM). If predicting more generalized spatial entities similar to soil associations, the most useful input data and methods are likely to differ from those used to predict individual soil classes.

10.4 Availability of Suitable Input Data for DSM

Input data required for DSM can be differentiated into field-obtained training data needed to establish how soils or soil properties vary across the landscape and predictor data sets required to support application of predictions of that variation across entire areas (Fig. 10.1). Both of these may be in short supply in areas of sparse spatial data but, of the two, geo-referenced point samples or detailed soil-landform field observations are likely to be the most limiting (see also Sections 1.3, 9.1 and 20.1). Carré et al., (2007) and Minasny and McBratney (2007) provide some guidelines for methods of selecting geo-located point observations to support prediction of individual soil properties. Odgers et al. (2008) have proposed a random catena method of obtaining soil class observations along a topographic profile from the top to the base of a hillslope in order to collect data suitable for establishing relations between soil classes and landform controls. Field samples and observations are expensive and difficult to collect and may be entirely lacking in many less-developed areas. In some cases, it may be necessary to infer the nature of differences in soil classes or soil properties based on theoretical or empirical understanding of soil forming processes only, with no recourse to local field observations to confirm or correct these assumptions.

Spatially extensive data sets that completely cover entire areas of interest are necessary to support predictions of the spatial pattern of distribution of soil classes or soil properties. Most protocols for predicting soil classes or properties have identified and used more or less similar sets of input data layers consisting of measures extracted from DEMs, imagery of varying spatial, spectral and, more recently, temporal resolution and what may be termed ancillary environmental maps depicting regional or local variation in controls such as geology or bedrock lithology, parent material type and texture, climate or vegetation (see also Sections 1.2, 2.2 and 3.3.1). The scale and level of detail of these input predictor layers have varied in response to both availability of input data and the nature and scale of the target features to be predicted. A limited number of discrete scales for which predictive mapping has been attempted can be recognized and associated with specific grid resolutions of input data layers (Table 10.1).

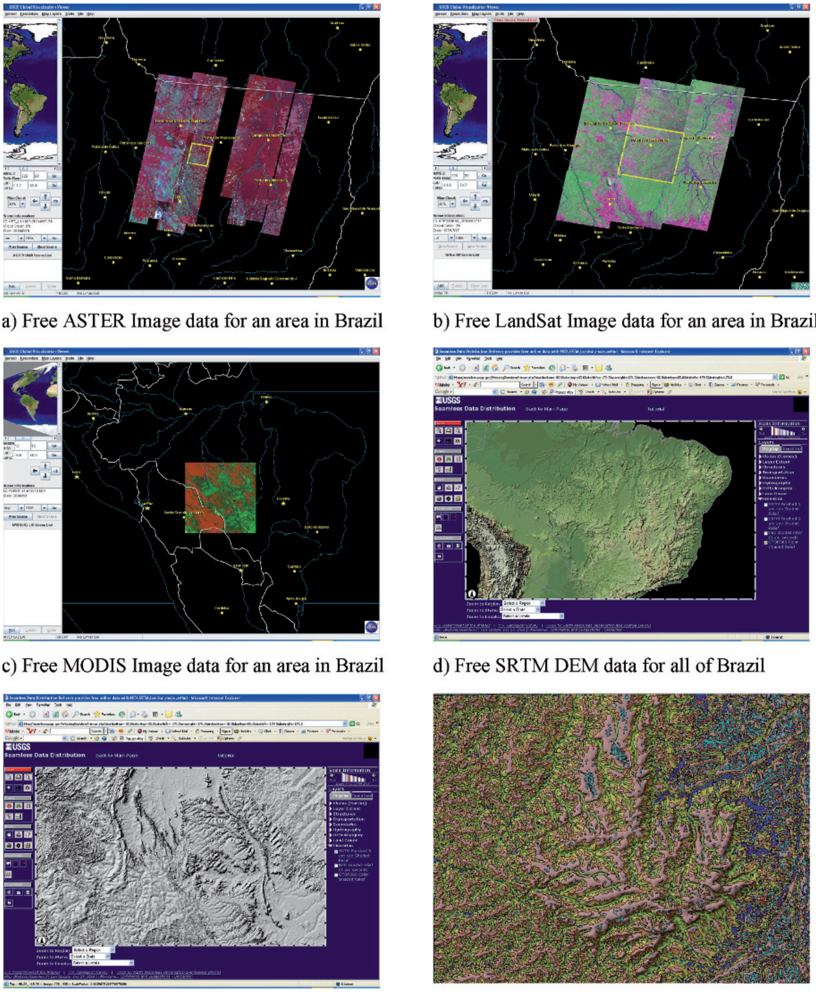
Many potential practitioners of DSM in regions with sparse spatial data infrastructures have expressed frustration with the lack of availability and lack of quality of predictor data layers, particularly DEMs, for their areas of interest. Clearly, many developing regions lack access to fine (5–10 m) to medium (25–50 m) resolution DEM data layers, but efforts such as those of Moran and Bui (2002) and Bui et al. (2002) have demonstrated that useful and effective regional predictive maps can be produced using DEM data with a grid resolution as course as 250 m. Their key realization was that local 3×3 window variables (slope, aspect, curvatures) were not very meaningful or useful predictors when derived from course resolution DEM grids but that measures of regional hydrological context and of variation in elevation, slope and morphology within larger search windows of varying size did prove to be useful.

The point here is that it is important to compute and use derived variables or predictive measures that are matched, or suited, to the dimensions of the DEM that is available, the size and scale of the output entities to be predicted and the intended use of the resulting map. Most areas now have access to free 90 m SRTM DEM data (Fig. 10.2d,e) and should also be able to order 30 m ASTER DEM data at a reasonable cost (for examples see Sections 2.2.1, 4.2 and 16.1). Admittedly these DEM data sets have limitations in that they often portray a mixture of the bare ground surface and a canopy surface in areas of dense tree cover. However, they can still provide valid and useful measures of regional context, relative drainage conditions and local surface form and orientation. With proper care not to assume the surfaces portrayed by these data are exact representations of the true topographic surface, these medium to moderately coarse resolution DEM data sets can be highly useful for making predictions at scales as fine as 1:50,000 to 1:100,000. Finally, increasing

Table 10.1 Major types and sources of input data for DSM at different scales and grid resolutions

Grid Size(m)	<5m	5–10 m	25–50 m	100–250 m	>250m (1000m)
Scale	1:1,000–1:5:000	1:5,000–1:20,000	1:20,000–1:100,000	1:100,000–1:250,000	1:1–1.2 M
Application areas	Site-level operational management	Site-level operational management to detailed planning	Non site-specific regional planning, Management of minimum sized areas	Regional planning and policy development	State to national planning and policy development
DEM Data sources	LiDAR, DGPS, Photogrammetric, lkonos 2–5 m	USGS 10 m from contours, LiDAR, Photogrammetric SPOT 5–10 m	Photogrametric SRTM 30 m, ASTER 30 m, SPOT 20 m	SRTM 90 m, DTED Level 1 90 m, National DEMs from contours	SRTM 90 m generalized GTOPO30, DTED Level 0
Image Data Sources	SPOT 5 2–5 m, EROS 2–10 m, IKONOS 1–4 m, Quickbird 1–2.5 m	IRS 5–20 m, SPOT 4 10–20 m, ASTER 15 m, Landsat7 15 m	Landsat 7 15–30 m, SPOT MS, ASTER 30 m	MODIS 250 m, Landsat 4 80 m,	AVHRR, MODIS 500–1000 m
Ancillary Data Sources	On-site field sensing such as EM, GPR	On-site field sensing such as EM, GPR	Airbome geophysics for parent material depth and texture	Regional Bedrock and Surficial geology maps	State to national bedrock and surficial geology maps
	Site specific direct field measurements	Airbome thermal or multi-temporal to detect local climate	Airbome thermal of multi-temporal to detect local climate	Manually prepared maps of vegetation or land system	Manually prepared maps of physiographic regions

* Some useful sites for downloading SRTM and ASTER DEM data and satellite imagery
Corrected SRTM: http://srtm.csi.cgiar.org
Corrected SRTM: http://www.ambiotek.com/topoview
Hydro-corrected SRTM: http://hydrosheds.cr.usgs.gov/
ASTER DEM: http://edcimswww.cr.usgs.gov/pub/imswelcome/
USGS Seamless SRTM and Image Data: http://seamless.usgs.gov/
Canada Toporama Satellite Imagery: http://toporama.cits.rncan.gc.ca/toporama_en.html
UK LandMap DEMs and Satellite Imagery: http://www.landmap.ac.uk/
US NOAA/NGDC GLOBE DEM: http://www.ngdc.noaa.gov/mgg/topo/gltiles.html
US National Map, DEMs, Imagery, vectors: http://nmviewogc.cr.usgs.gov/viewer.htm



a) Free ASTER Image data for an area in Brazil
b) Free LandSat Image data for an area in Brazil
c) Free MODIS Image data for an area in Brazil
d) Free SRTM DEM data for all of Brazil
e) Free SRTM DEM data for an area of Brazil
f) A Simple landform classification of SRTM data

Fig. 10.2 Illustration of access to and use of free spatial data that is widely available for most areas, even those considered to have sparse spatial data availability (See also Plate 13 in the Colour Plate Section)

availability and lower costs for LIDAR DEMs of very fine spatial resolution may rapidly turn the problem from one of insufficient amounts of data to one of more data than can be handled effectively.

Even in areas of sparse spatial data infrastructures, the quantity and quality of remotely sensed image data is rapidly improving and free data at resolutions of 30–90 m is becoming widely available (Fig. 10.2a,b,c). These data are adequate to support initial regional scale (1:25,000–1:100,000) predictive mapping efforts (see Fig. 10.2e) (for examples see also Chapters 26, 30 and 34). The author has found that ancillary information describing regional scale variation in environmental conditions such as bedrock and surficial geology, vegetation, climate, and physiography can usually be extracted from available secondary source thematic maps or can, if necessary, be interpreted manually using manual visual analysis of available imagery and DEM derivatives. However, at finer resolutions, improved predictive abilities will require more reliable and spatially precise information on parent material texture and depth, perhaps obtained from analysis of airborne radiometric data (see Chapters 14 and 15), on local variation in climate, perhaps obtained from analysis of thermal or multi-temporal imagery and on other subsurface conditions, such as soil depth, salinity or moisture content, that may be detected and mapped using proximal field sensing tools (see Chapters 2 and 13 for examples). Ancillary data sources can be vital inputs for predictive mapping and they can often be approximated using manual interpretation, if necessary or obtained directly using additional sensing technologies, if available.

Of the two main types of data required to support DSM, it is therefore the author's opinion (supported by comments in Sections 1.2 and 20.1) that the most limiting is that which permits description and elaboration of rules that describe the spatial arrangement of soils in the landscape and the conditions or criteria that control this distribution. Field observations of soil-landscape relationships or well developed tacit knowledge of these relationships are essential to support construction, application and review of classification rules for predictive maps. In areas of sparse spatial data, the most important requirement may well be to find ways to collect and assimilate information on soil-landscape patterns so that this can be related to available input data layers in digital format to create predictive rules. (examples of approaches to building knowledge of soil-landscape relationships are presented in Chapters 9, 20 and 25)

10.5 Protocols or Methods for Developing Predictive Rules for DSM

Having decided upon, and conceptualized, what to map and selected and obtained suitable input data it is then necessary to develop procedures to predict outcomes given predictor input maps and suitable training data. Some of the main approaches that have been reported for developing rules to classify soil objects to produce area-class maps are reviewed below. The three main approaches used to develop and apply classification rules (Fig. 10.3) can be categorized as unsupervised, supervised and heuristic (or knowledge-based) (see also Sections 1.4 and 26.2).

10.5.1 Unsupervised Classification Approaches

Unsupervised classification approaches are data driven and make the least use of local expert knowledge or judgment. Still, users influence the resulting classification

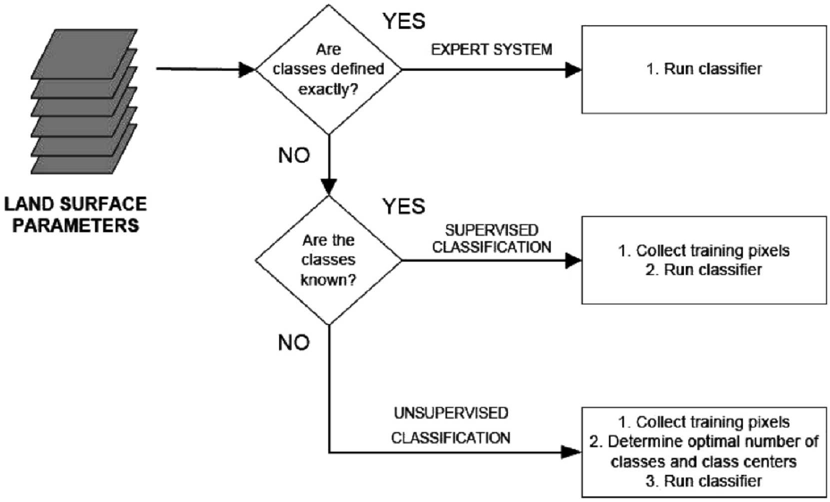


Fig. 10.3 Schematic diagram illustrating the most widely-used approaches for developing and applying rules for classifying soil entities using DSM methods

in several ways. The choice of what input variables and what sample locations to make available for input into unsupervised classification procedures can greatly influence the type and nature of classes that result. User identification of the total number of classes to predict also exerts an influence on the final classification as does selection of the clustering or classification algorithm implemented by the procedures. Finally, users must exercise local knowledge or judgment when assigning descriptions or attributes to each output class produced by a supervised classification.

The principal advantages of an un-supervised classification approach are that it is systematic and unbiased and it generates classes that exhibit a maximum amount of difference with respect to the input variables used in the classification. The principal disadvantage is that, since it is data driven, it is almost impossible to produce classes that closely match those defined for a locally derived heuristic classification system. Unsupervised classifications are also not generally capable of differentiating classes that exhibit only subtle, but often significant, differences with respect to one or more input variables or site conditions.

Unsupervised approaches are among the least commonly used protocols for developing and applying rules for producing soil class maps for operational use. However, the approach has been demonstrated to be capable of producing useful and meaningful classified maps by Burrough et al., (2000, 2001); Irwin et al., (1997) and others (see Chapter 26). If one does not have strong pre-existing knowledge of the main classes (of soil) that occur in an area and of the conditions or criteria that control the spatial distribution of these classes, then unsupervised approaches can help define optimum classes and map their spatial extent.

10.5.2 Supervised Classification Approaches

A number of different forms of supervised classification have been used to develop and apply rules to produce soil class maps. All of these approaches can be considered to represent a form of data mining. All data mining approaches extract rules for consistently recognizing class entities by analyzing training data from representative sites or example areas to detect relationships between each desired output class and a set of predictor variables or classes from a suite of user-selected input maps. In this approach, the map producer is called upon to exercise local expert judgment and opinion in selecting and classifying a sufficient number of reference sites or example locations for each and every class that may exist in an area. The map maker is considered able to recognize an instance of a class when it is encountered in the field, or observed on a display, but is not considered able to completely and systematically identify the environmental site conditions or rules that control the spatial distribution of any given class. In order to use this approach, a map maker must be able to obtain or generate a large number of reference sites for each class of interest. These reference sites should ideally encompass the full range of environmental conditions under which each class of interest is known to, or can, occur. There must be a sufficient number of reference sites for each class and the reference sites must be selected in as valid a manner as possible so that they do not provide a biased representation of the conditions under which a class occurs.

The principal disadvantage of most supervised classification approaches is that they require assembly of large data sets of spatially-located reference or training data. This assembly can be time consuming and expensive if it requires field sampling. If it involves on-screen selection of training locations, classes are not verified by field observations. Classified training sites can lead to misleading or false rules if the spatial size (support) of the reference locations is incompatible with the resolution and spatial accuracy of the main input data layers (e.g. mainly the DEM grid mesh). The principal advantage of all supervised approaches is that they provide a formal, systematic framework to identify which values of which input variables or classes are most strongly associated with (predictive of) each desired output class. These rules can uncover local tacit knowledge about where in the landscape certain soils are most likely to occur and why and can codify this knowledge systematically, formally and quantitatively (for examples see Chapters 2, 19 and 32).

Decision trees have been shown to be capable of extracting classification rules by analyzing patterns in input data values relative to classified reference areas (Bui et al., 1999; Bui and Moran, 2001; Lagacherie and Holmes, 1997; Moran and Bui, 2002; Scull et al., 2003, 2005; Zhou et al., 2005; Zambon et al., 2006; also Section 2.3). Decision trees work by splitting data sets into more homogeneous subsets. Splitting rules attempt to minimize entropy or variance within reference sites included in any node produced by a split. Different kinds of splitting rules can be used (Zambon et al., 2006) but the objective is always to identify splits that produce more homogeneous groupings for each node of the tree. Advantages of decision trees include the fact that they require no assumptions about the data, they can analyze both classified and continuous data, they can deal with non-linearity in input data and they are quite easy to interpret. Decision trees can be used to predict the spatial patterns of either classed entities or continuous variables. These predictions can have attached to them an estimate of the likelihood of occurrence of the predicted class (or individual property value) at each node based on consideration of the proportion of each node that is occupied by instances of the class or value being predicted. Decision trees leave behind a set of splitting rules for each node that can be easily interpreted to understand the criteria used to define each node and the conditions that are likely to occur within each defined class. Decision trees may not produce desired results if the selected predictor variables do not exhibit a strong spatial relationship with the classes or values to be predicted (as for example, in trying to use local measures of slope gradient and curvature to predict regional patterns typical of soil associations or complexes – see Moran and Bui, 2002 and Bui and Moran, 2003 for a discussion of this).

Bayesian analysis of evidence (BME) has been used to extract rules for recognizing soil classes by Bui et al. (1999); Cook et al. (1996); Corner et al. (1996) and Zhou et al. (2005). Bayesian analysis provides two useful sets of information that relate classes on input maps to predicted output classes. It provides a systematic method for quantifying the relative utility, or predictive strength, of each input layer relative to all other layers available to predict output classes. It also provides quantitative values for the probability of occurrence of any given evidence class given each of K possible output classes. Additionally, final predictions of expected output classes are constrained by consideration of *a priori* estimates of the relative proportions of each of K possible output classes, such that the final extent of each predicted output class matches the proportional extent provided by the estimate of prior probabilities. A limitation of BME is that it analyzes the frequency of occurrence of ***classes*** of input data relative to desired output classes. Therefore input layers of continuous variables must first be generalized into classes with decisions on the number of classes and the class boundaries having a potentially large effect on the subsequent results. Some advantages of BME are that it can produce estimates for the relative likelihood of occurrence of every defined output class at every location and it can identify and quantify which layers of input data are most useful for predicting output classes and which classes on each layer are associated with the highest likelihood of occurrence of any given output class of interest. BME is a powerful data mining tool that can uncover and quantify relationships between input data layers and output classes to be predicted (similar conclusion in Section 2.3.3).

Supervised approaches based on application of fuzzy logic have been described by Odeh et al. (1992); De Gruijter and McBratney (1988) and by several implementations of the SOLIM approach of Zhu (1994) (e.g. see Shi et al. (2004) (see Chapter 20).

10.5.3 Heuristic (Expert Knowledge) Classification Approaches

Heuristic approaches can be used where the user has identified all classes that are to be predicted and also has a well developed set of criteria that describe the conditions under which each potential output class may occur. The expert knowledge about what classes exist and what criteria control their spatial distribution may arise from extensive local field experience and analysis of local field observations to create locally appropriate classification rules (for examples see Qi and Zhu, (2006) and Chapters 9, 19, 20 and 31). Alternately, classes and classification criteria may be defined exclusively on the basis of theoretical knowledge and theoretical considerations about what parts of the landscape are likely to exhibit different environmental conditions and why.

Most implementations of the SOLIM approach of Zhu (1994) and Zhu et al. (2001) are based on capturing and applying local expert heuristic knowledge as fuzzy rules. SOLIM compares the values for predictor variables at each unclassified location to the values for those variables for a very limited number (1 or 2) of user-defined "instances" that define the central concepts of each class of soil to be mapped (Chapter 20). Each predictor variable is compared to the value of the variable for each defined instance and a degree of similarity is computed based on a similarity function. If the value of the predictor variable at the unclassified location is the same as for an instance, then the similarity of the two sites with respect to that variable is 1. Any difference in value can lead to calculation of a lower degree of similarity between the unclassified location and the instance with respect to that variable. In SOLIM, each unclassified site is compared to each instance with respect to each of N user-specified predictor variables. In SOLIM, the smallest or lowest value for fuzzy similarity from among consideration of all N predictor variables is used to establish the overall fuzzy similarity of an unclassified site to a particular reference site. The fuzzy similarity of a given unclassified location is computed for each of M available instances (usually less than 3) of a particular output class K and the highest value of fuzzy similarity is chosen via a max function to represent the similarity of the unclassified site to a particular output class K. A final hard classification is achieved by identifying which of the K output classes has the highest value for fuzzy similarity to any given instance and assigning this classification to that unclassified location. Published descriptions of these fuzzy methods do not make it clear whether the procedures can be made hierarchical so that every possible output class does not have to be predicted for every possible location in the data set. However, Zhu (personal communication, 2006) has indicated that different sets of rules for different groupings of instances are commonly developed and applied within different major land areas and within different types and scales of landforms.

MacMillan et al., (2007) describe another example of a fuzzy heuristic approach that identifies and maps classes whose existence and defining criteria have been recognized based on extensive local field experience and analysis of field observations. In this subjective classification, the criteria and conditions that control the spatial distribution of the desired classes have been described in a classification field guide and the principal requirement is to translate this existing set of classification rules into a corresponding set of formal quantitative machine rules that can be applied to digital inputs to predict the required output classes. Implementation of this approach is not strikingly different from supervised classification approaches described above. Fuzzy semantic rules are developed, applied, evaluated and revised in an iterative manner until such time as the spatial distribution of predicted output classes corresponds closely to an expert's expected distribution of those classes. This is conceptually similar to asking the same expert to select a large number of training sites that are deemed to be representative of each class of interest except here, instead of using the training sites to develop rules, we use successive refinements of heuristic rules to classify entire areas that are then treated as a single large set of training data and are reviewed to see if the resulting patterns correspond with expert expectations.

Shary et al. (2002) and Sharaya and Shary (2004) describe examples of a comprehensive system of classification of surface curvatures based entirely on objective theoretical considerations of expected relationships between curvature classes and anticipated environmental conditions. This approach demonstrates that it is possible to impose a set of theoretical classification rules even without any local, empirical knowledge to guide definition of classes of interest. The resulting maps are anticipated to differentiate portions of the landscape that can be expected to exhibit significant differences in soil processes and in patterns of development of soil properties and soil classes. So, if local expert knowledge of actual patterns of soil distribution is weak or absent, it may still be possible to produce useful maps based on application of theoretical considerations only.

10.6 Protocols or Tools for Applying Predictive Rules for DSM

Just about every published description of DSM procedures has used a different, and often a custom, application engine or computational tool to apply whatever predictive rules were developed to predict output classes from selected predictor input data. This is a concern for new practitioners trying to apply the methods themselves and trying to decide which protocols might be most applicable to their particular needs. Most described application engines are not tightly integrated into the GIS environment used to collate the input data, develop the rules and display the predictive output results. This is an area of concern as it restricts rapid appraisal and adoption of predictive mapping techniques. A considerable number of free or low cost software products have emerged that offer capabilities to compute a wide variety of predictor variables from a DEM data (e.g. TAS, SAGA, ILWIS, TauDEM, TAPES, TOPOG. TOPAZ) but most do not contain integrated classification engines for applying predictive classification rules. Some (SAGA, ILWIS, TAS) do provide some generic classification functionality but it has not yet been configured to easily develop and apply classification rules for predictive mapping. This is one area where it is hoped that some vertically integrated tools will emerge that can be used to apply the full sequence of predictive mapping activities from data preparation to rule development to application of rules to final map production (similar comments in Section 4.3.2).

10.7 Assessing the Quality of Input Data Layers for DSM and of the Resulting output maps

Potential users of predictive maps have expressed concerns about the quality of these maps in comparison to existing maps produced using traditional mapping methods.

The quality of predictive maps can be considered to be a function of the accuracy and relevance of the input layers used to produce the maps, the effectiveness of the predictive procedures used to create the maps and the thematic and positional accuracy of the final resulting maps.

10.7.1 Considerations of the Quality of Input Maps used to Make Predictions

Consider first issues of quality of the predictor input maps. With respect to the primary inputs derived from a DEM, considerations of quality have more to do with the ability of the DEM to faithfully portray the location, shape, size and pattern of surface features of relevance to the predictions than to any measures of absolute elevation accuracy, such as the commonly used root mean square (RMS) error in DEM elevation values relative to measured elevations at specific locations. The DEM needs to faithfully render relative point to point relations that capture and portray surface form at a scale and resolution appropriate for describing the landform entities that are to be mapped. The degree to which a DEM faithfully portrays a surface is related to its horizontal and vertical resolution and accuracy. Shary et al. (2002) showed that local measures of surface gradient and curvature computed within a 3×3 window were strongly influenced by grid resolution, with slope tending towards zero for large grid spacing and towards extremely large values for very small grid spacing. Moran and Bui (2002) noted that local measures of surface form (slope, aspect, curvatures) computed within a 3×3 window became less meaningful and useful as predictors of soil classes or properties as grid dimensions increased but that more regional measures of context or pattern (such as upslope catchment area) were less sensitive to grid resolution and were more reliable inputs for predictions of patterns of soil associations that used grids with larger horizontal dimensions. Coarser resolution DEM data sets are therefore best used to compute measures of regional context and texture or variance within a neighbourhood analysis window and are less useful for computing local measures of surface form (slope, aspect, curvatures).

Within the predictive mapping community, a consensus appears to have emerged that grids with a horizontal resolution of 5–10 m and a relative vertical accuracy of $\pm$0.5m or better appear to capture and portray meso scale terrain features at about the level of abstraction that they are most commonly appreciated by human observers (note however that in Chapter 4 Howell et al. did not achieve significant improvements using a 5 m DEM compared to a 25 m DEM). Grids of coarser resolution do not appear to capture and describe the correct location of landform features at the scale of hillslopes or portions of hillslopes that are of most common interest for predictive mapping. Coarser grids (25–100 m) can be thought of as a kind of regular sampling frame or mesh that can provide some relevant information on the approximate vertical range in elevation within a local neighbourhood and on relative values for slope gradient and curvatures. These values will always be under-estimates, because not all local terrain maxima and minima will be sampled, but the mesh will provide some relevant information on the size, scale and complexity of landform features that are at least 2 times the horizontal dimensions of the grid. Consequently, when using grids of coarser horizontal resolution (>10m), DSM practitioners should explicitly acknowledge that any predictions or classifications that are based on analysis of local surface form or context are not likely to be spatially accurate at point locations. The locations of local rises or hollows portrayed by a DEM of 25 m grid dimensions or more are likely to be displaced relative to the actual locations of these features in the real world and smaller local rises or hollows will be missed altogether. However the 25 m DEM data can be useful in indicating the frequency of occurrence of features of a particular relative size and scale within any small area.

Evaluations of the quality of input data layers should focus on establishing the size, scale, shape and context of terrain features that can be accurately described by an input layer of a particular resolution. The features to be predicted by analysis of these input layers need to be conceptualized as having horizontal dimensions that are at least 2x those of the grid postings. Any variation in the predicted classes or attributes that occurs over distances less than 2x the grid interval cannot be spatially located with any accuracy and can only be described in general terms. Another quality consideration is that many existing coarser resolution DEM sources do not provide a consistent and faithful portrayal of the bare ground surface, but rather, portray a digital surface model (DSM) that often describes the top of a forested canopy, in densely forested areas, or the tops of man-made features in built-up areas. Finally, many existing secondary source environmental maps are inadequate and need to be improved.

With respect to soil maps produced by predictive methods, quality can be defined as the ability of a map or product to correctly predict the characteristics of the landscape at particular points or within particular small areas (see also Section 1.5).

10.7.2 Considerations of Quality of Predictive Maps of Individual Soil Properties

The quality of maps that predict individual soil properties can be assessed by obtaining field observations or samples at randomly selected locations and computing RMS error between the predicted and observed values (see also Chapter 11). Uncertainty associated with the predictions of individual soil properties can be conveyed by preparing and presenting maps of the residuals arising from the predictive process. An alternative approach is to produce multiple realizations of each predictive map by varying the values of the input variables randomly with the range of expected accuracy of the input layers at point locations. The variation in predicted values observed in these multiple realizations can provide an illustration of the uncertainty of the predictions at any given location. Multiple realizations can also be achieved by using different predictive equations or techniques to produce each realization so as to illustrate the range of uncertainty in predictions arising from method error, as opposed to data error.

10.7.3 Considerations of Quality of Predictive Maps of Discrete Soil Classes

The quality of maps that predict discrete soil classes can be assessed in several ways. Kuhnert et al. (2005) Wang et al. (2005) recognized a need to distinguish between positional errors and thematic errors in raster maps. Similarly, Walker (2003) assessed accuracy in terms of ability to predict the correct classes at exact locations, termed locational ability and ability to predict proportions of classes within an area without consideration of location, termed predictive quality. The global accuracy or precision of a map is traditionally assessed by a preparing a contingency table that compares predicted values at specific sample locations to observed values at those same locations. The degree of similarity (or error) is then assessed by computing a measure such as the Kappa statistic that corrects for agreement that may arise from chance (see Section 19.2).

A problem that has been recognized with Kappa and its modifications is that it is entirely based on cell by cell comparison statistics (Kuhnert et al. (2005). Maps that have a bias or have similar patterns that are slightly distorted or misregistered may not agree well. The textbook example of this is a comparison of two identical chess boards displaced by exactly one cell in one direction. A point by point comparison would conclude that the two maps had zero similarity whereas, in reality, the patterns they predict are identical but the locational accuracy is off by one cell. Increasingly, new methods of assessing accuracy have been proposed that can identify and quantify positional and thematic accuracy errors separately (Pontius, 2000, 2002). The degree of positional error can be assessed by comparing proportions of classes estimated within search windows of ever increasing dimensions with proportions of classes on a map taken to represent ground truth. Kuhnert et al. (2005) calculated an index based on the difference between the total numbers of cells in each category in each size of moving window and a reference map. If the proportions of each class in two windows were identical the index was zero. The more mismatches identified, the larger the index. The window was systematically increased in size and new comparisons were made. As the window size grew, the granularity of maps was blurred, and eventually they obtained a perfect fit assuming that the numbers of cells in the same category was the same for the whole area for both the predicted and ground truth maps.

The thematic accuracy of categorical soil maps is typically assessed in terms of exact categorical match between specific predicted classes and hard classes observed at reference locations. This approach ignores the fact that soil varies continuously across the landscape and that soils at a particular location may be more or less similar to the central concept used to define any mapped class. This is the underlying assumption of fuzzy methods of soil classification and should be familiar to DSM practitioners by now. Liem et al. (2005) and Metzler and Sadler (2005) have both suggested methods for computing relaxed measures of agreement between hard classes and reference classes by adopting a fuzzy matching definition for a crisp classification which allows for varying levels of set membership for multiple map categories. In this approach, fuzzy sets were used to represent the "relative strengths"' of various membership categories for predicted classes relative to observed classes at a mapped pixel level (Metzler and Sadler, 2005).

Many applications of soil maps, especially of smaller scale soil maps, are non site-specific. For these maps, estimates of the proportions of particular soils within small areas equivalent to delineations of some minimum size for which management decisions are to be taken may be all that is required. Exact thematic accuracy at exact point locations is not necessary for these maps to be useful. Consequently, tests of map accuracy that evaluate the degree of exact match between predicted class and observed class at exact locations are neither appropriate nor desirable. We know that hard classifications assigned to reference or test locations in the field by local experts have to reflect decisions about what class best describes each reference location. We know that different experts are highly likely to assign the same location to different classes if the soil at the location is ambiguous and not completely representative of a specific class. These potential differences in classification of reference or test locations by a local expert represent sampling error in evaluation data sets which is not taken into account by any measures of exact categorical match at exact locations. Similarly, consider that many of the coarser resolution DEM data sets used to produce predictive maps do not portray the exact location of terrain features of interest correctly but displace them or only partially capture them. It is clear that this inaccurate depiction of landform features leads to spatial displacement of predicted classes relative to their true locations. Comparisons of thematic map accuracy should therefore assess relative (fuzzy) degrees of thematic correspondence between predicted and observed classes. They should also assess the degree of correspondence of predictions of proportions of classes within areas of different size, as well as just exact correspondence at specific point locations.

In general terms then, assessment of the accuracy of classed soil maps needs to take into account the scale of mapping and the intended use of the maps. The quality of fine resolution (large scale) maps intended for use for site specific operational management decisions may need to be evaluated for exact thematic match at exact point locations. The quality of coarser resolution (small scale) maps intended to support regional management or operational decision making for management areas of some minimum size that is larger than a single pixel may only need to be evaluated in terms of relative degree of predictions of proportions of classes within some minimum sized area of interest for management decisions. Lagacherie (Section 1.7.2) has recognized a need for improved and more formal protocols for assessing accuracy and error in predictive maps.

In the author's experience, while the pixilated appearance of raster predictive soil maps has often been criticized by users more accustomed to traditional, cartographically precise, polygonal soil maps, the accuracy of raster predictive maps has uniformly proven to be equal to, or superior to, that provided by manually prepared vector maps for the same areas. Zhu (see http://solim.geography.wisc.edu/projects/index.htm) has cited values for exact spatial and categorical match predictive accuracy of 77–89% for soil maps produced using the SOLIM method. Moon (2005a,b) reported accuracies of estimates of proportions of classes within small areas of 66–71% for predictive maps that compared favourably to accuracies of 55–60% for manually prepared vector maps for the same areas. Explicit evaluation of the accuracy of predictive maps, using a method appropriate to the scale of the maps and their intended use, is likely to document that these maps are at least as accurate as conventionally produced manual soil maps and that they are usually more accurate (see also Sections 1.5 and 7.1).

10.8 Building Capacity for Routine Operational Application of DSM

Finally, let us consider here opportunities and constraints for building capacity for routine operational application of DSM procedures.

A typical justification for not embracing and using DSM technologies is that most existing DSM procedures require expertise, data and software capabilities that are difficult to acquire and in short supply. It is easy for experienced DSM practitioners to forget how much they had to learn before they felt comfortable pursuing and applying DSM procedures. Potential new practitioners need access to documentation and training that will provide them with the necessary background skills and the confidence to use them. Training of new users has been recognized as a key requirement for increasing use of the SOLIM approach. Similarly, new users of the LandMapR toolkit have found it beneficial to undertake a few days of training in order to gain the necessary skills and confidence. It is incumbent on new practitioners to seek out training opportunities and on existing practitioners and software developers to provide it.

The author does not believe that either lack of data or lack of software capabilities presents a serious impediment to implementation of DSM activities. Even localities that lack access to fine resolution DEM data do tend to have access to some DEM and image data at reasonable resolutions of 25–100 m. These widely available free spatial data are perfectly adequate for applying and evaluating DSM concepts at a regional level at scales of 1:50,000 to 1:250,000. A lot can be learned and a lot accomplished by applying initial DSM efforts at these coarser resolutions (See Fig. 10.2f).

A number of free, or nearly free, software platforms exist that provide very full featured capabilities for processing input data sets, particularly DEMs, to compute a wide range of predictor input variables. Tools to analyze training data sets to extract classification rules are not commonly included within available GIS platforms. Such analysis tends to need to be conducted outside a GIS platform within some kind of specialised software (S-Plus, Expector, See 5, Netica by NorSys, Systat, SPSS) and the resulting rules implemented using a generic GIS calculator. In general, tighter integration of spatial data mining capabilities into available GIS platforms would be helpful in speeding up the development and adoption of DSM procedures. Similarly, most existing GIS platforms provide a calculator-like capability to apply Boolean classification rules (such as those generated by decision tree analysis) but most do not provide well developed engines for applying Bayesian or Fuzzy classification rules (except perhaps Idrisi). This shortcoming could be addressed by adding specialized extensions to any of a number of existing free GIS platforms or by acquiring the free SOLIM platform or the restricted availability LandMapR toolkit. Software and data do exist to support operational application of DSM procedures but the potential DSM practitioner has to be willing to work relatively hard to acquire and learn to use and apply both at the present time (see also Section 1.7.1).

A second requirement for building capacity for DSM is that of training potential users and funders of DSM products to better understand the relative costs, benefits, uses and limitations of DSM products. It is not uncommon to encounter scepticism and resistance to acceptance of DSM methods or products among institutions, organizations or companies that have a long history of using map products produced by conventional manual interpretation procedures. The raster DSM maps don't look like the neat, cartographically-precise, vector maps they have been accustomed to obtaining and using. There is often considerable suspicion of the "***auto-magic***" nature of DSM procedures. There is support for known products produced by conventional means and suspicion of new products produced by unconventional means.

The author has some experience with overcoming initial institutional resistance to DSM and converting it into active approval and support. The key to converting suspicion and opposition into active support was to demonstrate conclusively and convincingly to the various sceptical stakeholders that the new DSM methods could be relied upon to produce maps that were as accurate as, or more accurate than, any maps produced by existing methods and that this could be done for a fraction of the cost of existing manual methods or of other alternate automated methods. The approach taken was to devise and implement a systematic, staged process to apply different methodologies and to evaluate the relative predictive accuracy of various methods, their relative costs and the utility of the resulting products (see also Section 1.4).

Initially, a pilot project was undertaken for a relatively small area in which as many different approaches to producing the desired map products as it was possible to identify and implement were applied to map the exact same area of several map sheets. A comprehensive record was maintained of production time requirements and costs associated with each of the applied methods. A procedure for evaluating the relative predictive accuracy of each of the methods was devised and the same methods and data were used to evaluate the relative accuracies of all maps produced by each of the evaluated mapping methods. The procedures for evaluating accuracy were designed to enable a fair and equitable comparison of all maps produced by all methods. Since several of the methods tested produced polygons that were described in terms of the proportions of classed entities contained within their boundaries, it was necessary to devise a test procedure that could evaluate estimates of the proportions of predicted classes within a small test area. The evaluation procedures were never intended to measure or compare exact thematic correspondence at exact point locations and this was made abundantly clear to all stakeholders. This measure of accuracy was justifiable in view of the intended use of the resulting maps.

None of the maps was expected to be able to support detailed use for site-specific operational management but they were expected to be able to provide estimates of the proportions of map classes within small areas, equivalent in size to a minimum sized management area for forest management activities (e.g. a stand of 10 ha or more). Each of the several different methods selected for evaluation was applied to map the small (45,000 ha) pilot area and the maps resulting from application of each method were compared in terms of their recorded costs, time requirements and relative predictive accuracies. In this initial test, only one method produced maps that met the requirement for 65% minimum accuracy in predicting the proportions of map classes within minimum-sized test areas.

A second project was conducted to assess the ability of the mapping protocol identified as having the highest predictive accuracy to be applied successfully to a much larger area on an operational mapping basis. This was done to increase the confidence of the stakeholders that the identified methodology was capable of being scaled up for use for full-scale operational mapping and also to provide a second opportunity to evaluate the predictive accuracy of the maps produced by this method. Demonstration of an ability to meet the required minimum level of accuracy for a second project was believed to be beneficial for increasing the level of comfort of the stakeholders in the procedures prior to committing significant funding to map much larger areas at a much greater level of expense and risk. The second project again demonstrated an ability to exceed the minimum required level of predictive accuracy at a cost that was significantly lower than any previously implemented alternatives, including conventional manual mapping. The stakeholders were beginning to develop confidence in the new procedures. Activities were also undertaken to encourage and facilitate use of the initial map products so that stakeholders could feel assured that the maps would meet their needs for assessment of resources and planning for their management.

The final phase of this multi-step process was to proceed to full scale operational mapping of areas millions of hectares in extent (see MacMillan et al., 2007). Even this activity has been staged with mapping completed in blocks of 1–3 million ha per year. This reduces the risks and spreads the costs and the demands for personnel out over a period of several years. With completion of each successive operational project and successful production of maps that exceeded the minimum requirement for predictive accuracy, stakeholder support for, and acceptance of, the new predictive maps has increased. Most sceptics have now been convinced and have become active supporters of the methods.

10.9 Conclusions

A similar staged process to that described above is recommended to any DSM practitioner needing to build support for, and comfort with, adoption of new DSM technologies. Potential DSM practitioners are encouraged to apply and evaluate all viable mapping alternatives, to select the most appropriate method and to then demonstrate its capability to support full scale operational mapping in their locality at a scale that is most relevant to their needs.

Based on the author's experiences as described above, the single strongest impediment to widespread application of new DSM methods may well be our own hesitancy to believe in ourselves and to just "do it". You will not know what is possible until you try to produce maps for your own areas using data that are available to you. You may well be surprised at what can be achieved using existing data sources and existing methodologies. If we wait until perfection is possible, data are easy to acquire and models are easy to build and apply, we could end up waiting a long time. For the moment, a lot is possible with just a little effort and a little optimism.

It is hoped that this chapter will encourage individuals with an interest in applying new predictive mapping techniques to embrace change and to try to create useful, operational maps for large areas in their own regions of interest.

Acknowledgments The comments and analysis contained in this chapter are based largely on experience developing and applying predictive ecological mapping techniques in the province of BC, Canada. This predictive ecological mapping work was initiated and conceptualized by Dr. David Moon, managed by Nona Philips and financially supported by Tolko Industries Ltd., West Fraser Mills Ltd, and several other companies that are now assimilated into these two main forest industry clients. Ray Coupé provided all of the local ecological expertise that was used in developing methods and rules.

References

Bui, E. (Ed.)., 2000. A soil information strategy for the Murray Darling Basin (MDBSIS). URL: http://www.brs.gov.au/mdbsis/publications.html Last accessed June 3, 2004.

Bui, E., 2004. Soil survey as a knowledge system. Geoderma 120, 17–26.

Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. Geoderma 103, 79–94.

Bui, E.N., Moran, C.J., 2003. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. Geoderma 111, 21–44.

Bui, E.N., Loughhead, A., Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. Australian Journal of Soil Research. 37, 495–508.

Bui, E.N., Henderson, B., Moran, C.J., Johnson, R., 2002. Continental-scale spatial modeling of soil properties. Proceedings of the 17th World Congress of Soil Science, Bangkok, Thailand, August 14–20, 2002. Symposium no. 52, paper no. 1470, 8 pp.

Burrough P.A., van Gaans P.F.M., MacMillan R. A., 2000. High resolution landform classification using fuzzy-k means. Fuzzy Sets and Systems 113, 37–52.

Burrough, P.A., Wilson, J.P., van Gaans, P.F.M., and Hansen, A.J., 2001. Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. Landscape Ecology 16, 523–546.

Carré, F., McBratney, A.B., and Minasny, B., 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141 (1–2):1–14.

Cole, N.J., Boettinger, J.L., 2004. Pedogenic understanding raster-based classification methodology for mapping soils. Proceedings of the Western Regional Cooperative Soil Survey, June 13–18, 2004, Jackson, Wy. USDA, NRCS and U.S. Bureau of Land Management. CD-ROM.

Cook, S.E., Corner, R., Grealish, G.J., Gessler, P.E., Chartres, C.J., 1996. A rule-based system to map soil properties. Soil Science Society of America Journal 60, 1893–1900.

Corner, R.J., Moore, G.A., Grealish, G., Gessler, P.E., and Chartres, C.J., 1996. A manual for the Expector soil attribute mapping method. CSIRO Land and Water: Adelaide.

De Gruijter, J.J., McBratney, A.B., 1988. A modified fuzzy k-means method for predictive classification. In: Classification and related methods of data analysis. (Editor: H. H. Block), Elsevier, Amsterdam, pp. 97–104.

Hengl, T., Rossiter, D.G., 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Science Society of America Journal 67, 1810–1822.

Irwin, B.J., Ventura, S.J., Slater, B.K., 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77, 137–154.

Kuhnert, M., Volnov, A., and Seppelt, R., 2005. Comparing raster map comparison algorithms for spatial modeling and analysis. Photogrammetric Engineering & Remote Sensing 71, 975–984.

Lagacherie, P., Legros, J.P. and Burrough, P. A., 1995. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. Geoderma 65, 283–301.

Lagacherie, P., Holmes, S., 1997. Addressing geographical data errors in a classification tree for soil unit predictions. International Journal of Geographical Information Science 11, 183–198.

Liem T.T., Wickham J., Jarnagin S.T., Knight C.G., 2005. Mapping spatial thematic accuracy with fuzzy sets. Photogrammetric Engineering & Remote Sensing 71, 29–36.

MacMillan, R.A., Moon, D.E. and Coupé, R.A., 2007. Automated Predictive Ecological Mapping in a Forest Region of B.C., Canada, 2001–2005. Geoderma 140, 353–373.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Metzler, J.W., Sadler, S.A., 2005. Agreement assessment of spatially explicit regression-derived forest cover and traditional forest industry stand type maps. Photogrammetric Engineering & Remote Sensing 71, 1303–1309.

Minasny, B., McBratney, A.B., 2007. Chapter 12. Latin hypercube sampling as a tool for digital soil mapping. In: P. Lagacherie, A.B. McBratney & M. Voltz (Eds.), Digital Soil mapping: An Introductory Perspective. Developments in Soil Science, vol. 31. Elsevier, Amsterdam, pp. 153–165.

Moon, D.E., 2005a. Quesnel PEM: Accuracy Assessment Report. Prepared by: CDT – Core Decision Technologies Inc. Prepared for: Cariboo Site Productivity Assessment Working Group (C-SPAWG), 34pp.

Moon, D.E., 2005b. Quesnel PEM: Final Report. Prepared by: CDT – Core Decision Technologies Inc. Prepared for: Cariboo Site Productivity Assessment Working Group (C-SPAWG), 27pp.

Moran, C.J., Bui, E., 2002. Spatial data mining for enhanced soil map modelling. International Journal of Geographical Information Science 16, 533–542.

Odgers, N.P., McBratney, A.B., Minasny, B., 2008. Generation of kth-order random toposequences. Computers & Geosciences 34, 479–490.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Soil pattern recognition with fuzzy c-means: Applications to classification and soil-landform inter-relationships. Soil Science Society of America Journal 56, 505–516.

Pontius, R.G., 2000. Quantification error versus locational error in comparison of categorical maps. Photogrammetric Engineering & Remote Sensing 66, 1011–1016.

Pontius, R.G., 2002. Statistical methods to partition effects of quality and location during comparison of categorical maps at multiple resolutions. Photogrammetric Engineering & Remote Sensing 68, 1041–1049.

Qi, F., Zhu, A-X., 2006. A prototype theory-based approach to predictive soil mapping under fuzzy logic. Proceedings of the 18th World Congress of Soil Science, July 12–14, 2006, Philadelphia, Pennsylvania. Abstract.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography. 27, 171–197.

Scull, P., Franklin, J., Chadwick, O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. Ecological Modelling 181, 1–15.
Shary, P.A., Sharaya, L.S., Mitusova, A.V., 2002. Fundamental quantitative methods of land surface analysis. Geoderma, 107, 171–197.
Sharaya, L.S., Shary, P.A., 2004. Elementary forms in land form classifications. Izvestiya of Samara Scienific Center of RAS, Special Isssue "Natural Heritage of Russia", Part I, 2004, pp. 102–111. (In Russian).
Shi, X., Zhu, A-Xing, Burt, J.E., Qi, F., Simonson, D., 2004. A case-based approach to fuzzy soil mapping. Soil Science Society of America Journal 68, 885–894.
Skidmore, A.K., Ryan, P.J., Dawes, W., Short, D. and O'Loughlin, E., 1991. Use of an expert system to map forest soils from a geographical information system. International Journal of Geographical Information Systems 5, 431–445.
Thomas, A.L., King, D., Dambrine, E., Couturier, A., Roque, J., 1999. Predicting soil classes with parameters derived from relief and geological materials in a sandstone region of the Vosges mountains (Northeastern France). Geoderma 90, 291–305.
Walker, R.W., 2003. Evaluating the performance of spatially explicit models. Photogrammetric Engineering & Remote Sensing 69, 1271–1278.
Wang, G., Gertner, G.Z, Fang, S., Anderson, A.B., 2005. A methodology for spatial uncertainty analysis of remote sensing and GIS products. Photogrammetric Engineering & Remote Sensing Volume 71, 1423–1432.
Zambon, M., Lawrence, R., Bunn, A., Powell, S., 2006. Effect of alternative splitting rules on image processing using classification tree analysis. Photogrammetric Engineering & Remote Sensing. 72, 25–30.
Zhou Bin, Zhang Xin-gangm Wang Ren-chao, 2005., Automated soil resources mapping based on decision tree and Bayesian predictive modeling. Journal of Zhejiang University Science 5, 782–795. URL: www.zju.edu.cn/jzus/2004/0407/040706.pdf.
Zhu, A. 1994., Soil pattern inference using GIS under fuzzy logic. Ph.D. Dissertation, Department of Geography, University of Toronto.
Zhu, A-Xing, Hudson, B., Burt, J., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge and fuzzy logic. Soil Science Society of America Journal 65, 1463–1472.

Chapter 11
Towards a Data Quality Management Framework for Digital Soil Mapping with Limited Data

B.G.C.M. Krol

Abstract The re-use of legacy soil data together with increasing numbers of environmental co-variables becomes increasingly more interesting in digital soil mapping at intermediate scales, in areas with limited data. This poses important issues regarding the reliability of these data as well as of the final product of mapping. It also requires that the data and the manner in which they are (re-)used do not have a negative influence on the quality of the mapping product. Existing quality management approaches in soil mapping emphasise the producer perspective. In addition, rather than being preventive in nature they mainly rely on detection of defects in end-product testing. A shift is required from a focus on the quality of the end-product of mapping to quality control of the mapping process itself. The development of a framework for soil data quality management is proposed in this chapter.

11.1 Introduction

Adequate information about soil resources plays an important role in support of planning and decision making about the multi-functional use of land. Large-scale soil maps (1:25 000 and larger) resulting from detailed surveys provide data up to plot level (ranging from less than a hectare in subsistence level agriculture to hundreds of hectares in commercial farming) about soil conditions that are beneficial for decision making at local level, including farm planning and advisory. Small-scale maps (typical 1:250 000) provide reconnaissance level information about broad soil classes for indicative land zoning purposes over wide areas.

Planning and decision making at district to regional level requires information at intermediate scales on soil properties and their behaviour over relatively large area extents. Increasing numbers of new user groups including non-governmental

B.G.C.M. Krol
International Institute for Geo-Information Science and Earth Observation (ITC), Postbus 6, 7500 AA, Enschede, The Netherlands
e-mail: krol@itc.nl

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

organisations, environmental modellers, ecologists, and farmers are in demand of this soil-landscape information. However, in many regions that are generally data-poor, especially in developing countries, soil and other land resource information at the intermediate level (corresponding to map scales ranging from 1:50 000 to 1:100 000) are often missing. The high costs and time-consuming nature of soil sampling make the development of methods for the creation of soil maps from sparse soil data increasingly more important (Bishop et al., 2001).

The re-use of legacy soil data together with the use of cheap and easy-to-get ancillary data (e.g. SRTM-DEM, Aster images) becomes increasingly more interesting in digital soil mapping at intermediate scales, especially in data-poor environments. Examples of ancillary data, as major types and sources of input data for digital soil mapping at multiple scales and grid resolutions, are given in Section 10.4. Rather than just being what remains behind as legacy of previous and out-dated survey practice, legacy data pose new opportunities and challenges as input to digital soil mapping. Not only do they act as provider of soil data, they also function as a source for improved soil-geomorphic understanding. A study by Bui and Moran, (2003), for example, suggests that much value can be extracted from existing and even dated soil surveys. In another example Lagacherie et al. (1995) have used existing soil survey data as reference areas as well as knowledge base in a new soil survey over a wider area.

The practical use of legacy data and covariates in digital soil mapping, however, will involve the integration of multi-source data that have been generated with different objectives, using different methodological approaches, different systems for data description and classification and made available in different analogue and digital formats and at different spatial resolutions. The integrated use of legacy soil data together with increasing numbers of environmental co-variables for digital soil mapping poses important issues regarding the reliability of these data as well as of the final product of mapping.

Another risk involves the assumption of users of digital information that what is digital is correct. Although this misconception of perfect digital information is recognised by researchers, in digital soil mapping so far there is only limited attention for uncertainty analysis or error propagation, especially where the use of secondary information is considered (Bishop et al., 2006). This makes the quality of soil data and information a clear and present concern.

In this chapter the development of a data-quality management framework for digital soil mapping with limited data is proposed, as a basis for research work in progress. Three interrelated issues are emphasised:

1. An extension of focus on intrinsic data quality to other soil data quality dimensions;
2. A shift from end-product testing to mapping process control;
3. A change from a mainly producer-oriented focus to a user-oriented focus.

Thus, the proposed data-quality management framework intends to provide a mechanism for users of soil and ancillary environmental data to make decisions about data quality, and for producers of soil information to control the soil-mapping process.

11.2 Focus on Quality in Soil Mapping

In the context of soil survey there is a long-standing tradition of focus on quality, albeit often qualitative and mainly producer-oriented. In the pre-digital era published soil map sheets often came with an indication of quality, for example in the form of a reliability diagram. Considering the many possible questions a user may have concerning the quality of a soil survey Forbes et al. (1982) developed guidelines for the evaluation of the adequacy of soil-resource inventories. A proposal for further refinement of these guidelines is presented by Hengl and Husnjak (2006), as part of their evaluation of the adequacy and usability of soil maps in Croatia.

Modern predictive, digital soil maps are expected to be of more quality than conventional soil maps: they are made using data that are better defined and documented, in terms of their positional quality, currency and lineage (Finke 2007). Soil property maps that are produced by using geo-statistical approaches are typically accompanied by an 'error map' as by-product. It shows the prediction variance of the estimated soil property values and thus provides a way of expressing the uncertainty of the estimation (see also Davis, 2002). In other words, the error map gives an impression of the reliability of the map of estimated soil property values (Webster and Oliver, 2001).

At the same time, however, potential users of digital soil map products express concerns about the quality of DSM products as compared to maps produced using traditional mapping approaches (see Section 10.7).

11.2.1 End-Product Testing

Existing quality control approaches applied in mapping mainly rely on inspection to detect defects, whereas the goal should be to prevent them in the first place (Schmidley, 1997). Validation studies to assess model performance in soil mapping are typically carried out using independent validation sample sets. However, it is argued that the earlier mentioned data scarcity issue also holds for validation studies. Even in the overall more data-rich pedometric soil modelling environment model validation is controlled by the non-exhaustive availability of soil observation data, as is also recognised by Grunwald and co-workers (see Grunwald et al., 2005). Furthermore, it may be questioned whether to an end-user of soil geographical information it makes sense, or even is acceptable, to run an independent validation study to find out that the product of soil mapping appears to be uncertain or even erroneous. It is even argued by Finke (2007) that the presence of precision measures (i.e. error maps, see above) that accompany many modern soil property maps may lead users to conclude that these maps are of lesser quality than conventional maps. Moreover, few end-users may be able or even bother to consider the error map in evaluating the usefulness of a soil property map for their intended purpose. A problem with end-product testing in general is that validation samples, are often too small and come too late, to allow for any corrective measure (see also Godfrey and Howard, 2004).

11.2.2 Producers and Users

In soil mapping, as in mapping in general, data quality is in practice still mainly perceived from the producer perspective. The fitness for use of soil mapping products, however, is of prime interest to the user. The need for improved user orientation in dealing with data quality is also emphasised in Section 1.5.

Producers and users of data often have a different perception about data quality. Data producers typically identify a product of quality as conforming to specifications, whereas data users recognise quality if a product meets or even exceeds their expectations (Devillers et al., 2002; Kahn et al., 2002). The producer perspective in dealing with quality problems in soil data and geographical information is also illustrated by a still dominant focus on intrinsic data quality elements, such as accuracy and uncertainty (see for example: Bishop et al., 2006; McBratney et al., 2003; Zhang and Goodchild, 2002).

However, data quality problems go beyond the data values and their accuracy and also include other elements, such as accessibility and completeness (Dalcin, 2004; Wang and Strong, 1996). In addition, eventually it is the data user that must find the data accurate, for example in being correct, objective and from a reputable source (Wang and Strong, 1996). At the same time the role of users in dealing with aspects of data quality is getting more and more attention in soil mapping. Here Finke (2007) even considers the usage aspect of quality as a quality variable, together with other aspects that define data quality.

It is also recognised that there is a need for mechanisms that help users of soil information to make decisions about uncertainty (Lagacherie and McBratney, 2007), and data quality for that matter. Such mechanisms are also needed to guard against inappropriate use by inexperienced users. The gradual shift from qualitative to quantitative soil-landscape modelling that is taking place (see also Grunwald, 2006), combined with the increasing availability of automated tools for quantified soil modelling, such as offered by geographic information technology, involves that often 'inexperienced users are able to perform complex analyses without adequately considering issues of data quality' (Bishop and Minasny, 2006).

11.3 Data Quality, Definitions and Dimensions

The ISO standard for geographic information (ISO, 2002) defines data quality as the 'totality of characteristics of a product that bear on its ability to satisfy stated and implied needs'. Geographical data quality is also defined as the 'ability of data to adequately fulfil the purpose for which it will be applied', as its 'fitness for use' (Hunter et al. 2003), or as the 'fitness of data to a given purpose' (Chrisman, 1984).

What is clear from these and other definitions is the notion that data quality cannot be restricted to data values, it includes a data-use component as well. In fact, data quality can only be assessed in relation to the people who use the data (Strong et al., 1997). In a soil database, for example, data as such have no identified quality. Quality only exists when the data are being used by somebody (see also Dalcin, 2004). A high-resolution DEM available in a database, for example, may accurately represent elevation in a given geographical region. But this same DEM has only limited quality, if it just partially covers the project area of a particular soil mapping project. Schmidley (1997) even goes as far as to define quality as ‘customer satisfaction’. Since it emphasises the importance of taking a consumer viewpoint on quality, and because it is the user who ultimately judges whether or not a product is fit for use, the ‘fitness for use’ concept is now widely accepted and adopted (Wang and Strong, 1996).

Data quality is generally recognised as a multidimensional concept (Chapman, 2005). A *data quality dimension* (or characteristic) can be understood as a set of data quality attributes that capture a particular aspect of data quality (Wang and Strong, 1996).

In the geographical work field the following data quality dimensions are usually emphasised: positional accuracy, attribute accuracy, logical consistency, completeness, temporal accuracy, and lineage. According to Hunter et al. (2003) these are the elements of data quality that typically need to be reported to users of geographical data. An alternative grouping identifies five so called ‘PARCC’ attributes of (geographical) data quality as follows (Jones, 1999; EPA, 2003): namely, Precision, Accuracy, Representativeness, Completeness and Comparability.

The above mentioned data quality dimensions and PARCC attributes conform to the ISO standard for geographical information, ISO19113 (ISO, 2002). According to this standard the following two components must be used to describe the quality of a dataset: (1) data quality (sub-)elements, and (2) data quality overview elements.

Table 11.1 ISO data quality elements and data quality subelements (ISO, 2002)

Data quality element*(applicable data quality sub-elements)**	Description
Completeness (*commission, omission)*	Presence and absence of features, their attributes and relationships
Logical consistency (*conceptual consistency, domain consistency, format consistency, topological consistency)*	Degree of adherence to logical rules of data structure (conceptual, logical, physical), attribution and relationships
Positional accuracy (*absolute accuracy*, r*elative accuracy, gridded data position accuracy*)	Accuracy of the position of features
Temporal accuracy (*accuracy of a time measurement, temporal consistency, temporal validity*)	Accuracy of the temporal attributes and temporal relationships of features
Thematic accuracy (*classification correctness, non-quantitative attribute correctness, quantitative attribute accuracy*)	Accuracy of quantitative attributes and the correctness of non-quantitative attributes of the classifications of features and their relationships

* additional data quality elements may be created.
** additional data quality sub-elements may be created for any of the data quality elements.

Table 11.2 Data quality categories and dimensions (after Strong et al., 1997; Kahn et al., 2002)

Data Quality (DQ) category	Data quality dimension	Short description
Intrinsic	Accuracy	Data are correct and reliable
	Objectivity	Data are based on facts, impartial
	Believability	Data considered true and credible, trustworthy
	Reputation	Data come from a good source
Accessibility	Accessibility	Data easily and quickly obtainable when needed
	Access security	Data protected against unauthorized access
Contextual	Relevance	Data are useful, appropriate, applicable
	Value-added	Data are beneficial, provide advantages
	Timeliness	Data are sufficiently up-to-date
	Completeness	Data include all necessary values
	Amount of data	Data are of sufficient volume
Representational	Interpretability	Easy to interpret what the data mean
	Ease of understanding	Data is easy to comprehend
	Concise representation	Data are formatted in a compact and precise way
	Consistent representation	Data are represented in the same, consistent format

ISO data quality elements and related data quality sub-elements are used to describe, where applicable, 'how well a dataset meets the criteria set forth in its product specification' and thus provide quantitative quality information, see Table 11.1. ISO data quality overview elements, including *purpose*, *usage* and *lineage*, are used for a non-quantitative, general description of a dataset.

In an enterprise context Wang and Strong (1996) and Strong et al. (1997) tried to capture quality aspects that users of data and information consider as important. They introduced a data quality framework to assess the quality of organisational data. This framework identifies fifteen quality dimensions organised in four categories (Table 11.2). According to Wand and Wang (1996) the contextual and representational data quality categories specifically concern the usefulness of the data in an information system within an organisational setting. The intrinsic and accessibility data quality categories, on the other hand, are recognised as use-independent and related to the design and implementation of an information system.

11.4 Data Quality Dimensions in Soil Mapping Context

Where accuracy and objectivity are well known data quality dimensions for information system professionals (Wang and Strong, 1996), this also holds for the producers of soil geographical information, and of digital soil maps for that matter. The aforementioned positional accuracy, attribute accuracy and temporal accuracy emphasise that soil geographical data has quality in its own right, that it has an intrinsic value.

To users of soil data and soil map products, however, these are not sufficient to consider soil data and soil map information as being of good quality. They also view reputation and believability as important data quality dimensions. An example of the use of the reputation dimension is provided by Rossiter (Chapter 6) in his discussion about soil data renewal in a soil mapping example in Kenya: the composition of the soil survey team and the reputation of its members are detailed to infer the presumed quality of the survey data. The believability dimension is considered by users for data quality assessment if, for example, nothing is documented concerning the laboratory analysis protocol considered for the analysis and subsequent presentation of soil physical and soil chemical data. In the example of legacy soil observation data shown in Fig. 11.1 it is impossible for the user to find out how trustworthy the presented soil analytical information is: which data have been analysed but are not presented in the table? Which data have not been analysed at all? How credible are the data that are presented?

Contextual soil data quality dimensions emphasise that the quality of (legacy) soil data and environmental covariates has to be considered in the context of a task at hand. This can be, for example, their use in a particular case of soil mapping (e.g. in terms of a mapping area and/or targeted soil variables), or their application in a particular mapping approach (e.g. by environmental correlation, geo-statistical). An example: an environmental modeller with specific interest in modelling the spatial variation of organic matter content in the topsoil may well consider the corresponding 'OM' value in Fig. 11.1 as relevant, given that the database values are of quality in terms of their believability.

The data in a soil information system must be stored in a secure and accessible way. Accessibility soil data quality dimensions refer to the ease of user-access to digital soil and related environmental data and to digital soil mapping products, for example while querying a soil geographical database. In addition, they also refer to the secure storage and management of these data in a soil information system.

The representational soil-data quality dimensions refer to the soil information system that must organise the data both in a concise and consistent manner, and in such a way that these data are interpretable and easy to understand. For example, two multiple-source soil maps each representing part of a user's project area can only be successfully integrated into a new comprehensive area-wide soil map if they are both formatted in a similar, consistent and concise manner.

Sample no.	*Depth cm*	*Horizon*	*Texture*	*OM %*	*$CaCO_3$ %*
84a	0-15	Ap1	C	2.26	16.2
84b	15-40	Ap2	C	1.84	16.0
84c	40-67	Bk1	C	--	n.a.
84d	67-102	Bk2	C		20.6
84e	102-140	Bk3	SiC		25.9
84f	140+	BCk	C		

Fig. 11.1 Believability as soil data quality dimension: fragment of a soil analytical database table without any additional information available about laboratory analysis protocol

The data quality dimensions presented in Table 11.2 refer more to organisational data, to handle data quality in business processes. It is not necessarily complete where geographical data, and soil data more in particular, are concerned. Data lineage, for example, is not considered. It may well be that a further specification of data quality dimensions for digital soil mapping will result in an extension of those listed in Table 11.2. It remains to be identified which data quality dimensions are important in digital soil mapping, which ones have a more generic relevance, and which ones are relevant in a particular case only.

11.5 Managing Data Quality

The terms *quality management*, *quality assurance* and *quality control* are often loosely used to refer to activities where quality is considered in one way or the other. In particular, the difference between 'quality control' and 'quality assurance' does not seem clear in all cases, and the terms are often synonymously applied to describe the data quality management practice (Chapman, 2005). In quality research and practice, however, these are terms with an explicit and distinguishable meaning. And as such each of them plays a specific role in the dealing with the quality of goods, data, and services in order to guarantee a good quality and to prevent a defective quality as much as possible. In this sense quality control and quality assurance are crucial elements of any data quality management mechanism.

The ISO-9000 standard defines quality control as 'operational techniques and activities that are used to fulfil requirements for quality' (ISO, 2000). In a more comprehensive manner, EPA (2001) defines quality control as the 'overall system of technical activities that measures the attributes of performance of a process, item, or service against defined standards to verify that they meet the stated performance criteria established by the customer, operational techniques, and activities that are used to fulfil performance criteria for quality'. Data quality control is set up and carried out by the producer (or provider) of data (or services). This same data quality control is subject to quality assurance by another, external entity. In more practical terms, data quality control can be seen as a structured set up of preventive and corrective actions as integral part of, for example, a soil mapping process. It has to provide confidence that an information product or service will fulfil use requirements and is produced according to defined standards. This makes data quality control a continuous process rather than a one-time activity.

11.6 A Data Quality Management Framework for Soil Mapping

While considering its distinctive nature also soil geographical information can be treated as a product that moves through an information production chain. Wang (1998) recognises an analogy between quality issues in product manufacturing and information production. This analogy can also be extended to the generation of soil

Table 11.3 Analogy between product manufacturing and digital soil mapping (adapted from Wang, 1998)

	Product manufacturing	Digital soil mapping
Input	Raw materials/ingredients	Soil and ancillary data
Process	Assembly line	Soil mapping approach
Output	Physical product	Soil information product

information in soil information systems. Where product manufacturing is considered as a processing system that converts raw materials or ingredients into physical products (e.g. bread, packaged beer, bicycles), soil information production – as in digital soil mapping – can be seen as a processing system acting on input data to generate soil information products, as is shown in Table 11.3.

Starting from the analogy between manufactured goods and information products Wang (1998) has extended Total Quality Management principles to a Total Data Quality Management (TDQM) methodology. At a general level the TDQM cycle for continuous data quality analysis and improvement recognises four elements necessary for quality management of information products:

1. the *definition* of information requirements and of important data quality dimensions associated with a targeted information product;
2. the *measurement* of identified quality dimensions, which involves the production of quality metrics;
3. the *analysis* of root causes for data quality problems and the calculation of the impact of poor data quality on the information product;
4. the *improvement* of data quality in identified key areas for improvement.

Together with the application of identified data quality standards, of Good Manufacturing Practices (GMP) and of Standard Operation Procedures (SOP), elements of TDQM are now increasingly considered for the management of data, for example in organisational information systems (as in health services) and (spatial) data warehouses.

Application of the TDQM methodology in a digital soil mapping context involves that techniques for improving soil data quality are applied on identified soil data quality dimensions according to information requirements that have been specified by the soil information user. The definition, measurement, analysis and improvement of soil data quality in an iterative way is essential to ensure that soil map information of the highest possible quality is generated. A schematic presentation is given in Fig. 11.2.

Application of this TDQM methodology requires that first the characteristics (functionalities) for the targeted soil information product ('what must the soil map do for me') are defined, the soil information quality requirements are assessed, and the appropriate soil mapping approach is identified. Next, measurement, analysis, and improvement are considered. Thus, it may for example appear that an entity-type of soil map, as it was typically developed in the past, may not fit the needs for

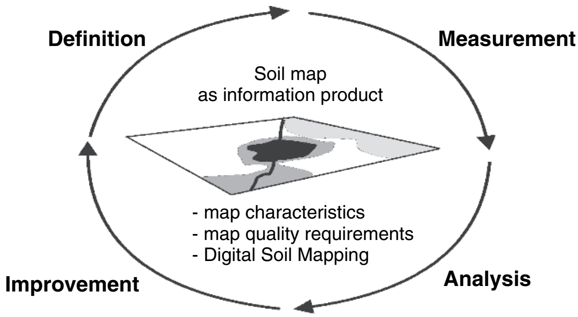


Fig. 11.2 The TDQM methodology applied for digital soil mapping

soil organic carbon information of environmental modellers today. In such case, this should become clear in the definition phase where this same soil information user group would be involved. In case they are not involved it is the soil mapper's responsibility that their needs are properly met in a later phase, otherwise this conventional soil map will not be 'fit for use' to the environmental modellers.

11.7 Conclusions and Perspectives

There is a structural lack of relevant soil information products at intermediate scales over wide areas. This is especially the case in regions in the world that appear as generally data scarce. To bridge this information gap methods for digital soil mapping are required that are cheap and only require limited data input.

Although legacy soil data and related environmental data, together with cheap and easy-to-get ancillary data, provide a very useful data source for digital soil mapping they also pose data quality problems.

As is indicated in this chapter, soil data quality problems are not restricted to uncertainty issues, they also include aspects like completeness and accessibility of data. It is recognised that soil and environmental data for digital soil mapping has value in its own right, has intrinsic data quality. In addition, digital soil mapping data must be considered in the context of the intended uses of the soil map product, highlighted by their contextual data quality. The organisation of soil and other environmental data in Soil Information Systems (Lagacherie and McBratney, 2007) can facilitate the convenience and ease of use of data, its accessibility and representational data quality (see also Kerr, 2006), for digital soil mapping.

To improve data quality in soil mapping, focus is required on quality aspects that are important to users of data. They are both soil specialists that apply existing, multi-source and multi-theme data in soil mapping, and end-users, consumers of soil information.

Existing quality management approaches in soil mapping so far merely have a producer-oriented focus and also mainly rely on inspection of end products and detection of defects. There is a need for a more user-oriented quality focus that aims at the prevention of errors. This calls for a systematic approach to data quality management of the soil mapping process itself.

It is argued that elements of TDQM can assist in the development of a soil data quality management framework that focuses on the prevention of defective mapping products as well as on increased user involvement in the generation of soil information. Issues that need further investigation in this context include:

- The definition of a soil data quality space: which data quality dimensions are important in digital soil mapping; which ones have a more generic relevance, which ones are relevant in a particular case only.
- The development of quality metrics for the measurement of identified soil data quality dimensions. Some of them, like 'accuracy', will be more easy to measure, whilst for others, like 'ease of understanding', it will be more difficult to develop quality metrics.
- The continued development of approaches, techniques and tools for the improvement of the quality of soil and related environmental data; this is specifically relevant in cases where the re-use of legacy data is considered, and where the spatial and semantic integration of multi-source datasets is emphasised.

It is beyond doubt that good use can be made here of the standing experience in soil mapping, such as in the application of (geo-)statistical approaches in dealing with spatial uncertainty. But also other multidisciplinary techniques and tools can be instrumental as part of a soil data quality management framework, for example elements of data mining (see for example Moran and Bui, 2002) and ontology-based semantic matching (Krol et al., 2007) that have been introduced in digital soil mapping.

The digital soil mapping community cannot leave problems with soil information products for the users to be recognised and resolved. Any team involved in digital soil mapping should pro-actively and continuously improve the quality of the soil information product. Since users are more likely to encounter problems (particularly concerning contextual data quality) with the soil information they use, as producers and/or suppliers of soil information we, therefore, need to continuously expand our knowledge about how and why our products of digital soil mapping are used.

References

Bishop, T.F.A., McBratney, A.B., Whelan, B.M., 2001. Measuring the quality of digital soil maps using information criteria. Geoderma 103, 95–111.

Bishop, T.F.A., Minasny, B., 2006. Digital soil-terrain modelling: the predictive potential and uncertainty. In: S. Grunwald (Ed.), Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Bishop, T.F.A., Minasny, B., McBratney, A.B., 2006. Uncertainty analysis for soil-terrain models. International Journal of Geographical Information Sciences 20, 117–134.

Bui, E.N., Moran, C.J., 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.

Chapman, A.D., 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Global Biodiversity Information Facility, Copenhagen.

Chrisman, N.R., 1984. The role of quality information in the long-term functioning of a geographic information system. Cartographica, 21, 79–87.

Dalcin, E.C., 2004. Data Quality Concepts and Techniques Applied to Taxonomic Databases. Ph.D. Thesis, University of Southampton, Southampton.

Davis, J.C., 2002. Statistics and data analysis in geology, third edition. John Wiley & Sons Inc., New York.

Devillers, R., Gervais, M., Bedard, Y., Jeansoulin, R., 2002. Spatial data quality: from metadata to quality indicators and contextual end-user manual, OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management, 21–22 March 2002, Istanbul.

EPA, 2001. EPA Requirements for quality assurance project plans, United States Environmental Protection Agency, Washington, DC.

EPA, 2003. Guidance for geospatial data quality assurance project plans, United States Environmental Protection Agency, Washington, DC.

Finke, P.A., 2007. Quality assessment of digital soil maps: producers and users perspectives. In: P. Lagacherie, A.B. McBratney, M. Voltz, (Eds.), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science, Vol. 31. Elsevier, Amsterdam.

Forbes, T., Rossiter, D., Van Wambeke, A., 1982. Guidelines for evaluating the adequacy of soil resource inventories, 1987th edition. SMSS Technical Monograph No. 4. Cornell University, Department of Agronomy, Ithaca, NY.

Godfrey, S., Howard, G., 2004. Water Safety Plans (WSP) for urban piped water supplies in developing countries. Water, Engineering and Development Centre, Loughborough University, Leicesterhsire, UK.

Grunwald, S., Bruland, G.L., Goovaerts, P., Bloom, S.A., 2005. Independent validation of soil predictions – the act of testing the truth. Abstract book, Pedometrics 2005 meeting, 12–14 September 2005, Naples, FL., USA.

Grunwald, S., 2006. What do we really know about the space-time continuum of soil-landscapes. In: S. Grunwald (Ed.), Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics. CRC Press, New York.

Hengl, T., Husnjak, S., 2006. Evaluating adequacy and usability of soil maps in Croatia. Soil Science Society of America Journal 70, 920–929.

Hunter, G.J., Jones, S.D., Bregt, A., Masters, E.G., 2003. Spatial data quality. In: C.M. Bauzer Medeiros (Ed.), Advanced Geographic Information Systems. Encyclopedia of Life Support Systems (EOLSS), Oxford, UK.

ISO, 2000. Quality Management Systems – Fundamentals and Vocabulary (ISO 9000:2000), International Organization for Standardization, Geneva, Switzerland.

ISO, 2002. Geographic Information – Quality Principles; first edition (ISO 19113:2002), International Organization for Standardization, Geneva, Switzerland.

Jones, B.E., 1999. Principles and practices for quality assurance and quality control. US Geological Service Open File Report 98–636. http//ma.water.usgs.gov/fhwa/products/ofr98-636.pdf.

Kahn, B.K., Strong, D.M., Wang, R.Y., 2002. Information quality benchmarks: product and service performance. Communications of the ACM 45, 184–192.

Kerr, K., 2006. The institutionalisation of data quality in the New Zealand health sector. PhD thesis. University of Auckland, New Zealand.

Krol, B., Rossiter, D.G., Siderius, W., 2007. Ontology-based multi-source data integration for digital soil mapping. In: P. Lagacherie, A.B. McBratney, M. Voltz (Eds.), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science, Vol. 31. Elsevier, Amsterdam, 119–133.

Lagacherie, P., Legros, J.P., Burrough, P.A., 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., McBratney, A.B., 2007. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: P. Lagacherie, A.B. McBratney, M. Voltz (Eds.), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science, Vol. 31. Elsevier, Amsterdam.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
Moran, C.J., Bui, E.N., 2002. Spatial data mining for enhanced soil map modelling. International Journal of Geographical Information Science 16, 533–550.
Schmidley, R.W., 1997. Quality control in mapping: some fundamental concepts. Surveying and Land Information Systems 57, 31–36.
Strong, D.M., Lee, Y.W., Wang, R.Y., 1997. Data quality in context. Communications of the ACM 40, 103–110.
Wand, Y., Wang, R.Y., 1996. Anchoring data quality dimensions in ontological foundations. Communications of the ACM 39, 86–95.
Wang, R.Y., Strong, D.M., 1996. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems 12(4), 5–34.
Wang, R.Y., 1998. A product perspective on total data quality management. Communications of the ACM 41, 58–65.
Webster, R., Oliver, M., 2001. Geostatistics for Environmental Scientists. John Wiley & Sons Ltd, Chichester.
Zhang, J., Goodchild, M., 2002. Uncertainty in Geographical Information. Research Monographs in Geographic Information Systems. Taylor and Francis, London.

Chapter 12
Demand-Driven Land Evaluation

I.L.Z. Bacic

Abstract Land evaluation is the prediction of land performance over time under specific uses. These predictions are then used to guide strategic land-use decisions. Modern land evaluation has a 30-year history, yet the results are generally accepted to be disappointing. Land users and planners are inclined to ignore land evaluators, reflecting the poor quality and low relevance of many actual land evaluations, as well as poor communication with users. The main objective of this research was to improve use and usefulness of information for rural land use decisions based on an operational demand-driven approach for land evaluation with case studies in Santa Catarina State, Brazil. First, the use and usefulness of soil surveys and land evaluation reports to land use planners were described and quantified and the relation between latent demand and actual supply was observed. Then, the farmers' decision environment and its implications for land evaluation were studied. These were the basis for the subsequent steps of this research. Next, the applicability of a data-intensive distributed environmental model (AgNPS) in a relatively data-poor environment was evaluated. This model and other tools for visualization of scenarios were used with community participation, to test their effects on collective understanding of shared environmental problems. Finally, the potential of a participatory approach for integrating risk analysis into decision making for rural land use was evaluated. This research showed that a demand-driven approach to make the information more relevant and useful to rural decision makers for land use planning is possible in practice and should be further explored, but its effectiveness needs time to be confirmed. Applying the proposed approach, new demands were raised and considering that the number of soil scientists and financial resources are scarce in the region, digital soil mapping based on existing data emerges as a potential alternative to help to answer to the increasing rural decision makers demands.

I.L.Z. Bacic
Epagri/Ciram – Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S.A., Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina, Rodovia Admar Gonzaga 1347, C.P. 502. 88034-901, Florianópolis, Santa Catarina, Brazil
e-mail: bacic@epagri.sc.gov.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

12.1 Introduction

Land evaluation is the process of predicting land performance over time according to specific types of uses (Van Diepen et al., 1991; Rossiter, 1996). These predictions are intended to guide strategic land use decisions. So, one would expect that land use planners and other decision-makers who influence rural land use would be eager to use the results of land evaluation. Unfortunately this is often not happening in practice. One of the main motivations for this study was the author's disappointment resulting from more than 15 years working as a soil surveyor and land evaluator with the feeling that the work was not useful to and used by the potential clients. This apparent irrelevance of land evaluation is also observed internationally. Several authors have stated that decision makers do not in general make use of these results, nor are they particularly satisfied with them, if indeed they know of their existence (Rossiter, 1996; Bouma, 1997, 1999). To date, more attention has been paid to land evaluation methods themselves than to their relevance and the utilization of the information that they generate.

It has been suggested that to change the existing situation, adapted land use options and planning strategies should be formulated with the participation of the stakeholders and in accordance with their possibilities (Bouma, 1999), i.e. a participatory and demand-driven approach. It is crucial to know what are the problems, the needs and possibilities of the stakeholders before starting the land evaluation process, otherwise there is a risk that questions may be answered that have no priority or relevance, and/or that questions may not be answered properly according to the community expectation. The shift towards more participatory research is, however, not only inspired by pragmatic reasoning. The modern farmer, especially in developed countries, as well as the land use planner, is a well-trained professional who is less interested in receiving "definite" answers to questions than in having a presentation of a series of realistic options with accurate predictions from which he or she can make a selection. Modern agronomic and soil research has a clear challenge in developing such options in close consultation and interaction with the stakeholders, be it farmers, planners or politicians themselves. That is also the case in Santa Catarina state, Brazil. One of the reasons that the planners do not use current land evaluations is that these do not present a range of realistic, relevant alternatives. Usually the land evaluation procedures show what is wrong in the land use, what and where the conflicts are, but do not give realistic options from which the stakeholders can choose. In particular, these options should include uncertainties about each land use alternative and the risks to change the current activity, since decision makers always must decide on the basis of uncertain and incomplete information.

Information technology continues to improve rapidly, in particular GIS, remote sensing, expert systems, as well as geostatistical models, digital soil mapping and pedometric techniques (Cook et al., 2006; Lagacherie, 2006; Chapter 33). In the context of participatory land evaluation, these should be used as much as possible during the whole process, taking into account local conditions, e.g. the readiness of decision makers to interact with information technology. Bouma (1999) points out that modern information technology has an important role to play in stimulating interaction with decision makers. Visualisation of alternative land use patterns associated with different options is a very powerful tool to involve them in the land use planning process. Interactive computer technology allows, for instance, joint generation of alternative land use scenarios with all associated input data by researchers and decision makers.

The general objective of this research was to improve use and usefulness of information for rural land use decisions based on an operational demand-driven approach for land evaluation with case studies in Santa Catarina State, Brazil. The work is a collection of case studies in Santa Catarina, Brazil, related to demand-driven land evaluation. For more details, see Bacic (2003), Bacic et al. (2003, 2006a,b).

12.2 Material and Methods

A diagram adapted from Rossiter, unpublished (Fig. 12.1) illustrates the emerging demand-driven paradigm in land evaluation used in this work.

The first step of the research was to evaluate the use and usefulness of soil surveys and land evaluation reports to land use planners. To test the success of a large land evaluation exercise undertaken as part of micro-catchment project in Santa Catarina State, southern Brazil, agricultural extensionists, considered as the primary land evaluation clients, were queried. A questionnaire was used with both structured and open questions, to determine their experiences with, and attitudes to, the current land evaluation method. The relation between latent demand and actual supply was observed.

Next, the farmers' decision environment and its implications for land evaluation were studied. To understand the environment for agricultural land use decisions and review its implications for land evaluation, literature, including local documentation, and semi-structured interviews with farmers and extensionists, were used.

One of the types of information demanded by the decision makers was the environmental degradation risks assessment. Therefore, the applicability of a data-intensive distributed environmental model – AgNPS (Young et al., 1987, 1989) – in a relatively data-poor environment was evaluated. This included data preparation, cell size selection, sensitivity analysis, model calibration and application to different management scenarios. The model was calibrated by making a "best guess" for model parameters and a pragmatic sensitivity analysis was performed. The parameters were adjusted so that the model outputs (runoff volume, peak runoff rate and sediment concentration) closely matched observed data. This model was used along with other tools (synoptic satellite image, ortophoto mosaic, location of pig producers) for visualization of scenarios in a participatory community workshop, designed to test their effects on collective understanding of shared environmental problems. Workshops were organized with extensionists and farmers directly involved in the land uses that were thought to be related to the perceived environmental problems. Questionnaires were administered at four different times during

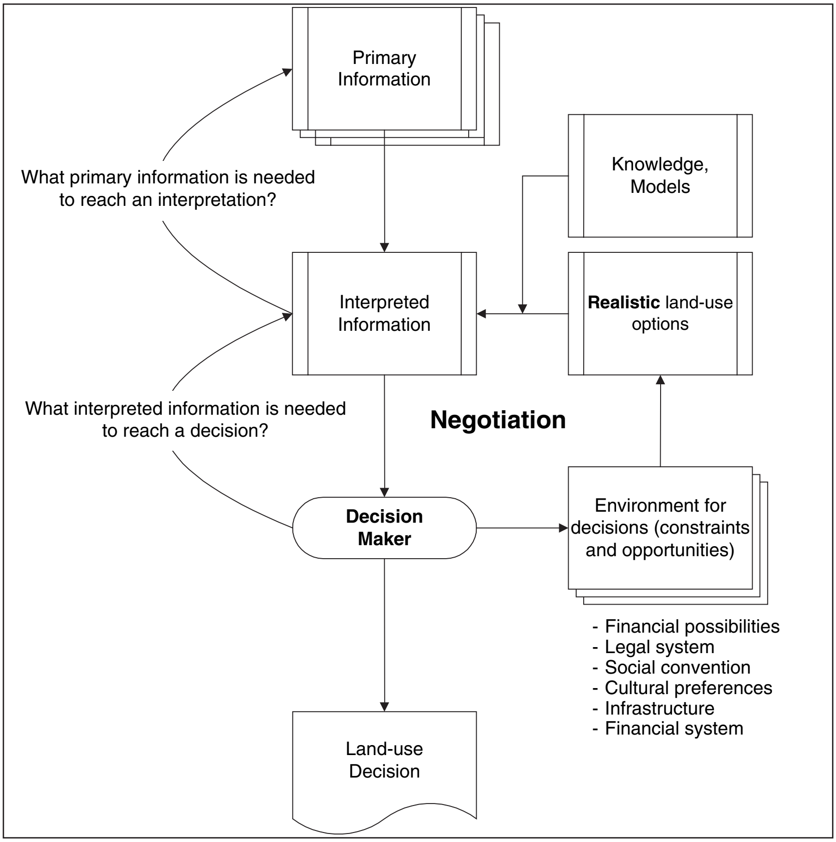


Fig. 12.1 Demand-driven land evaluation and resource inventory (adapted from Rossiter, unpublished)

the workshops, to test participants' reactions to, and opinions of, the information provided.

Finally, formal risk concepts were introduced at decision makers meetings according to local demands and following a participatory approach, as a first step towards integrating risk assessment into rural decision making in Santa Catarina, Brazil. Semi-structured interviews and meetings were conducted with extensionists and farmers. The following information was presented and discussed: (1) the time series frequency distribution of maize yield predictions, simulated by GAPS computer program (Buttler et al., 1997) for 16 feasible planting dates, representing climatic risks, both within and between years; and (2) a simple economic analysis (gross margin) and income probabilities for seven land use options over a recent five-year period, followed by an interactive exercise where probabilities of achieving user-supplied target gross margins were calculated according to participants actual information, using the @RISK computer program (Palisade Corporation, 1998).

Decision-makers' attitudes towards risk, and how these were influenced by objective information, were also investigated.

12.3 Results and Discussion

12.3.1 The Use of Land Evaluation Information by Land Use Planners and Decision-Makers

The soil resource inventory and associated land evaluation were considered to have some utility, but were not in general used for their intended purpose, namely farm planning. This was mainly because they did not contain crucial information necessary to such planning in the actual context in which the farmer had to take decisions. The additional required information, reflecting the primary deficiencies on the reports, is shown in Table 12.1. These deficiencies could have been avoided with a demand-driven approach, evaluating and reporting according to the true needs and opportunities of the decision-makers.

12.3.2 The Environment for Farmers' Land Use Decisions

Different groups of farmers expressed different needs for information and thus should be approached in different manner. Some farmers would welcome any information on improving their current farming systems, whilst others are also interested in innovative crops or agricultural processes. Yet another group seemed to need motivation more than information. The study suggests that if the land evaluation process is begun with a careful analysis of the decision environment of rural land users (farmers) and follows a demand-driven approach, the results will likely be more realistic and therefore more useful to both policy/planning institutions and direct land users. This should lead to more demand and a "virtuous cycle" where planning, land evaluation and clients' needs and possibilities are increasingly interlinked.

12.3.3 Applicability of a Distributed Watershed Pollution Model in a Data-Poor Environment

The work showed one example of the applicability of a data-intensive water quality model (AgNPS) in a relatively data-poor environment, demonstrating that it is possible to consistently apply such model even without expensive procedures for data measurement and collection, at least for relative risk assessment. In this case it was shown that expert knowledge of the area in addition to local knowledge and literature information compensates in part for poor data. It was possible to apply a distributed environmental model like AgNPS, for relative ranking of environmental

Table 12.1 Summarized from the rural extensionists responses to the questionnaires

Land inventory usefulness						
Very useful/Useful (%)			Slightly useful/Useless (%)			No answer (%)
59			32			9
How the planners are using the land inventories						
Displaying maps (%)	Meeting with farmers (%)		Planning of land use and management (%)		Other (%)	Not using (%)
36	30		43		25	17
Additional information required						
Socio-economic analysis (%)	More land use alternatives (%)	Uncertainities and risks assessment for alternatives (%)	Environmental degradation risks assessment (%)	Other (%)	None (%)	No answer (%)
42	21	34	59	8	11	8

management scenarios in a comparative way (Fig. 12.2). This can be useful in many areas of the world where data and financial or human resources for detailed model calibration are lacking. This work also demonstrated that supposedly unprepared decision makers were able to properly understand and react to new tools, even though it was the first attempt to introduce these in the region. This shows the potential in the region for the use of GIS, expert systems and digital soil mapping techniques.

12.3.4 Using Spatial Information to Improve Collective Understanding of Shared Environmental Problems at Watershed Level

Visualization of scenarios (e.g. Fig. 12.2) with community participation was useful to increase participants' understanding of the water pollution problem, improve their perceptions (Table 12.2), stimulate the search for solutions (Table 12.3) and generate new demands. This was the case even taking into account that rural decision makers are not well educated and not used to visualizing scenarios. In this, Santa Catarina is similar to many areas of the world. Participants, in general, liked the material presented and the methods of the meetings.

12.3.5 A Participatory Approach for Integrating Risk Assessment into Rural Decision Making

The case study particularly focused on two of the main risk-oriented information demands in the region: (1) yield predictions for maize on different planting dates and (2) economic information for different land use options (Table 12.4). It evaluated the potential of a participatory approach for integrating risk analysis into decision making for rural land use and decision makers' view of the supplied information (e.g. Table 12.5). It also investigated decision makers' attitudes towards risk, and the degree to which these could be changed by objective information. Different groups had markedly different levels of knowledge, analytic capacity, economic conditions, perspectives and needs, and therefore should be approached differently and with group-specific information. Farmers were mostly moderately or extremely risk averse. However, at the end they declared themselves willing to take risks if they have adequate information. The results suggest that a participatory approach, by gathering, presenting and periodically discussing demanded information with decision makers is certainly a practice to be further explored to effectively integrate risk assessment into rural decision making.

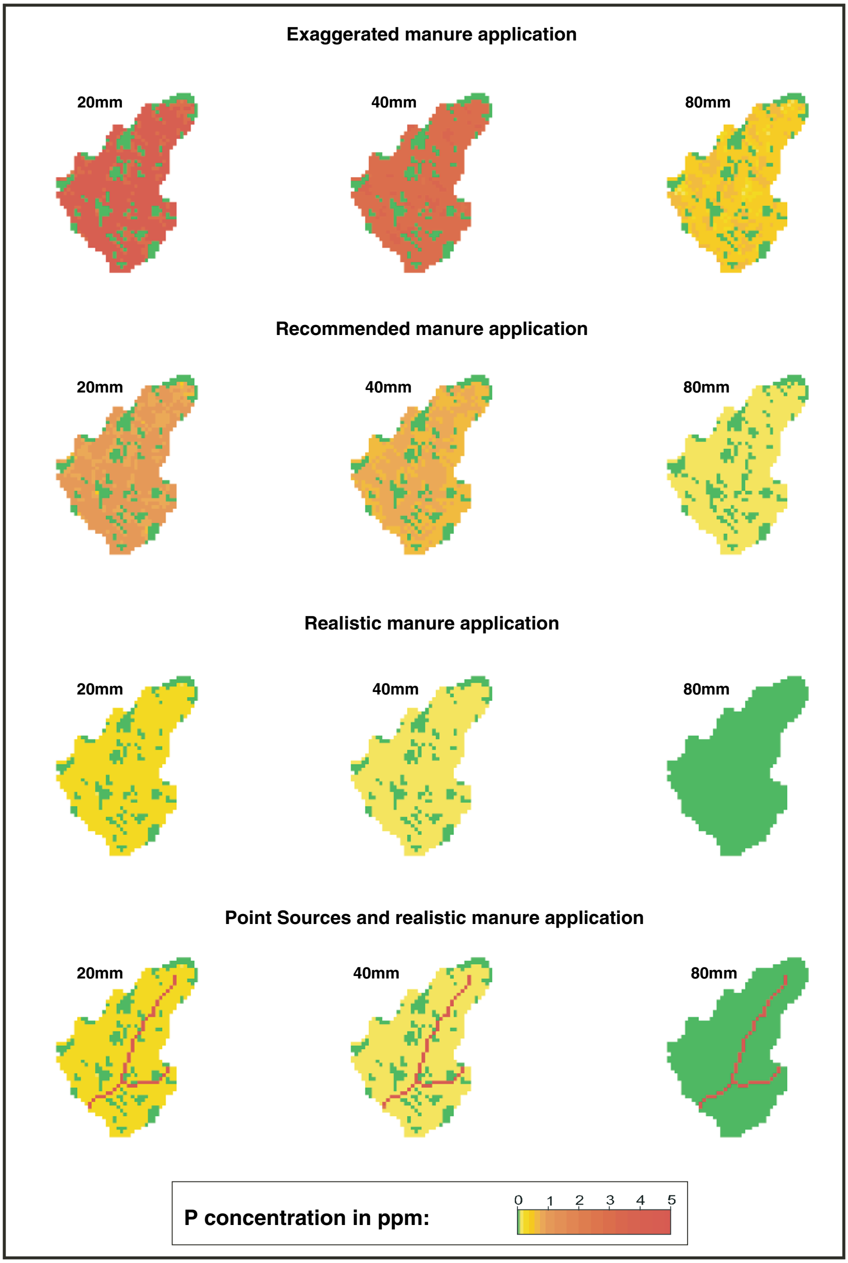


Fig. 12.2 P concentration (ppm) in runoff for four scenarios and three different storm sizes (See also Plate 14 in the Colour Plate Section)

Table 12.2 Main causes and general perceptions of pig manure pollution

	Marginal farmers			Consolidated farmers			Extensionist		
	Q1 (%)	Q2 (%)	Q3 (%)	Q1 (%)	Q2 (%)	Q3 (%)	Q1 (%)	Q2 (%)	Q3 (%)
Main causes[a]									
General high number of animals	100	64	0	80	40	20	28	44	39
Animal concentration	100	91	100	90	90	100	89	83	89
Ponds location	100	73	91	50	80	60	39	22	28
Inappropriate ponds building	91	73	100	60	80	80	17	28	28
Direct flow to the streams	91	82	100	50	80	70	78	67	83
Management	36	55	91	30	80	80	33	72	72
General perception									
Severity level[b]	64	45	27	20	20	20	44	56	44
Urgency for solutions[c]	91	45	27	10	0	10	56	56	44
Possibility for solutions[d]	91	18	55	60	40	80	89	72	83

Q1: Questionnaire 1, pre-visualization; Q2: Questionnaire 2, post-visualization; and Q3: Questionnaire 3, post-discussion.
[a] Proportion of respondents considering the cause to be very important (from three options: very important, important and slightly important).
[b] Proportion of respondents considering pig manure pollution to be very and extremely severe in the region (from five options: extremely severe, very severe, severe, slightly severe and not severe).
[c] Proportion of respondents considering the search for solutions to be very and extremely urgent (from five options: extremely urgent, very urgent, urgent, slightly urgent and not urgent).
[d] Proportion of respondents considering to be very difficult and difficult to find solutions (from four options: very difficult, difficult, easy and very easy).

12.4 Conclusions

Demand driven land evaluation has been suggested by several authors as an attempt to make the information more relevant and useful to rural decision makers for land use planning. The interactive approach presented here was clearly valued by the decision makers. It was the first time in the author's more than 15 years as a land evaluator that decision makers reacted to information presented by land evaluators. They praised, criticised, changed their perceptions, made suggestions and requested more information. Even the negative reactions were a positive achievement of this work, as it is better to correct the path earlier than to invest time and resources to realise later that the work was not useful. Results presented in this work showed that a number of new demands were raised by decision makers. One of the types of information requested was the existing, but previously ignored, physical land evaluation. In this case, as the number of soil scientists and financial resources are scarce in the region, digital soil mapping based on existing data (Mayr and Palmer, 2006; Chapter 4) is a potential alternative to help to answer to the increasing expected demands. Yet, the use of digital soil mapping techniques could have a key role in the rural decision making process, helping to give rapid answers to the decision makers, improving the efficiency of traditional soil surveys (see also Chapters 4 and 20), improving the quality of the given information

Table 12.3 Effectiveness and feasibility of measures to decrease pollution problems caused by pig manure

	Marginal farmers			Consolidated farmers			Extensionist		
	Q1 (%)	Q2 (%)	Q3 (%)	Q1 (%)	Q2 (%)	Q3 (%)	Q1 (%)	Q2 (%)	Q3 (%)
Effectiveness of measures for improvements[a]									
Decrease number of animals	73	73	9	30	10	0	17	17	33
Decrease concentration of animals	91	82	100	60	90	90	89	67	89
Change location of ponds	82	91	100	90	80	70	33	33	28
Vegetal streams protection	55	82	100	70	70	80	39	44	33
Manure transportation	91	100	100	70	90	90	33	72	67
Improve management of manure	91	100	100	90	90	90	89	89	94
Avoid direct flow	91	91	82	90	80	80	78	94	83
Feasibility of measures for improvements[b]									
Decrease number of animals	55	73	9	40	40	20	11	11	17
Decrease concentration of animals	27	73	0	70	60	60	17	44	33
Change location of ponds	64	82	9	90	80	80	39	50	39
Vegetal streams protection	82	100	91	80	90	90	72	89	83
Manure transportation	91	100	91	80	80	80	39	61	67
Improve management of manure	91	91	82	80	90	90	83	89	83
Avoid direct flow	73	91	91	100	80	80	89	94	100

Q1: Questionnaire 1, pre-visualization; Q2: Questionnaire 2, post-visualization; and Q3: Questionnaire 3, post-discussion.

[a] Proportion of respondents considering the measure to be very effective (from three options: very effective, slightly effective and ineffective).

[b] Proportion of respondents considering the measure to be feasible (from three options: feasible, slightly feasible and infeasible).

(Chapter 25) and improving the communication among the actors involved in the planning process through visualization tools, simulation models, etc. Finally, this research showed that demand-driven land evaluation approach is possible in practice and should be further explored, but its effectiveness needs time to be definitely confirmed.

Table 12.4 Gross margins for historical scenarios 1995–1999, for seven land-use options

	Gross margin (R$)[a]						
	Bean	Soybean	Onion	Garlic	Swine	Milk	Maize
1995	53.69	28.57	2031.65	−1144.76	16000.00	3497.29	74.70
1996	77.19	233.33	573.80	−783.78	9510.00	2513.39	139.54
1997	88.87	325.80	1267.60	2222.11	19045.90	2421.59	101.08
1998	14.77	168.83	1781.75	−2327.86	10916.10	2302.38	122.13
1999	−28.79	−53.18	637.81	−3962.84	−6258.40	−711.96	−6.90
Mean	41.15	140.67	1258.52	−1199.42	9842.72	2004.54	86.11
CV[b]	85.24	91.86	191.61	−52.61	100.57	125.97	150.14

[a] Brazilian Reais (€ 1,00 = R$1,80 in December 1999).

[b] Coefficient of variation.

Table 12.5 Preferences for land-use options under the hypothetical condition of a newly established farm (multiple answers allowed)

New activity	Marginal farmers			Consolidated farmers			Extensionist		
	Q1[a] (%)	Q2[b] (%)	Q3[c] (%)	Q1[a] (%)	Q2[b] (%)	Q3[c] (%)	Q1[a] (%)	Q2[b] (%)	Q3[c] (%)
Pig	50	38	–	43	71	71	17	33	39
Maize	50	13	50	86	57	86	67	72	67
Milk	38	38	13	43	43	71	83	89	83
Bean	13	13	–	–	–	14	17	6	6
Soybean	–	–	–	–	–	14	33	28	22
Onion	–	63	63	–	–	–	11	22	22
Garlic	–	–	–	–	–	–	11	–	–

[a] Questionnaire 1: before presentation
[b] Questionnaire 2: after presentation
[c] Questionnaire 3: after discussion

Acknowledgments The work presented is based on the author's PhD research. I would like to thank the promotor Prof. Dr. Ir. Arnold K. Bregt, the co-promotor Dr. David G. Rossiter and the collaborators Dr. Chris M. Mannaerts and Dr. M. K. McCall. The research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the International Institute for Geo-Information Science and Earth Observation (ITC) and the Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S.A. (EPAGRI).

References

Bacic, I.L.Z., 2003. Demand-driven land evaluation: with case studies in Santa Catarina, Brazil. Ph.D. Thesis, Wageningen University and ITC.

Bacic, I.L.Z., Rossiter, D.G., Bregt, A.K., 2003. The use of land evaluation information by land use planners and decision-makers: a case study in Santa Catarina, Brazil. Soil Use and Management 19, 12–18.

Bacic, I.L.Z., Bregt, A.K., Rossiter, D.G., 2006a. A participatory approach for integrating risk assessment into rural decision-making: a case study in Santa Catarina, Brazil. Agricultural Systems 87, 229–244.

Bacic, I.L.Z., Rossiter, D.G., Bregt, A.K., 2006b. Using spatial information to improve collective understanding of shared environmental problems at watershed level. Landscape and Urban Planning 77, 54–66.

Bouma, J., 1997. The land use systems approach to planning sustainable land management at several scales. ITC Journal 1997-3/4, 237–242.

Bouma, J., 1999. Land evaluation for landscape units. In: Sumner, M.E. (Ed.), Handbook of Soil Science. CRC Press, Boca Raton, FL, pp. E393–E412.

Buttler, I., Riha, S., Wilkens, P., Rossiter, D.G., Sampath, R., Philips, J., Simoens, P., Melkonian, J., 1997. General-purpose Atmosphere-plant-soil Simulator – GAPS. Cornell University, Ithaca, NY.

Cook, S., Jarvis, A., Gonzalez, J.P., 2006. Digital soil mapping for the 21st Century: where to from here? In: Mendonça-Santos, M.L. & McBratney, A.B. (Eds.), 2nd GWDSM, 2006: Program and Full Papers for the 2nd Global Workshop On Digital Soil Mapping – DSM for regions and countries with sparse spatial data infrastructure. Embrapa, 4–7 July 2006.

Lagacherie, P., 2006. Digital soil mapping: a state of art. In: Mendonça-Santos, M.L. & McBratney, A.B. (Eds.), 2nd GWDSM, 2006: Program and Full Papers for the 2nd Global

Workshop On Digital Soil Mapping – DSM for regions and countries with sparse spatial data infrastructure. Embrapa, 4–7 July 2006.
Mayr, T., Palmer, R., 2006. Digital soil mapping using existing data – difficulties and potentials. In: Mendonça-Santos, M.L. & McBratney, A.B. (Eds.), 2nd GWDSM, 2006: Program and Full Papers for the 2nd Global Workshop On Digital Soil Mapping – DSM for regions and countries with sparse spatial data infrastructure. Embrapa, 4–7 July 2006.
Palisade Corporation, 1998. @RISK. Release 3.5.2. Palisade Corporation, Newfield, NY.
Rossiter, D.G., 1996. A theoretical framework for land evaluation (with discussion). Geoderma 72, 165–202.
Van Diepen, C.A., Van Keulen, H., Wolf, J., Berkhout, J.A.A., 1991. Land evaluation: from intuition to quantification. In: Stewart, B.A. (Ed.), Advances in Soil Science. Springer, New York, pp. 139–204.
Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1987. AgNPS, Agricultural Non-point Source Pollution Model: A Watershed Analysis Tool. Conservation Research Report 35. United States Department of Agriculture, Washington DC.
Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1989. AgNPS: a nonpoint-source pollution model for evaluating agricultural watersheds. Journal of Soil and Water Conservation 44, 168–173.

Part III
Digital Soil Mapping – Methodologies

Chapter 13
Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping

R.A. Viscarra Rossel and A.B. McBratney

Abstract This paper discusses the potential of soil diffuse reflectance spectroscopy (DRS) for rapid and cheap soil analysis and its application to digital soil mapping. We consider both visible-near infrared (vis-NIR) and mid infrared (mid-IR) spectroscopy, the use of multivariate calibrations, the development of soil spectral libraries and the cost and benefits of soil DRS. Finally, we conclude with some thoughts on the potential use of the techniques for digital soil mapping and soil science generally.

13.1 Introduction

The acquisition of quantitative soil information is essential for effective soil management. Obtaining these data by systematic sampling using conventional survey and laboratory analyses is ineffective and prohibitively expensive. Further adoption of digital soil mapping demands the development and adoption of techniques to replace and/or enhance conventional soil survey and laboratory analysis (see Chapter 2). In this paper we propose the use of diffuse reflectance spectroscopy (DRS) in the visible-near infrared (vis-NIR: 400–2500 nm) and mid infrared (mid-IR: 2500–25000 nm) as a worthy candidates for this.

Both vis-NIR and mid-IR techniques are rapid, accurate and more economical than conventional methods of soil analysis, they do not use environmentally harmful chemicals, require fewer pretreatments, are non destructive and when combined with multivariate calibrations a single spectrum can provide estimates of a number of soil properties. The techniques are highly sensitive to both organic and inorganic soil composition, making them potentially useful and powerful tools for the assessment and monitoring of soil, its quality and function. The mid-IR contains a lot more information on soil mineral and organic composition than the vis-NIR (e.g. Janik and Skjemstad, 1995) and their multivariate calibrations across a wide range of soil types are more robust (Viscarra Rossel et al., 2006). The reason is that the fundamental

R.A. Viscarra Rossel
CSIRO Land and Water, Bruce E. Butler Laboratory, GPO Box 1666, Canberra ACT 2601, Australia
e-mail: raphael.viscarra-rossel@csiro.au

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

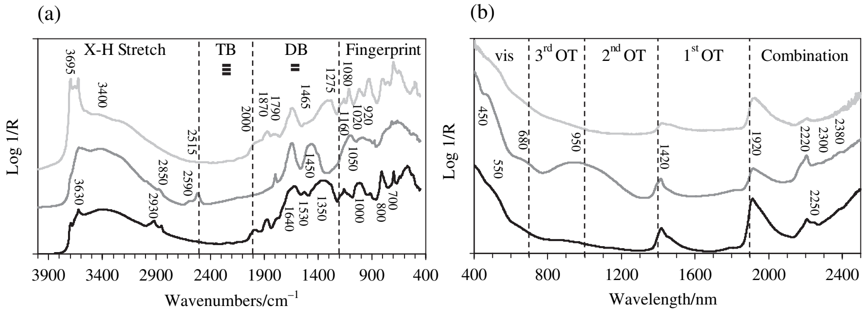


Fig. 13.1 Soil diffuse reflectance spectra in (**a**) the mid infrared 4000–400 cm^{-1} (or 2500–25000 nm) showing approximately where the fingerprint, double bond (DB), triple bond (TB) and X-H stretch regions, and (**b**) the visible and near infrared 400–2500 nm (or 25,000–4000 cm^{-1}) showing approximately where the combination, first, second and third overtone (OT) vibrations occur as well as the visible (vis) range

molecular vibrations of soil components occur in the mid-IR, while only their overtones and combinations are detected in the NIR. Hence soil NIR spectra display fewer and much broader absorption features compared to mid-IR spectra (Fig. 13.1).

Soil mid-IR spectroscopy is particularly well suited for the analysis of soil organic matter and mineral composition because absorption bands associated with both organic functional groups and soil minerals can be readily identified in soil mid-IR spectra (Fig. 13.1a). Visible-NIR spectroscopy can also be used to analyse soil, however, vis-NIR spectra (Fig. 13.1b) are visually much less interesting and difficult to interpret than those of the mid-IR. Absorptions bands are broad and they tend to overlap. Nevertheless, this region does contain useful information on organic and inorganic materials in the soil. Absorptions in the visible region (400–700 nm) are mostly due to electronic excitations and are primarily associated with the darkness of soil organic matter and to chromophores of iron containing minerals (Fig. 13.1b).

Diffuse reflectance spectroscopy has been used in soil science research since the 1950's and 60's (e.g. Brooks, 1952; Bowers and Hanks, 1965). However, it is only in approximately the last 20 years, most likely coinciding with the establishment of chemometrics and multivariate statistical techniques in analytical chemistry, that their usefulness and importance in soil science have been realised. So far, the published literature contains a vast number of investigations on the use of DRS as an analytical technique to complement soil analyses (e.g. Dalal and Henry, 1986; Janik and Skjemstad, 1995; Viscarra Rossel et al., 2006). The aim of this paper is to discuss the potential of soil DRS as an effective analytical tool for surveying and digital soil mapping.

13.2 Multivariate calibrations of soil diffuse reflectance spectra

Diffuse reflectance spectra of soil are largely non-specific due to interferences resulting from the overlapping spectra of soil constituents that are themselves varied and interrelated. This is particularly significant with vis-NIR spectra, which result from weaker overtones and combinations of vibrations occurring in the mid-IR. This lack of selectivity may be compounded by instrumental noise and drift and light-scatter and pathlength variations that occur during measurements. All of these factors result in complex absorption patterns that need to be mathematically extracted from the spectra so that they may be correlated with soil properties. Hence, the analysis of soil diffuse reflectance spectra requires the use of chemometric techniques and multivariate calibration (e.g. Martens and Naes, 1989). Chemometrics refers to the use of techniques for the mathematical or statistical treatment of chemical data, while multivariate calibration, in this instance, refers to the use empirical data and prior knowledge to predict an unknown soil property $\mathbf{y}$ from many spectroscopic measurements $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, simultaneously via a mathematical transfer function. Thus, soil spectroscopic calibrations are pedotransfer functions with large numbers of spectroscopic predictor variables.

Early research in DRS for soil analysis used calibrations based on only specific wavelengths that were selected using variable selection techniques such as step-wise multiple linear regression (SMLR) (e.g. Dalal and Henry, 1986; BenDor, 1995). The main reason for the use of SMLR was the inadequacy of more conventional regression techniques like multiple linear regression (MLR) and the unawareness of soil scientists of the existence of full spectrum data compression techniques such as principal components regression (PCR) and partial least squares regression (PLSR). Unlike MLR, PCR and PLSR can cope with data containing large numbers of predictor variables that are highly collinear. PCR and PLSR are related techniques and in most situations prediction errors will be similar. However, PLSR is often preferred by analysts because PLSR relates the response and predictor variables so that the models explain more of the variance in the response with fewer components, the models are more interpretable and the algorithm is computationally faster. Other multivariate and data mining techniques have also been used to calibrate soil spectra, however, mostly, with only limited advantages over PLSR. For example, Fidêncio et al. (2002) employed radial basis function networks (RBFN) to relate soil organic matter to soil spectra in the NIR region. Shepherd and Walsh (2002) used multivariate adaptive regression splines (MARS) for the estimation of soil properties from soil spectral libraries. Daniel et al. (2003) implemented artificial neural networks to estimate soil organic matter, phosphorus and potassium from their vis-NIR spectra. Brown et al., (2006) used boosted regression trees with vis-NIR spectra. Viscarra Rossel (2007) combined PLSR with bootstrap aggregation (bagging-PLSR) to improve the robustness of the PLSR models and produce predictions with uncertainty.

The literature has shown the potential of DRS and multivariate calibration for predictions of soil properties (Table 13.1). This summary shows that DRS can be used for the prediction of soil properties and that generally, the mid-IR produces better predictions than the NIR and the vis-NIR, and that the latter produced better predictions than the NIR or visible alone.

The studies in Table 13.1 use surface and subsurface soils and report results collected from single soil types with few samples (e.g. Masserschmidt et al., 1999; Walvoort and McBratney, 2001) to many soil types from different continents

Table 13.1 Predictions of soil properties using visible (vis), near infrared (NIR), mid infrared (MIR) and vis-NIR DRS. Data shown are average R^2 values of validation results reported by various authors from 1986 to 2006

Soil property	Vis	vis-NIR	NIR	mid-IR
Acid (exch.)		0.65^b	0.61^b	0.56^c
Al (exch.)	0.05^d		0.61^b	0.43^c
C (inorg.)		0.96^a	0.87^a	0.98^a
C (total)		0.89^a	0.90^a	0.95^a
C:N ratio		0.88^a		
CEC	0.16^d	0.76^b	0.68^b	0.79^b
Ca (exch.)	0.35^d	0.80^a	0.45^c	0.89^a
Carbonate			0.69^b	0.95^a
EC	0.10^d	0.38^d		0.31^d
Fe (DTPA)		0.69^b	0.49^c	0.55^c
K (exch.)	0.29^d	0.52^c	0.47^c	0.36^d
LR	0.25^d		0.62^b	0.81^a
Mg (exch.)		0.76^b	0.59^c	0.76^b
N (NO3)		0.63^b	0.04^d	0.06^d
N (total)		0.86^a	0.94^a	0.86^a
Na (exch.)		0.22^d		0.33^d
P (avail.)	0.06^d	0.81^a	0.10^d	0.14^d
pH_{Ca}	0.36^d	0.63^b	0.68^b	0.75^b
pH_w	0.36^d	0.61^b	0.62^b	0.66^b
Metal content: Cd, Cr, Cu, Pb, Zn		$0.45–0.93^{c-a}$		$0.66–0.99^{b-a}$
Clay	0.43^c	0.76^b	0.64^b	0.78^b
Sand	0.47^c	0.70^b	0.59^b	0.84^a
Silt	0.31^d	0.59^c	0.41^c	0.67^b
SSA			0.70^b	
Water		0.78^b	0.80^b	0.81^a
Biomass		0.60^c	0.75^b	0.69^b
Enzyme activity			0.55^c	0.70^b
OC	0.60^c	0.79^b	0.76^b	0.91^a
Respiration rate		0.66^b		

Note: R^2 values for predictions of soil properties are classified as: [a] very good (>0.81), [b] good (0.61–0.8), [c] fair (0.41–0.6) and [d] poor (<0.4). Other more appropriate statistics like the root mean squared error (RMSE) of predictions are not reported because they were missing in a lot of the original papers. Adapted from Viscarra Rossel et al. (2006).

(including tropical soils) with thousands of samples (Brown et al., 2006). Mostly, however, the studies include two to four different soil types and calibrations with 100–200 samples.

13.3 Development of diffuse reflectance spectral libraries

There is widespread interest for the development of soil spectral libraries. However, only a few that are geographically diverse and specific to soil properties have been described in the literature (e.g. Shepherd and Walsh, 2002; Dunn et al., 2002: Brown et al., 2006). Three important requirements for the development of a soil spectral library are: (i) it should contain as many samples as are needed to adequately describe the soil variability in the region in which the library is to be used; (ii) the samples should be carefully subsampled, handled, prepared, stored and scanned. Everything that has happened to the sample up to the time of scanning will be embodied in the sample and recorded in the spectra; and (iii) the reference soil analytical data used in the calibrations should be acquired using reliable and accredited analytical procedures. As with any type of modelling, the dictum here is ‘garbage in = garbage out’ and hopefully the converse ‘quality in = quality out’.

With relation to (i) above, if the library is being developed for a particular region from scratch, then the soil sampling strategy used will be critical (e.g. de Gruijter et al., 2006). If the library is being developed from a legacy soil sampling, then the recommendation is to scan all of the samples, then correlate the spectra with the relevant soil properties and use these to assess the quality of the legacy soil data. Figure 13.2a, illustrates the procedure used for the development of a soil spectral library from a legacy soil sampling of cotton-growing soils from northern NSW, Australia, and Fig. 13.2b the scheme for the development of a spectral library from a new soil sampling (Viscarra Rossel et al., 2008).

The schemes in Fig. 13.2 are divided into two sections: (1) for building the library and (2) for using it. The soil samples are prepared and scanned. The spectra are compressed using PCA to identify structure, patterns and possible clustering. Outlier analysis is performed on the multivariate data, and if outliers are present, the samples are rescanned for verification. From Fig. 13.2a, if there are no outliers the spectra are combined with the legacy soil data and their correlations assessed. If these are within acceptable limits, then the spectral library may be used to predict the properties of new soil samples that belong to the same spectral population as the

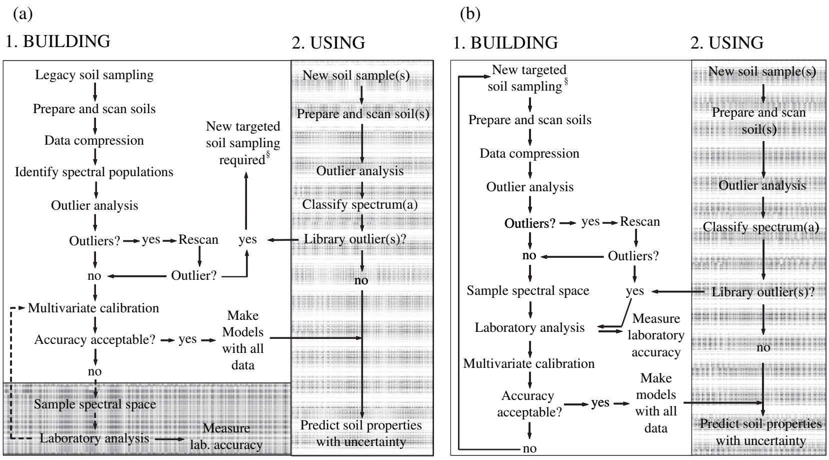


Fig. 13.2 Schemes for the development of a soil diffuse reflectance spectral library from (**a**) a legacy soil sampling and (**b**) a new soil sampling. The building of the library is shown by (1) and its use by (2). Adapted from Viscarra Rossel et al. (2008)

soils in the library. Conversely, if these correlations are unexpectedly poor, e.g. there is little or no correlation between the organic carbon and clay contents of the legacy samples and their spectra, then most likely the quality of the soil data are poor. If this is the case, then the spectral data space of samples in the library is sampled to select a representative subset that adequately describes the spectral diversity in the library. These samples are then subjected to the relevant soil laboratory analysis before the calibrations are derived. If the accuracies of the calibrations are acceptable, then they may be used to predict values of the reference soil properties for the entire library, thus repopulating the legacy data with good quality information. Clearly, these calibrations may also be used to predict soil property values for new samples that belong to the same population as the soils in the library. If samples are not represented by the spectral library, i.e. they are spectral outliers, they may be removed and a new targeted soil sampling mission may be required to populate the library with samples with similar characteristics. Then a similar procedure may be followed (Fig. 13.2b). In this case, if the new spectra classify poorly because they are not well represented by the library, they may be added to the library after reference laboratory analysis. Hence, the development of soil spectral libraries should be a continual process.

13.4 Costs and benefits of soil DRS for DSM

Two major costs associated with DRS include the initial investment in a spectrometer and the development of the spectral library. A bench top FT-IR spectrometer with an extended wavelength range to also include the NIR can cost somewhere in the vicinity of US$50,000, as can a portable vis-NIR instrument. The development of a spectral library with well characterised reference soil samples can be expensive. For instance, for a set of 500 well characterised soil samples analysed in an accredited laboratory with data on soil pH, organic carbon, total carbon and nitrogen, CEC and exchangeable cations, P, clay, sand and silt, can cost as much as a spectrometer. However, the number of samples needed for the calibrations may be reduced by sampling to select a representative subset of samples for conventional laboratory analysis (see 13.3 above). The more important point to stress here is that the quality of the library calibrations can only be as good as the quality of the data used to derive them.

Although US$100,000 may seem like a hefty investment, it is important to realise that the potential savings are immense. DRS is one of the simplest, most efficient and powerful spectroscopic techniques. Potentially, DRS can improve the analytical capabilities and efficiencies of either commercial or research laboratories and at the same time drastically reduce their costs. It can do this because: (i) measurements require minimal sample preparation as only a few grams of air-dry ground soil are needed; usually less than 2 mm for vis-NIR (although it can be used on unprepared soil due to the higher energy of their source and more sensitive detectors) and less than 200 μm for mid-IR; (ii) large numbers of samples can be scanned rapidly; for instance in our small research laboratory, using our mid-IR spectrometer we can manually scan 100 samples in 7 h and twice as many using our portable vis-NIR instrument. Obviously with automation these numbers could be substantially increased. Furthermore, predictions of soil properties can be made either in real-time or in batches after the spectra are collected, taking only a couple of extra minutes; and (iii) a single spectrum can be used to predict various soil properties, some of which are listed in Table 13.1. There can also be some value adding by combining DRS with soil inference systems (SPEC-SINFERS) to predict other important and functional soil properties via pedotransfer functions (McBratney et al., 2006).

13.5 Conclusions

DSM has much to gain from the adoption of soil DRS. Applications include not only soil analysis to improve soil survey and mapping but also soil classification, precision agriculture, and contaminated site assessment and management. There is great potential for the use of soil spectra in different areas of soil science, not only the analytical stream. For example, for soil organic carbon research spectra could be used directly as input into carbon turnover models, or in precision agriculture where spectra could be used directly in decision support systems to derive fertiliser recommendation, etc. However, training of young (and old) soil scientists is needed. The collection of soil spectra is easy. Spectra can be collected by the press of a button. However, it is much harder to manage and analyse the large volumes of data that the instruments generate. Soil scientist need to acquire the necessary quantitative skills to analyse and interpret these data. There is plenty in the literature on both DRS (e.g. Williams and Norris, 2001) and multivariate statistics and chemometrics (e.g. Martens and Naes, 1989). There are various commercial software products that can be used for the spectroscopic analyses and calibration, which although expensive, provide you with the full range of tools that are needed to analyse soil spectra. Alternatively, shareware and freeware packages that provide most of the functionality provided by commercial packages are also available (e.g. Viscarra Rossel, 2008), not to mention the free software environments such as the R-project (www.r-project.org/) and the readily available and downloadable codes for the mathematical and statistical processing of the data.

Acknowledgments We wish to acknowledge the Cotton Catchment and Communities CRC (CCC CRC) and the Grains Research and Development Corporation (GRDC) for their financial support.

References

Bowers, S.A., Hanks, R.J. 1965. Reflection of radiant energy from soils. Soil Science 100, 130–138
Brooks, F.A. 1952. Atmospheric radiation and its reflection from the ground. Journal of Meteorology 9, 41–52
Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of Americal Journal 59, 364–372

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G. 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273–290

De Gruijter, J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for natural resource monitoring. Springer Verlag

Dalal, R.C., Henry, R.J. 1986. Simultaneous determination of moisture, organic carbon and total nitrogen by near infrared reflectance spectrophotometry. Soil Science Society of America Journal 50, 120–123

Daniel, K.W., Tripathi, N.K., Honda, K. 2003. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). Australian Journal of Soil Research 41, 47–59

Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis - a case study from the Riverine Plain of south-eastern Australia. Australian Journal of Experimental Agriculture, 42, 607–614

Fidêncio, P.H., Poppi, R.J., De Andrade, J.C. 2002. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. Analytica Chimica Acta 453, 125–134

Janik, L.J., Skjemstad, J.O. 1995. Characterisation and analysis of soils using mid-infrared partial least squares. II. Correlations with some laboratory data. Australian Journal of Soil Research 33, 637–650

Martens H., Næs, T. 1989. Multivariate Calibration. John Wiley & Sons, Chichester

Masserschmidt, I., Cuelbas, C.J., Poppi, R.J., De Andrade, J.C., De Abreu, C.A., Davanzo, C.U. 1999. Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. Journal of Chemometrics 13, 265–273

McBratney, A.B., Minasny, B., Viscarra Rossel, R.A. 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. Geoderma 136, 272–278

Shepherd, K.D. & Walsh, M.G. 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal 66, 988–998

Viscarra Rossel, R.A. 2007. Robust modelling of soil diffuse reflectance spectra by 'bagging-'PLSR'. Journal of Near Infrared Spectroscopy 15, 39–47

Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics and Intelligent Laboratory Systems 90, 72–83

Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A. and McBratney A.B. 2008. Using a legacy soil sample to develop a mid-IR spectral library. Australian Journal of Soil Research, 46:1–16. doi: 10.1071/SR07099

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75

Walvoort, D.J.J., McBratney, A.B. 2001. Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. In G. Grenier and S. Blackmore (Eds), ECPA 2001, Third European Conference on Precision Agriculture, Vol. 1, Agro Montpellier, pp. 503–507

Williams, P., Norris, K. 2001. Near-infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, St. Paul, Minnesota

Chapter 14
Digital Soil Mapping at a National Scale: A Knowledge and GIS Based Approach to Improving Parent Material and Property Information

R. Lawley and B. Smith

Abstract One of the fundamental parameters in the soil formation equation is that relating to the parent material from which the soils have been derived. Such information is typically derived from geological surveys and paper maps. However, an increasing propensity to directly produce digital geological maps and associated data bases means that a far greater range of information can be made available to assist the soil scientist in mapping and predicting soil characteristics. Such geoinformation typically can include, detailed lithological parameters, geochemistry of soils and sediments, engineering parameters and remotely sensed information.

In this paper we describe on-going work at the British Geological Survey in which we are actively developing a national digital parent material map and property data base at a scale of 1:50 000. The main aim in doing this is to support the development of national soil data sets at a similar scale by those responsible for soil survey in the UK. However, our experience to date suggests that an adoption of similar strategies in regions and countries with sparse, soil orientated, data infrastructures could be of considerable value. For example many countries have, or are receiving, aid in support of the development and licensing of mineral resources (i.e. Madagascar, Afghanistan and Mauritania) which include not only significant improvements in geological mapping and associated GIS infrastructure, but also remote sensing and geochemical survey.

14.1 Introduction

In 2005 the British Geological Survey (BGS) began a five-year programme of research entitled ‘sustainable soil management’. The aim of the programme is to produce an integrated analysis of the UK’s near surface environment by developing (a) improved soil-parent material maps at 1:50 000 scale, (b) 3-D modelling

R. Lawley
British Geological Survey, Nottingham, NG12 5GG UK
e-mail: rslaw@bgs.ac.uk

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

capability of the near surface environment (see Chapter 16) and (c) increased knowledge of processes that occur within this near surface zone.

Drivers for funding this new work programme included an increasing scientific (Anderson et al., 2004; Zanner and Graham, 2005 and Wysocki et al., 2005) and legislative interest (DEFRA, 2004) in the characteristics of the near surface (the region typically less than 5 m below ground level). In undertaking this programme of activities BGS also hoped to be in a position to assist the UK soil-science community in meeting the needs of emerging soil policies that are based on the assessment of soil function.

In the UK it has been recognised that whilst the 1:250 000 scale soil maps (Mackney et al., 1983) offer full national coverage of soil information, there is also a need for similar data at 1:50 000 scale. This is required to support increasingly integrated European framework directives, England's soil action plan (DEFRA, 2004) and for managing the impact of climate change of soil resources. Unfortunately soil data at this scale is not systematically available in the UK (or indeed in many other countries). These factors together with the completion of the UK digital geological map at a scale of 1:50 000 and the implementation of a policy to digital archive all legacy geo-data at BGS (Jackson and Green, 2003) provided the catalyst for producing an improved soil-parent material map (see Fig. 14.1) as part of the BGS sustainable soil management programme.

As a result of the digitisation of legacy data, and publication of a digital geological map for the whole of the UK (solid and drift) it is now possible to offer a multitude of datasets and models, rather than 'just' the traditional single-purpose geology map. It is also possible for the first time to interactively trace the source information back to scanned archive material. However, in making digital maps geologists have had to discard traditional cartographic production values, and move toward more 'mechanistic' map data. Like soil maps, geological maps are a synthesis of empirical knowledge; the subtleties and detail of the rocks, traditionally shown by cartographic elegance of shading, or spatially unconstrained comments on the map face, are often lost once the map is digitised and made available through GIS. When that data is subsequently used by non-geologists or reprocessed by rule-based GIS systems to make derivative maps, the geological data often becomes problematic with a tendency towards overly complex or poorly defined classifications.

The Parent Material Map (PMM) shown in Fig. 14.2, is an attempt by BGS to remove the complexity of geology description as detailed in the BGS lexicon of named rock units (BGS, 2006) from its baseline datasets and at the same time re-introduce spatial data previously unpublished because it was regarded as 'outside the remit' of the national geological survey.

Once completed, it is intended that the PMM will become a fundamental baseline dataset for soil scientists to use in the creation of their own soil models for the UK (potentially via the *scorpan* variables as described by McBratney et al., 2003) where data relating to climate, land use, relief and parent material is available at similar, or better scales.

For the past decade a significant amount of World Bank funding has been focused on the upgrading of geological mapping via digital methodologies, remote sensing and geochemical survey in response to mineral exploration and mining sector devel-

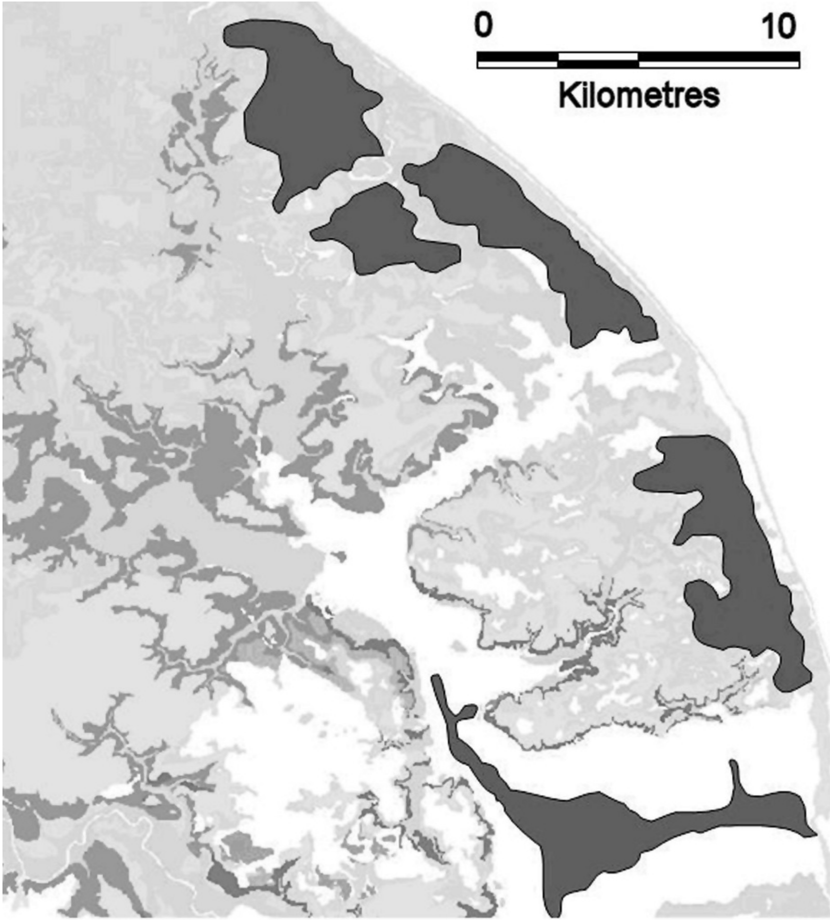


Fig. 14.1 Shows current geological map, with shaded area of unpublished loess information extracted from original fieldslips

opment. We propose that methodologies BGS have employed in the development of the UK PMM, together with the existence of upgraded information from the mining sector provide an opportunity for the production of digital parent material maps and associated databases. These digital maps in turn could represent a first, but significant step in closing the information gap, which hinders the development of digital soil models in countries with poor or non-existent data infrastructure.

14.2 Methodologies, Results and Discussion

The Parent Material Map derives its core spatial framework and descriptive content from main BGS dataset: DigMapGB50 (BGS, 2005). DigmapGB50 is a typical lithostratigraphical geology map and dataset. It has been derived from empirical

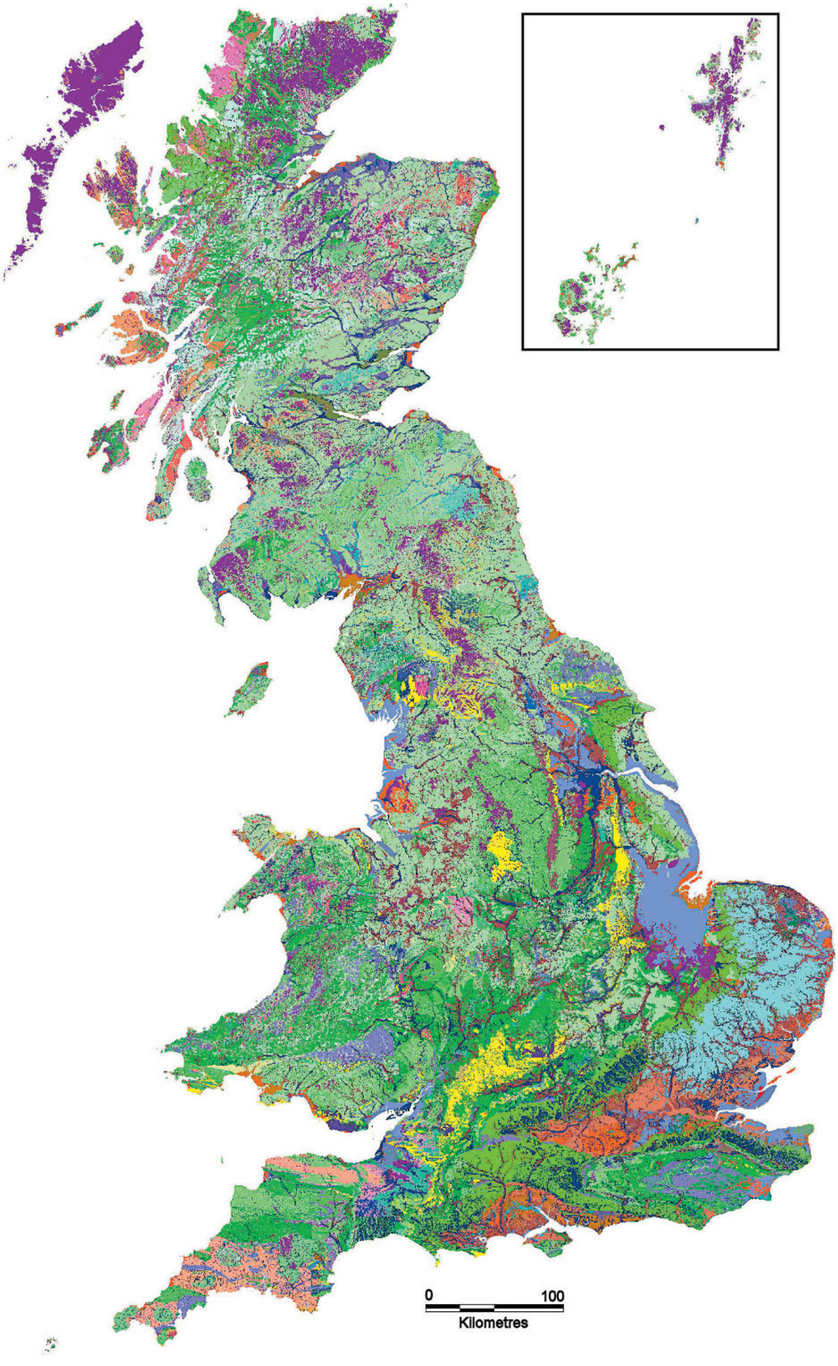


Fig. 14.2 Parent Material Map for Britain (version 0.1) (See also Plate 15 in the Colour Plate Section)

methods of geological survey over many years, at scales of between 1:10 000 and 1:25 000, with cartographic generalisation for final publication at 1:50 000 scale. Like all geological maps there is reliance upon the skills and ability of the surveyor to 'interpret' landscape, soils and outcrop to create the map. However, map quality is ultimately constrained by cost, complexity and survey objectives. In the UK (and in many international surveys) geological maps exhibit four common flaws:

1. A tendency to concentrate on hard rock (bedrock) geology or economic geology, at the expense of superficial (near–surface) deposits or un-economic strata.
2. A tendency for lithological descriptions to concentrate on 'fresh' material characteristics, with only brief descriptions of weathered material. (Soil information can rarely, if ever, be found on a published geological map).
3. A lack of freely accessible, *quantified* analysis and description of the rocks and associated properties
4. Highly clustered sample and survey patterns (databases are a 'patchwork quilt' of information rather than a seamless entity)

As a result of these flaws, some geological maps do not necessarily define what immediately lies beneath the soil and thus they can make poor parent-material maps. However, mining the archive datasets and reconstructing the geological map with an emphasis on near surface materials makes it is possible to correct these issues. For example, in the production of our UK PMM we aim to (1) use data mining and terrain modelling to spatially delineate unpublished surficial deposits, and (2) reattribute and simplify geological terminology to make data contained within the geological map more useful to potential users as a parent material map. In the longer term we also aim to incorporate information from:

1. The statistical integration of geochemical, engineering and other quantified survey information
2. The integration of remotely sensed data. For example DTMs, Spectral Gamma, Landsat, SRTM, ASTER, GPR etc.
3. The integration of 'traditional' soil survey and profile data (if available at similar scales, for improving and testing the simplification process on a regional basis)
4. Confidence mapping of the PMM

These short and longer terms aims are discussed in more detail below.

14.2.1 Data Mining for Unpublished Surficial Deposits

In the UK, geology maps are traditionally divided into two groups: Bedrock and Superficial. Bedrock mapping delineates rocks more than ∼2.6 million years old, Superficial mapping delineates deposits less than ∼2.6 million years old (also termed 'Quaternary' deposits). For much of BGS' 170-year history, bedrock mapping in the UK was driven primarily by a need to find mineral and coal resources. It has only been in the last 40 or so years that Quaternary and non-economic deposits have been surveyed as ‘primary targets’ for research. The Quaternary history of the UK comprises a cyclical succession of ice advance and retreat, the most recent of which, known as the ‘Devensian’, began ~80 000 years BP and ended ~10 000 years BP. During the latest stages of the cold period, northern UK was buried under glaciers and ice fields, whilst southern UK was subject to periglacial conditions. The Quaternary geology of the UK is a complex system of ice contact, outwash and soliflucted deposits. These deposits present specific difficulties for geological survey; they are generally heterolithic and spatially complex. Previous attempts to use geology maps for soil modelling (Mayr et al., 2001) have indicated that the geological map for the UK under-represents Quaternary and Holocene (recent) deposits at surface, and that three specific deposits form the bulk of the missing deposits: Peat, Colluvium and Loess (including coversand). These deposits exhibit blanket-like forms and where present are generally less than 1 or 2 m thick. Typically, these deposits have been observed during the early surveys and commonly their presence has been ‘indicated’ with map-face annotation, but not necessarily identified via a specific map boundary. As a result, a lot of field-based evidence within these legacy records is ‘unpublished’ and missing from the modern digital output. These unpublished data can be partly ‘restored’ to the PMM by undertaking an extensive data rescue and renewal process to seek out and digitally capture them (see Chapter 6).

Additionally, Peat, Loess and coversands have been exploited as resources in the UK and so some sparse sample information is available from BGS national and regional surveys for resources. Where archive searches fail, remote-sensing techniques can assist. Several satellite and airborne sensors provide spectra suitable for deriving pseudo-maps of Peat distribution, as well as Aerial photo interpretations. However, Colluvium presents a problem in that its heterolithic composition and ubiquitous and complex habit make accurate and consistent survey difficult. Descriptions of this soliflucted material are often deliberately ‘vague’, and in areas of Till deposits, delineation of Colluvium is extremely difficult. To resolve the possible extent of this deposit, we are using terrain analysis to determine areas where such deposits may have accumulated over time and analysis of ‘upslope areas’ for these locations to determine potential lithological characteristics (see Fig. 14.3).

14.2.2 Re-Attribution and Simplification of Geological Terminology

The reattribution and simplification of geological terminology is crucial to the success of the PMM. This is to allow a wider audience to easily use the dataset without having to become too geologically aware or spend resources on lengthy ‘background’ research. There are two reasons why this is a critical issue for BGS. Firstly, fewer environmental-science students are gaining the geological training needed to understand traditional geological datasets. Secondly, the creation of near-surface

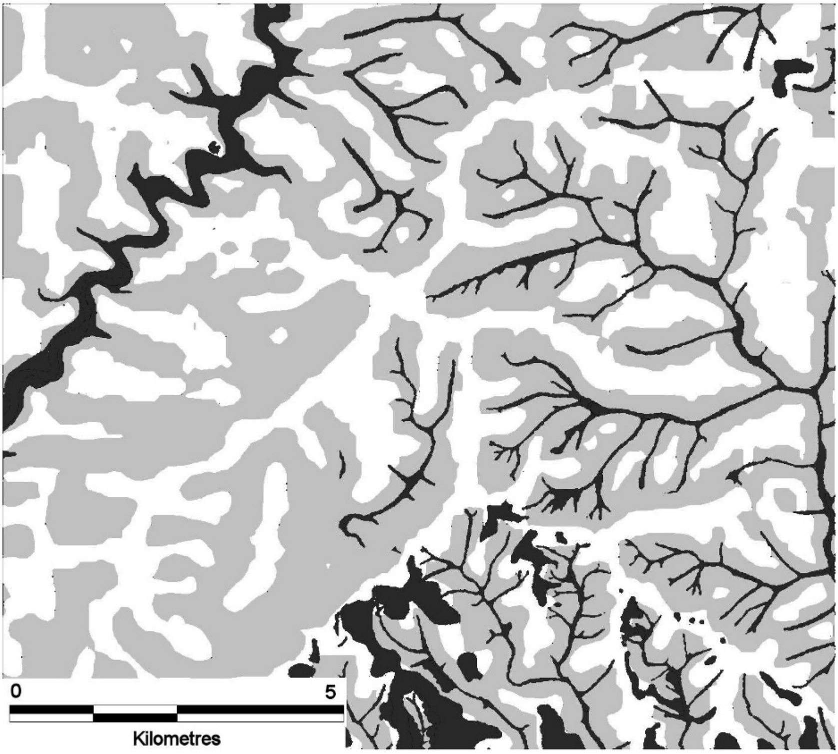


Fig. 14.3 Shows mapped superficial deposits (in *dark grey* colours) with modified 'zones' of modelled colluvium in *lighter grey*. Note the absence of mapped units on the *left-hand* side of the image, as a result of mapping policy at time of survey. Geologists have subsequently resurveyed this area of Devizes, UK and the new survey data indicates the modelled data offers a good representation of the colluvium distribution

and soils data is increasingly computer-based (specifically GIS) and for best results, computer models require simple binary rules for quantifying variables. Finally the number of geological units described in the BGS Lexicon, and mapped in the UK's 1:50 000 digital geological map has increased to over 9 000. This compares to the 26 parent material classifications used in the UK's national scale 1:250 000 soil mapping programme (Mackney et al., 1983). It is important; therefore, that we deliver a balanced level of attribute information that integrates with other users needs in a form that is optimised for computer manipulation.

Simplification of attributes for the PMM begins with the standard lithological description of the rock/deposits. All the deposits shown on the UK geology map are re-described using a hierarchical rock classification scheme developed by BGS (Gillespie and Styles, 1999). The user can thus 'step back' from the detailed description through the hierarchy, to a simpler root description. i.e. Troctolite can be reattributed as gabbroic rock, coarse crystalline igneous basic rock, igneous basic rock, or simply, igneous rock. At the same time, detailed information that is considered implicit in geology descriptions (i.e., texture or mineralogy) can be made more explicit for other users.

In its current version the PMM offers a series of relatively simple semi-qualitative attributes for each mapped deposit. These include: texture, colouration, mineralogy (including specific minerals of importance to soil properties), porosity, permeability, strength, density, carbon content, hardness and weathering structure. The range of attribution and quantified data within the PMM will increase over time as more BGS and external datasets are trawled for contributing information.

14.2.3 Integrating Other Data into the PMM

Most geological survey projects incorporate an element of geochemical and engineering sampling (as well as survey specific sampling). Geochemical surveys can be stream sediment or soil sample based, and whilst designed to expose mineral/chemical properties, they will invariably describe some soil and weathered-zone characteristics and so can be used to improve the geological map. Engineering surveys typically target weathered material as part of their remit as well as unweathered rock. Either dataset offers potential for soil scientists trying to find quantified parent material characteristics. Typically the data will be site specific (i.e. a single point in 3d space) and will require cautious application and/or extrapolation to the extents of available spatial geological data. However, basic information of texture, discontinuities colour, dominant mineralogy and bulk chemistry can all potentially serve as co-variables to assist in confirming/improving the more qualitative geological map classification. Satellite and airborne sensors have played fundamental roles in geological surveys, particularly in areas of limited vegetation cover, standard scenes of Landsat 7, ASTER are being used to improve the PMM. Airborne Hyperspectral and radiometric surveys are becoming more widely used in mineral surveys, these tools offer equally good opportunities for soil scientists to derive soil characteristics (see Section 2.2.1). These sensors will be tested in the UK to assess their integration with the PMM to determine soil mineralogy and possibly moisture content.

14.2.4 Confidence Mapping for the PMM

The PMM will evolve in the next four years from a ‘geology map’ into a ‘weathered zone’ model. It will comprise a combination of empirical observation, quantified point data and statistical models. In order to manage, and promote the use of this data, the PMM requires metadata and ‘confidence’ modelling to inform the user about how reliable the information is, and the limitations of its derivation. All the input layers to the PMM carry some form of quality assessment. Initial testing of the current version of the PMM has begun with a joint-research project involving the National Soil Research Institute in which soil scientists are testing and training the PMM by comparison with extensive soil sample datasets and soil survey archives.

14.2.5 Application in Data Poor Environments

The vast majority of soil surveying, on a regional scale, was performed in Africa and the Far East during the middle of the last century. This typically used geological information for differentiation of parent material type at a scale of 1:500 000 or larger as this was simply the only spatial framework available. Since that time higher resolution, systematic data has slowly become available as a result of mineral and resource exploration culminating with an intense period of activity during the early 21st century. These new data sources, for example in Mozambique, Morocco, Mauritania, Madagascar, Ghana and Papua New Guinea to name but a few, typically include geological survey at a scale of 1:100 000 or smaller; multi-element, multimedia geochemical survey and a suite of high resolution DTM and associated remotely sensed imagery including hyper-spectral, gamma spectrometric and radar based techniques. Whilst these mineral surveys have not explicitly deployed soil surveying, and therefore lack the micro-scale data needed to make a traditional soil map; they potentially contain vital data of use to the digital soil modeller. The issue becomes one of demonstrating the need for, value of, and capability of transferring the knowledge from mineralogical surveys into digital soil models, as discussed in Section 3.2.

14.3 Conclusions

Parent material mapping requires an extensive reconstruction of traditional geological linework. For the UK this involves a substantial data mining exercise, and has highlighted the need for better archiving of soil and weathering related observations. It is believed that the PMM will offer soil scientist in the UK a much-improved spatial framework on which to build 1:50 000 scale soil maps. Although the UK has abundant geological (and soil) data, it is evident from the work done so far, that many geological surveys across the world are acquiring abundant parent material data, of which only a small proportion is being ‘published’ within the geological map. For example many countries in Africa are revising, and in many cases remapping their geology in support of a worldwide resurgence in mineral exploration. This newly acquired material is often supported by high-resolution remote sensing data (airborne and satellite digital terrain maps, hyper spectral imagery and gamma spectrometry) which when coupled with geology offers a vast amount of information to those wishing to undertake digital soil mapping. Our experience is that geologists or soil scientists could relatively easily apply the simple data mining techniques and map reconstruction being used here to any such legacy geo-data. However, to do this, data needs to become much more widely available, with simpler and clearer terminology of relevance to soil science. It also has to be recognised that some geological maps may never make perfect parent material maps.

Acknowledgments The authors would like to acknowledge the assistance of the many BGS staff that have contributed to the development of the PMM. R. Lawley and B. Smith publish with the approval of the Director, British Geological Survey.

References

Anderson, S.P., Blum, J., Brantley, S.L., White, A.F., Chadwick, O., Chorover, J., Derry, L.A., Drever, J.I., Hering, J.G., Kirchner, J.W., Kump, L.R., Richter, D., 2004. Proposed initiative would study Earth's Weathering Engine. EoS Transactions of the American Geophysical Union 86(28), 265–272.

British Geological Survey, 2005. Digital Geological Map of Great Britain 1:50 000 scale (DiGMapGB-50). Version 2.11. Keyworth, Nottingham. UK.

British Geological Survey, 2006. The British Geological Survey Lexicon of Named Rock Units. (www.bgs.ac.uk/lexicon/lexicon_intro.html).

DEFRA, 2004. The first soil action plan for England 2004–2006. Department for the Environment, Food and Rural Affairs, London, UK, 36pp.

Gillespie, M.R., Styles, M.T., 1999. BGS Rock Classification Scheme Volume 1. Classification of Igneous Rocks, British Geological Survey Research Report, (2nd edition) RR 99–06.

Jackson, I., Green, C., 2003. DiGMapGB – The Digital Geological Map of Great Britain. Geoscientist 13(2), 4–7.

Mackney, D., Hodgson, J.M., Hollis, J.M., Staines, S.J., 1983. The 1:250 000 National Soil Map of England and Wales. Soil Survey of England and Wales, Harpenden.

Mayr, T.R., Palmer, R., Lawley, R., Fletcher, P., 2001. New Methods of Soil Mapping. Final Report: SR0120. National Soil Resources Institute, Silsoe, UK.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Wysocki, D.A., Schoeneberger, P.J., LaGarry, H.E., 2005. Soil surveys: a window to the subsurface. Geoderma 126, 167–180.

Zanner, C.W., Graham, R.C., 2005. Deep regolith: exploring the lower reaches of soil. Geoderma 126, 1–3.

Chapter 15
3D Modelling of Geology and Soils – A Case Study from the UK

B. Smith, H. Kessler, A.J. Scheib, S.E. Brown, R.C. Palmer, O. Kuras, C. Scheib and C.J. Jordan

Abstract Developments in GIS based technology have greatly aided the routine production of three-dimensional geological maps. Similarly the continued development of airborne remote sensing, geophysics and infrared measurement now provide tools that can assist in the mapping of soil structure and properties rapidly in 2D, 3D and even 4D. Whilst the combined use of such techniques have grown popular for performing site investigations and developing conceptual models of contaminated sites their use in determining and mapping soil has been restricted.

In this paper, we describe ongoing work at the British Geological Survey in which we have combined a variety of remote sensing, soil, geological and geophysical survey techniques to assist in the production of site specific, 3D digital soil models and geological maps. We were particularly interested in investigating (a) to what extent do methodological differences between the UK's soil and geological communities hinder the development of an integrated near surface model (b) whether technologies to map geology in 3D can be used to develop spatial models of the soil; and (c) can technologies used in digital soil mapping assist in reducing uncertainties associated with such models at a range of scales.

To date we have found clear evidence that differences in terminology do hinder the development of linked models of the near surface environment; but that such differences can be resolved by dialog between field surveyors from each discipline at an early stage in the process. The GSI3D software used in this work performed well in this, relatively simple usage and a successful 3D model of the Brakenhurst surface environment was obtained. However our attempt to use digital soil mapping techniques was compromised by the relatively poor contrast in soil properties across this specific site. Further investigations across representative soil landscapes are being carried out that should address this issue and provide more insight into the adoption of digital soil mapping techniques at a local scale.

B. Smith
British Geological Survey, Sir Kinsley Dunham Centre, Keyworth, Nottingham, NG12 5GG, United Kingdom
e-mail: B.Smith@bgs.ac.uk

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

15.1 Introduction

Digital soil mapping (McBratney et al., 2003, see also Chapter 1) is a rapidly expanding field that has similar aims to those being developed by the wider Earth science community. For example both soil science and geology rely on the use of a variety of digital techniques (e.g. satellite and airborne remote sensing and geophysical surveys) coupled to, spatial and numerical analysis and observational data. In addition, developments in portable GIS and computing technology within soil science and geological science communities allows interactive capture of field notes and development of spatially attributed maps and models in the field rather than on return to the laboratory (BGS, 2002).

Following its completion of the 1:50,000 digital geological survey of the UK (Jackson and Green, 2003) BGS began the development and subsequent licensing of 3D geological models across a range of scales (BGS, 2005a). This undertaking was commensurate with the commissioning of the BGS sustainable soils programme (BGS, 2005b, Chapter 14) whose aim, in response to drivers from existing and prospective European Union Framework directives (DEFRA, 2004), was the provision of better and more relevant geological information for soil science and survey. The simultaneous development of these two areas led to the hypothesis that both near surface pedological and geological information could be incorporated into future 3D geological models. In testing this hypothesis we were particularly interested in investigating (a) to what extent do methodological differences between the UK's soil and geological communities hinder this (b) whether technologies developed to map geology in 3D can be used to routinely develop spatial models of the soil environment at a site specific and catchment scale; and (c) can technologies used in digital soil mapping assist in reducing uncertainties associated with such models at a range of scales.

The first issue that we considered when testing our hypothesis was the potential impact of differences in classification that might prevent the effective use of existing field information. The basis for these concerns were that the Soil Survey and Geological Survey of the UK, have emerged into the 21st century with significant differences in methodology, nomenclature and scientific rationale about the genesis of the shallow subsurface. For example, in the UK soil maps have traditionally portrayed information for the agricultural community and other land based industries where topsoil (A-horizons) and Subsoil (B-horizons) and their properties were of most interest. Consequently, the majority of investigations were restricted to 1.5 m depth (Hodgson, 1997). Because of this interpretation, process orientated research and derived products have tended to focus on customers interested in nutrient and water availability, workability and soil erodibility.

Geologists in the UK, on the other hand, have investigated the shallow geosphere in a different manner. The aim of a geological survey is to map, describe and characterise the material that makes up the Earth, through survey and process orientated research. In the 19th century the BGS was mainly concerned with mapping the bedrock in order to find resources. However, in the latter part of the 20th century quaternary deposits and water resources became the dominant interest, but only in

Table 15.1 Examples of differing terminologies used for geological and pedological purposes in the UK

Geological terminology (Rock Classification Scheme, BGS 2006)	Pedological terminology (NSRI classification, Hodgson, 1997)
Mud	Loam
Clayey, sandy silt	Loam
Sand and Gravel	Drift with siliceous stones
Diamicton	Drift with siliceous stones
Marl	Calcareous mudstone
Coarse-grained crystalline intrusive rock	Igneous basic rock

the past few decades has the focus of interest shifted to the very shallow subsurface (<4 m). As a result, geological maps and field descriptions (BGS, 2000) describe the geology consistently, only below a metre, with thin intervening deposits often being ignored. A few striking examples of the differing nomenclature for parent material lithologies between Soil and Geological Surveys are listed in Table 15.1. At an international level the situation is even more complex and consequently a key challenge of any integrated soil-geological model, or to develop new data sets from existing geological and pedological data is the development of a consistent language.

The impact of these differences in approach to mapping and collecting property information on the UK's near subsurface (see also Chapters 14 and 25) may be minimised by undertaking multidisciplinary study and interpretation, as has been undertaken on this project. However, they have also resulted in fundamental knowledge gaps at the boundary of each surveys limits of investigation (Fig. 15.1).

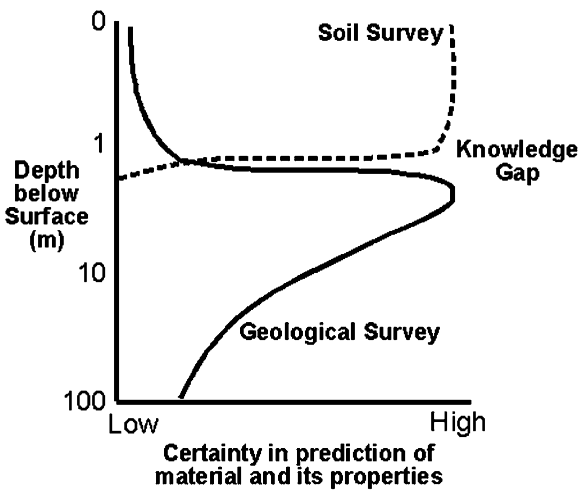


Fig. 15.1 Schematic diagram illustrating variation in the reliability of subsurface information with depth for a typical spatial survey site in the UK

Similar differences in approach have also been identified elsewhere. For example in the USA gaps in our understanding of processes in, and the spatial nature of, the deeper soil horizons and near surface geology have been identified (Anderson et al., 2004; Wysocki et al., 2005 and Zanner and Graham, 2005). The filling of this knowledge gap is critical for the development of any *subsurface information system*, which attempts to deliver solutions to problems that cut across the whole near surface environment, such as, groundwater vulnerability, corrosivity to infrastructure and carbon cycling.

It is also important to acknowledge that that the numerical assessment of uncertainty or confidence in spatial information data derived from field surveys (of soil and/or geology) still provides a challenge for the scientific community (e.g. Chapter 18). This is due to the interpretative nature of field survey, which is based not only on observed factual data from auger holes, trial pits and exposures but on a conceptual understanding of the evolution of the landscape and the processes that act and have acted in the shallow subsurface environment. A currently favoured approach in the Earth Sciences (Nordlund, 1996; Hwang et al., 1998) is to use fuzzy logic to produce numerical estimates of uncertainty by combining: data density and quality; geological and pedological complexity; the robustness of the conceptual model and expertise of a given surveyor.

We aim to test the hypothesis that near surface pedological and geological information could usefully be incorporated into future 3D geological models, during site-specific studies across a range of spatial scales and complexities representative of the main UK geo/soilscapes. At each site it is our intention to survey the environment, collecting a wide variety of shallow subsurface data and then subsequently model all geological units and pedological horizons as polyhedrons or volume objects (Grunwald, 2006) within a virtual soil-geology modelling environment. The 3D modelling software used in our tests is "Geological Surveying and Investigation in 3D" (GSI3D) developed by INSIGHT GmbH (Sobisch, 2000).

Recent work at BGS in conjunction with INSIGHT GmbH has demonstrated the feasibility of producing highly detailed 3D models of geological structures across a variety of scales and geoscapes. Similarly these developments have shown that 3D models better enable the non-specialist, who commonly represent the majority of end users for geological, and potentially also soil data to understand the complexities of, and uncertainties in the subsurface environment. 3D spatial models also represent a rapid and more accurate methodology from which to develop accurate conceptualised models for input into numerical modelling of groundwater infiltration and transport. In recognition of the positive benefits of this approach GSI3D has now been deployed across the BGS as the standard tool for 3D geological surveying (Kessler and Mathers, 2004) and exists as a tool on most geologists desks.

This paper describes our initial test undertaken in 2005 following the survey at the Brackenhurst campus of Nottingham Trent University near Southwell in the East Midlands region of the UK. The site was chosen for this study because of its relatively simple geological and pedological environment as evidenced by preliminary site surveys and larger scale mapping.

15.2 Methodology and Results

The Brackenhurst site is 2.5 km^2 in size and lies between 20 and 45 m above sea level. It is situated on typical red Triassic mudstones with some interbedded greenish grey silt- and sandstones. The area has been glaciated but was ice free in the latest glaciation, when the site was exposed to periglacial processes such as frost shattering and solifluction, which has resulted in head deposits covering the slopes and valley floors. During the Holocene the lower slopes and valley floors were filled with colluvial deposits. The soils on the site are mainly pelosols, brown earths and surface water gleys on the tops and slopes and some groundwater influenced soils on the valley floors (Palmer, 2006).

During the summer and autumn of 2005 the BGS undertook a complete site survey including many investigative and remotely sensed surveys listed in Table 15.2.

Table 15.2 Listing of surveys and methodologies used on the Brackenhurst site

Type of survey	Main methods used
GPS survey of boreholes and pits	Differential GPS System
Remote sensing and Terrain analysis	25 cm air photos, 5 m cell size DTM
Geological Survey	Walk over survey with drilling and pitting
Soil Survey	Walk over survey with 100 augerholes and 6 trial pits,
Geochemical Survey	200 m grid, three sampling depths (0.2, 0.5 and 1 metre), analysed for a range of major and trace metals and pH.
Geophysical Survey	2D Electrical resistivity tomography, electromagnetic mapping, Ground penetrating radar, downhole geophysical logging
Gamma Spectrometry Survey	Walk over survey with GPS; K, Th and U were interpreted
Hydrogeological Survey	Piezometer installations, soil moisture measurements

In addition to each survey, listed in the table above, delivering their own results in form of a map and a report, data was shared between surveys and collated into one software environment (GSI3D) for soil horizon and geological modelling.

In simple terms, GSI3D utilizes a Digital Terrain Model, geologically (and in this case soil) mapped linework and borehole or augerhole data to enable the geoscientist to construct regularly spaced intersecting cross-sections by correlating boreholes and the outcrops-subcrops of units to produce a fence diagram of the area. Mathematical interpolation between the nodes along the sections and the limits of the units or horizons produces a solid model comprised of a series of stacked triangulated volume objects.

The users draw their sections based on observational information such as borehole logs and surface features linking or correlating them with regard to, the morphology of the terrain, auxiliary information from geophysical-geochemical maps and measurements. Most importantly the shape of the correlated unit can be readily constrained or modified by the geoscientist. This process harnesses the modellers' professional understanding of geological and pedological processes, examination of exposures and theoretical knowledge gathered over a lifetime of fieldwork. Figure 15.2 illustrates typical stages in the development of a GSI3D model of the near surface.

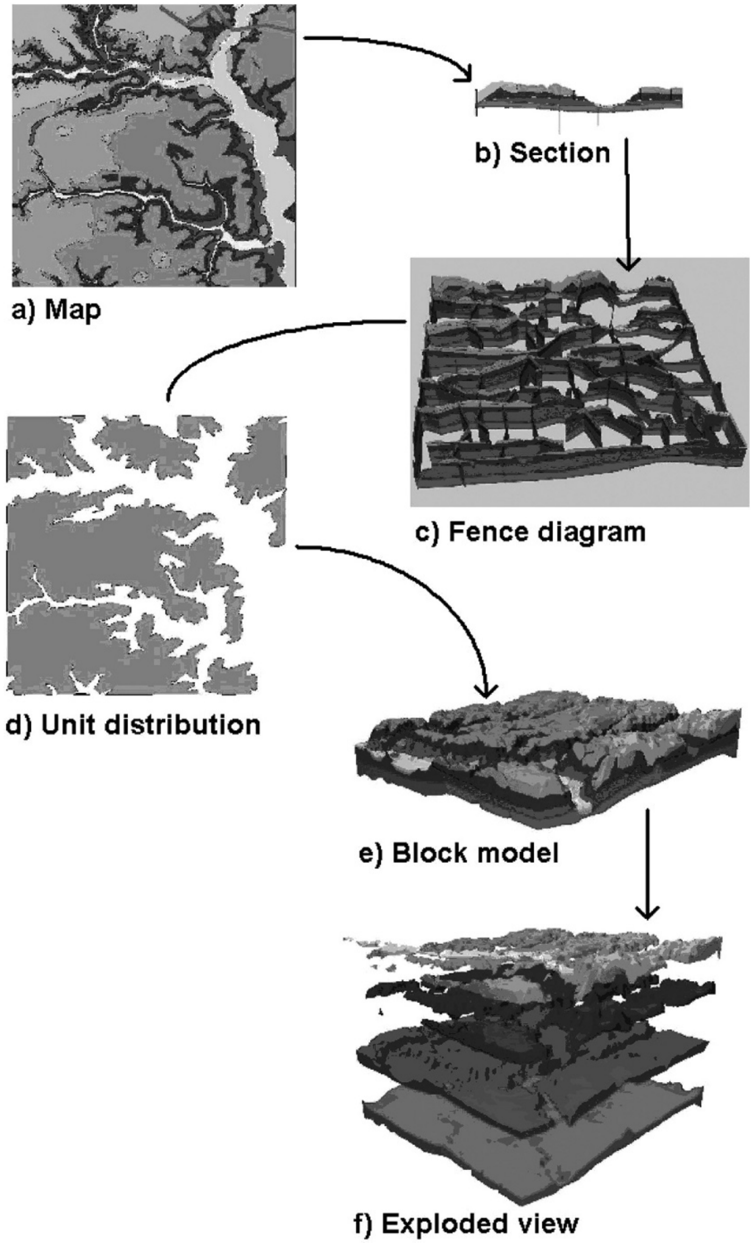


Fig. 15.2 Schematic diagram illustrating the development stages of a typical GSI3D Model

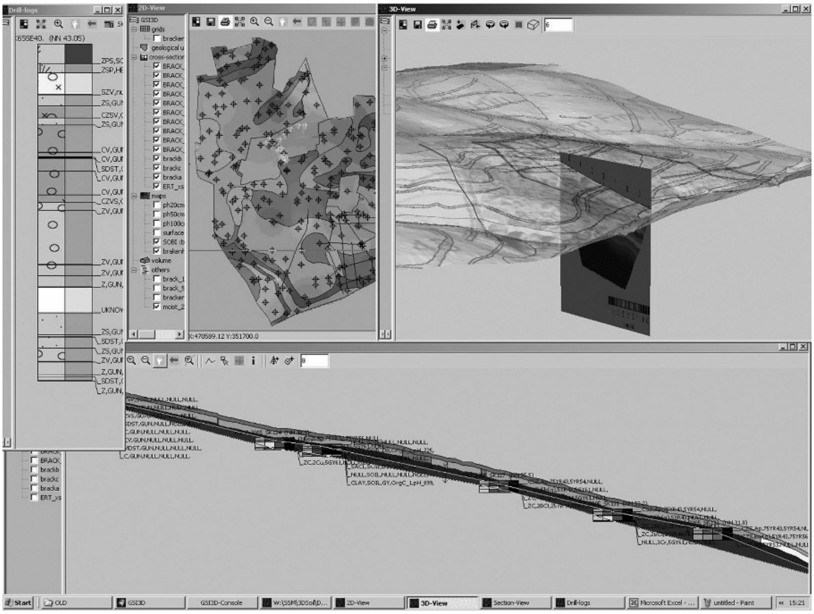


Fig. 15.3 GSI3D desktop screen shot showing views of data gathered in the field trial. These include geophysical logged borehole and soil map with all sample points in the 2D map window, geological data draped onto DTM with radiometrics survey and electrical tomography section in the 3D window and a section through several augerholes in the section window

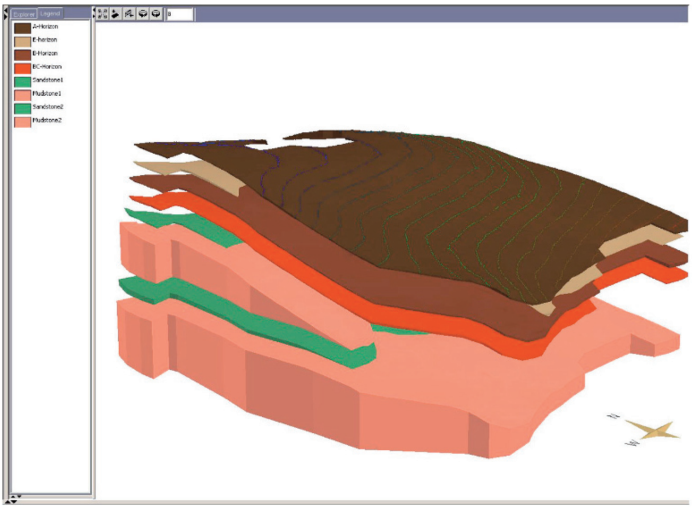


Fig. 15.4 Final 3D soil-geology model for the SW of the Brakenhurst site (500 × 500 m, 8x vertical exaggeration). The model shows the topsoil A-Horizon in dark brown, partially overlying the E-Horizon of a surface water gley in beige. The brown horizon is an amalgamation of all subsoil B-Horizons, which are underlain be the red B/C horizon which constitutes mainly weathered and soliflucted bedrock. Bedrock sandstone is shown in green, and mudstone in pink (See also Plate 16 in the Colour Plate Section)

Using the GSI3D methodology all data was assembled in one workspace (Fig. 15.3) and 4 soil horizons were correlated. For this exercise these were the topsoil (A horizon), the alluvial E horizon of a surface water gley, all combined subsoil horizons (B), and the transitional horizon between unaltered parent material and the soil profile (BC-horizon). In developing a 3D model (Fig. 15.4) the geo-scientist needs to be able to draw on all data gathered during the field campaign. Consequently a major task was the translation of all data into a geo-referenced framework.

Also all map and downhole data sets had to comply with one general standardised classification system with common attributes (nomenclature, property description and associated legends) such as the BGS rock classification scheme (BGS, 2006).

15.3 Discussion and Conclusions

Whilst study of the Brackenhurst site and its 3D model is ongoing we have been able to demonstrate:

1. That despite differences in nomenclature and methodologies between soil survey and geological surveys it is possible – provided cognisance is taken of the need to use a multidisciplinary approach – to produce fully attributed 3D models (and associated XML database) of the shallow subsurface.
2. The studies have shown, that spatial soil horizon modelling can use methods and software developed originally for shallow geological modelling, when soil horizons follow a unique super positional order. Ongoing studies in other more complex soilscapes at a site specific and catchment scale will be used to test this assumption and cases where super positional order breaks down.
   The modelling of horizons that do not have one single super positional order or are even overturned and convoluted e.g. in periglacially disturbed soils, can be solved using techniques developed for modelling faulted and overturned geology. Each single horizon being discretised as a unique object, therefore overcoming the need for a ordered system, just as in a fractured or faulted geological model. The problem is therefore not technological, but more importantly how to gather spatial information on these structures as part of a soil-geological field investigation. It is our belief that advances in geophysical techniques may, in favourable conditions, may assist in this regard.
3. Soil and Geology are to be seen as a continuum and must be studied and surveyed in an integrated manner, as customer requirements move beyond the traditional boundaries of compartmentalised science.
4. The potential for geophysical methodologies to assist in the interpolation and scaling up of models has yet to be fully tested because the available contrast in properties at the Brackenhurst site was limited. However, this is the consequence of the site being specifically selected for its low geological and pedological complexity. Information from recent investigations at more complex sites confirms this hypothesis.

Future research includes the attribution of the soil horizon volumes with properties and assessing their confidence, and gradually increasing emphasis on real-time data capture, temporal monitoring and modelling, site complexity and scale.

Acknowledgments The authors would like to acknowledge the assistance of Nottingham Trent University for granting access to the Brackenhurst College site for the purpose of this study. B. Smith, H. Kessler, A.J. Scheib, S.E. Brown, O. Kuras, C. Scheib, and C.J. Jordan publish with the approval of the Director, British Geological Survey.

References

Anderson, S.P., Blum, J., Brantley, S.L., White, A.F., Chadwick, O., Chorover, J., Derry, L. A., Drever, J.I., Hering, J.G., Kirchner, J.W., Kump, L.R. and Richter, D. (2004). Proposed initiative would study Earth's Weathering Engine. EoS Transactions of the American Geophysical Union 86(28): 265–272.

BGS (2000). Specification for the preparation of 1:10,000 scale geological maps (2nd Edition). British Geological Survey Research Report RR/00/02, Keyworth, UK.

BGS (2002). Proceedings of the workshop "Capturing digital data in the field", Keyworth, 25th and 26th April, 2002. (www.bgs.ac.uk/dfdc/).

BGS (2005a). 3D Framework (www.bgs.ac.uk/science/3dmodelling/).

BGS (2005b). Sustainable Soils Programme (http://www.bgs.ac.uk/sustainablesoils/index.html).

BGS (2006). The British Geological Survey Rock Classification Scheme (http://www.bgs.ac.uk/bgsrcs/home.html).

DEFRA (2004). The first soil action plan for England 2004–2006. Department for the Environment, Food and Rural Affairs, London, UK.

Grunwald, S. (2006). Three-dimensional Reconstruction and Scientific Visualisation of Soil-Landscapes. 373–392. In: Grunwald S. (ed.) Environmental Soil-Landscape Modelling, 488pp.

Hodgson, J.M. (1997). Soil Survey field handbook. Soil Survey Technical Monograph No. 5. NSRI, Silsoe, UK.

Hwang, D.M., Karimi, H.A. and Byun, D.W. (1998). Uncertainty analysis of environmental models within GIS environments. Computers & Geosciences 24(2): 119–130.

Jackson, I. and Green, C. (2003). DiGMapGB – The Digital Geological Map of Great Britain. Geoscientist 13,(2): 4–7.

McBratney, A.B., Mendonca Santos, M.L. and Minasny, B., (2003). On digital soil mapping. Geoderma 117(1–2): 3–52.

Kessler, H. and Mathers, S.J. 2004. From geological maps to models. Geoscientist 14: 4– 6.

Nordlund, U. (1996). Formalizing geological knowledge – With an example of modeling stratigraphy using fuzzy logic. Journal of Sedimentary Research 66(4): 689–698.

Palmer, R.C. (2006). The soils of Nottingham Trent University farm at Brackenhurst, Nottinghamshire. NSRI unpublished research report no. YE20006E for the British Geological Survey

Sobisch, H-G. (2000). Ein digitales raeumliches Modell des Quartaers der GK25 Blatt 3508, Nordhorn auf der Basis vernetzter Profilschnitte. Shaker Verlag, Aachen. 113pp.

Wysocki, D.A., Schoeneberger, P.J. and LaGarry, H.E. (2005). Soil surveys: A window to the subsurface. Geoderma 126: 167–180.

Zanner, C.W. and Graham, R.C. (2005). Deep regolith: exploring the lower reaches of soil. Geoderma 126(1–2):1–3.

Chapter 16
Landsat Spectral Data for Digital Soil Mapping

J.L. Boettinger, R.D. Ramsey, J.M. Bodily, N.J. Cole, S. Kienast-Brown, S.J. Nield, A.M. Saunders and A.K. Stum

Abstract We propose that Landsat remotely sensed spectral data represent useful environmental covariates for digitally mapping soil distribution on the landscape, especially in arid and semiarid areas. Based on the common conceptual model that unique soils are the products of unique sets of soil-forming factors, Landsat spectral data can represent environmental covariates for vegetation (e.g., normalized difference vegetation index, fractional vegetation cover) and parent material and/or soil (e.g., band ratios diagnostic for gypsic and calcareous materials). In areas with sufficient relief, topographic data (e.g., slope, compound topographic index) derived from digital elevation models (DEMs) can be combined with Landsat-derived data to quantitatively model soil distribution on the landscape. These digital data can by analyzed using commercially available image processing software. Various classification and analysis methods (e.g., optimum index factor; principle component analysis; unsupervised and supervised classification) can be used to recognize meaningful soil-landscape patterns. . Training sites can be selected from existing soil surveys or from areas that have actual field data collection points. Accuracy assessment with independent field observation can be performed, and various classification methods can be used to generate estimates of prediction error. Landsat scenes are spatially explicit, physical representations of environmental covariates on the land surface. While the 30-m spatial resolution and fairly coarse spectral resolution may limit some applications, the wide availability and low expense should facilitate the utility of Landsat spectral data in digital soil mapping.

16.1 Introduction

The reflectance or emissions of electromagnetic radiation from the Earth can be quantified using satellite sensor platforms (Lillesand and Kiefer, 2000). Reflectance and emission data can be analyzed to extract information about the Earth and its resources, as the physical and chemical properties of different surfaces vary across the

J.L. Boettinger
Department of Plants, Soils, and Climate, Utah State University, 4820 Old Main Hill, Logan, UT 84322-4820 USA
e-mail: Janis.Boettinger@usu.edu

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

electromagnetic spectrum. Satellite-derived remote-sensing imagery from a particular sensor can have a variety of spectral and spatial properties. Spectral resolution refers to the number of spectral bands and the width of the electromagnetic spectrum sensed in each band. Spatial resolution refers to the area on the ground represented by an individual pixel. Spectral bands with contrasting spectral characteristics for the minimum area (pixel size) on a given surface can be compared to differentiate features on the Earth's surface.

The sensors belonging to the Landsat satellite program of the National Aeronautics and Space Administration (NASA) have been used for many land surface applications for more than 30 years (NASA 2006; USGS 2006). The most recent sensor, Landsat 7 Enhanced Thematic Mapper Plus (ETM+), was launched in 1999. Landsat 7 ETM+ has seven spectral bands that integrate specific wavelength segments of the electromagnetic spectrum with a minimum spatial resolution of 30 m in the visible (3 bands), near-infrared (NIR; 1 band), short-wave infrared (SWIR; 2 bands), and 60 m in the thermal infrared (TIR; 1 band) (Table 16.1). The spectral and spatial resolution of Landsat 7 ETM+ bands 1–7 are similar to the older but still commissioned Landsat 5 Thematic Mapper (TM), which was launched in 1984. Landsat imagery is very affordable, with an ever-increasing amount of no-cost to low-cost data available (e.g., see no-cost Landsat images available for an area of Brazil in Fig. 10.2).

Jenny (1941) conceptualized soils (S) on a landscape as a function of five environmental factors: climate (cl), organisms (o), relief (r), parent material (p), and time (t). Conceptual models of soil formation have been used to predict patterns of soil map units in traditional soil survey, usually based on interpretation of aerial photography with field verification of soils and associated landscape feature (e.g., Soil Survey Division Staff, 1993). With the increasing availability of spatially explicit digital data, such as remotely sensed spectral data and digital elevation models, and the hardware and software for processing and analyzing vast amounts of spatial data,

Table 16.1 Landsat 7 ETM+ spatial and spectral band resolutions. Bands 1–5 and 7 are often used in digital soil mapping

Band	Spatial Resolution m	Spectral Range μm	Common Name
1	30	0.450–0.515	Blue
2	30	0.525–0.605	Green
3	30	0.630–0.690	Red
4	30	0.775–0.900	NIR
5	30	1.550–1.750	SWIR (MIR)
6	60	10.40–12.50	TIR
7	30	2.090–2.350	SWIR (MIR)
Pan	15	0.520–0.900	Visible + NIR

Abbreviations: NIR = near infrared; SWIR = short-wave infrared (formerly MIR = middle infrared, 1–3 μm; Lillesand and Kiefer, 2000); TIR = thermal infrared.

we can quantitatively predict soil distribution on the landscape. To represent soil and the related environment factors in a spatial context and express these relationships quantitatively, McBratney et al., (2003) proposed the *scorpan* model: At a point in space and time, soil (as either soil classes, *Sc*, or soil attributes, *Sa*) is an empirical quantitative function of the soil (s, as a class or as a directly or remotely sensed property), climate (c), organisms (o), relief (r), parent material (p), age (a), and spatial position (n).

Landsat remotely sensed spectral data can serve as useful environmental covariates for digitally mapping soil distribution on the landscape. This is particularly true in arid and semiarid areas where there is a range in vegetation cover and mineralogical properties of the soil surface and/or parent material are not completely covered by vegetation. Landsat spectral data represent real physical properties, and useful environmental covariates can be derived for vegetation, soil, and parent material, and their quantitative relationships used to predict soil distribution. In areas with sufficient relief, topographic data derived from digital elevation models (DEMs) can also be combined with Landsat-derived environmental covariates to develop predictive models of soil distribution. Various types of training and input data can be used, along with various methods of data classification and analysis using commercially available software. The objective of this chapter is to demonstrate the usefulness of Landsat spectral data in digital soil mapping for both initial and update soil surveys, using examples from our research in soil survey projects in the western USA and from this book.

16.2 Landsat Data as Environmental Covariates

Landsat spectral bands (Table 16.1) most commonly used in digital soil mapping are bands 1–5, and 7. Environmental covariates can be digitally represented using raster data layers derived from various band ratios of Landsat 7 ETM+ data. These derivations are known in remote sensing image processing literature as image enhancements, which are used to visually explore the imagery and/or for subsequent analysis (Jensen, 2005).

For the following examples presented from Utah and Wyoming, USA, we analyzed Landsat with ERDAS Imagine image processing software (Leica Geosystems, 2003). All data were projected in the same geographic space using the Universal Transverse Mercator (UTM) system, clipped to the extent of the study area, and converted into ERDAS Imagine file format (.img).

16.2.1 Vegetation

Probably the most common use of Landsat data for digital soil mapping is the Normalized Difference Vegetation Index (NDVI), which can represent the environmental covariate of vegetation. The NDVI is a normalized difference ratio model of the near infrared (NIR) and red bands of a multispectral image (Rouse et al., 1973).

Fig. 16.1 Normalized Difference Vegetation Index (NDVI) of an area along the Powder River, northeastern Wyoming. *Darker* areas have low vegetation cover (e.g., narrow, steep ridgetops and sideslopes), whereas *lighter* areas have high vegetation cover (e.g., relatively high-moisture areas adjacent to the Powder River). *Black* areas do not have vegetation, such as the Powder River and the major highway that transects the area from east to west

Using Landsat data, the NDVI is determined by using bands 3 (Red) and 4 (NIR): NDVI $= (4 - 3)/(4 + 3)$. This results in values ranging from –1.0 to 1.0, where higher values indicate higher vegetation density (e.g., Fig. 16.1). Recently, Landsat 7 ETM+ bands 7, 4, 2, and the NDVI calculated by substituting band 2 for band 3 were used as environmental covariates for vegetation (*o* in *scorpan*) in models predicting soil class distribution in Rio de Janeiro state in Brazil (see Chapter 34 for further discussion).

It is often useful to re-normalize the values of the NDVI to provide an estimated value of percent of vegetation cover (Zeng et al., 2000), referred to as fractional vegetation cover (FVC): FVC = [(NDVI – min NDVI)/(max NDVI – min NDVI)] * 100. Cole and Boettinger (2007) and Saunders and Boettinger (2007) used fractional vegetation cover as an environmental covariate to predict soil class distribution in arid rangelands in Wyoming, USA.

Vegetation types may also be identified using Landsat spectral data. For example, vegetation typically associated with soil classes in a region of the Central Amazon, Brazil, was identified optically using Landsat 5 TM data (see Section 29.2 in Chapter 29).

16.2.2 Parent Material and/or Soil

Landsat spectral bands, particularly in the short-wave infrared (SWIR) range (see Table 16.1), can be used to represent the environmental covariates of parent material and/or soil. Different mineral assemblages will have different spectral reflectances, which may be separable by analyzing bands 1–5 and 7.

Landsat images may be visually explored using only three bands at one time (assigned to red, green, and blue color guns). The 3-band combination that has the maximum variance and minimum duplication within the scene can be selected by calculating the optimum index factor (OIF) (Jensen, 2005). Nield et al. (2007) used the OIF to select Landsat 7 ETM+ band combinations of 1, 5, 7 for visually analyzing areas with gypsic soils, and 4, 5, 7 for areas with natric soils.

The soil enhancement ratios of Landsat spectral band ratios 3/2, 3/7, and 5/7 have been interpreted to accentuate carbonate radicals, ferrous iron, and hydroxyl radicals, respectively, in exposed soil and geologic materials (Amen and Blaszczynski, 2001). Cole and Boettinger (2007) and Saunders and Boettinger (2007) incorporated these soil enhancement ratios with topographic data to predict soil map unit distribution in the Powder River Basin and Green River Basin of Wyoming, respectively.

Normalized difference ratios of Landsat spectral bands, similar in form to the NDVI, may be developed to target specific mineralogical signatures of soils and/or parent materials. Nield et al. (2007) successfully mapped gypsic soil areas (50% or more gypsum by weight within a few cm of the soil surface) by focusing on the spectral response of gypsum, indicated by the high reflectance in band 5 relative to low reflectance in band 7: (5–7)/(5+7). The normalized difference ratio of Landsat bands 5 and 2 can be diagnostic for calcareous sedimentary rocks: (5–2)/(5+2). In a Basin and Range landscape with sedimentary rocks intruded by Tertiary volcanic rocks, the (5–2)/(5+2) normalized difference ratio clearly distinguished sedimentary rocks from igneous rocks (Fig. 16.2).

The British Geological Survey is augmenting existing geology maps with Landsat spectral data to create the Parent Material Map for Britain (see Fig. 14.2). This parent material map is intended to become a fundamental environmental covariate for modeling soil distribution in the UK (see Chapter 14 for further discussion).

Principal components analysis (PCA) can be valuable in the enhancement of Landsat spectral data. Raw spectral data are transformed into new PCA images that can compress vast amounts of information contained in the data scene (e.g., bands 1–5, 7) into a few principal components. This transformation can make the image easier to interpret visually for distinguishing parent material, as well as vegetation density (Fig. 16.3, B).

16.2.3 Land Use and Land Cover

Landsat spectral data, particularly bands 5, 4, and 3, are commonly used to characterize land use and land cover (as described in Chapter 22 for Rio de Janeiro, Brazil). Land cover can be strongly related to specific soil properties or classes.

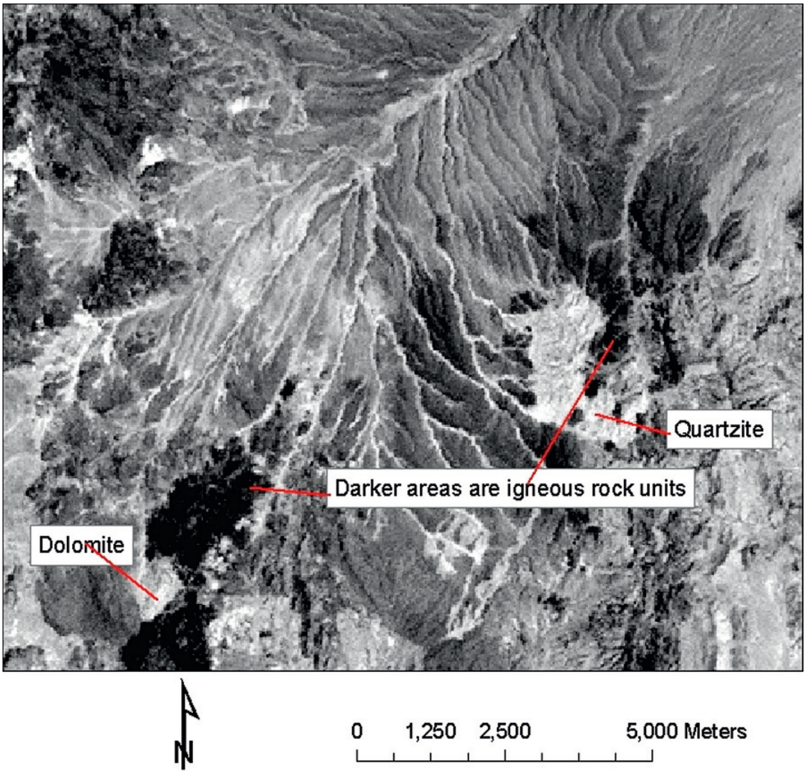


Fig. 16.2 Normalized difference ratio of Landsat 7 bands (5–2)/(5+2) clearly differentiates igneous rocks such as andesite (*darker* areas) from sedimentary rocks (*lighter* areas) such as dolomite and sandstone/quartzite at the northern edge of the Tonoquits Volcanic Field in the Basin and Range physiographic province, southwestern Utah

For example, land cover classified from Landsat 7 ETM+ bands 1–5 and 7 was strongly related to soil type on the margins of the Great Salt Lake in Utah, USA (Kienast-Brown and Boettinger, 2007). Land use and land cover are also strong indicators of human activity, which is often related to soil properties (see Chapter 30 for further discussion in a study in Brittany, France).

16.3 Classification

16.3.1 Layer Stacking

The NDVI, fractional vegetation cover, other specifically developed normalized difference ratios, PCA, and other Landsat data layers can be stacked with additional raster data layers for further analysis and classification. In the examples that follow, ancillary data representing the environmental covariate of relief were derived

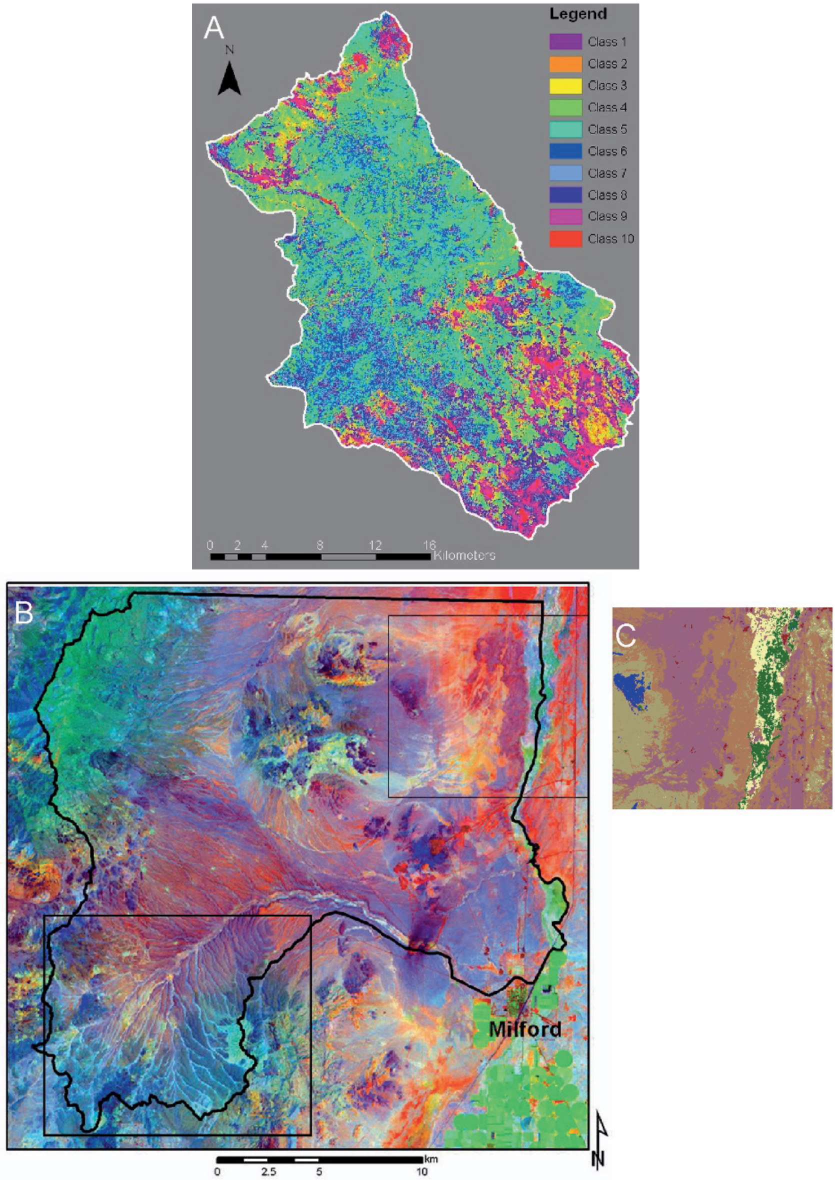


Fig. 16.3 (**A**) Map illustrating a 10-class unsupervised classification of a raster data layer stack containing the soil enhancement ratios of Landsat spectral band ratios 3/2, 3/7, and 5/7; slope; compound topographic index; fractional vegetation cover derived from the NDVI in the Green River Basin of Wyoming, USA. (**B**) Image of the first three components of the principal components analysis (PCA) of the Landsat 7 ETM+ image of a Basin and Range landscape in southwestern Utah. The box at lower left indicates the approximate location of the area shown in Fig. 16.2. The PCA image distinguishes areas of different parent materials (e.g., sedimentary vs. igneous rocks as indicated in Figure 16.2), and different vegetation density (e.g., high vegetation density at mountains tops in *upper left* and irrigated fields in *lower right* [*green* in Color Plate],

from U.S. Geological Survey (USGS) 10-m DEMs. For example, Saunders and Boettinger (2007) stacked raster data layers representing parent material (the soil enhancement ratios of Landsat spectral band ratios 3/2, 3/7, and 5/7), relief (slope; compound topographic index), and vegetation (fractional vegetation cover derived from the NDVI) for further classification for digital soil mapping using ERDAS Imagine image processing software.

16.3.2 Unsupervised Classification

Unsupervised classification, a type of cluster analysis (Leica Geosystems, 2005) can be performed on raster data layer stacks. Unsupervised classifications are generally considered to be unbiased and data driven. Layer stacks of various types of raster data can be classified using different numbers of classes until meaningful patterns are produced that could represent map units on the landscape. The goal is to seek patterns similar to the soil-vegetation-landform patterns interpreted from aerial photography in traditional soil survey. Unsupervised classification is particularly useful for developing an initial predictive map and/or stratifying the project area into manageable physiographic units. Saunders and Boettinger (2007) used unsupervised classification of a layer stack containing spectral and topographic raster data to elucidate patterns useful for developing an initial soil sampling plan in an arid shrubland of western Wyoming that lacked previous soil survey data (Fig. 16.3, A).

16.3.3 Supervised Classification

Supervised classification of layer stacks containing Landsat images, PCA, normalized difference band ratios, and/or ancillary data can also be performed. Supervised classification requires training sites selected from existing soil maps and/or in the field. The image is classified around cluster means derived from the training data (Leica Geosystems, 2005). Supervised classification of Landsat bands 1–5 and 7 and selection of training sites identified in the field was useful for refining maps of wet and saline areas in a soil survey update along the shore of the Great Salt Lake in northern Utah, USA (Kienast-Brown and Boettinger, 2007). In a Basin and Range landscape of southwestern Utah, we trained a supervised classification of the PCA of a Landsat image on field sites representing typical concepts of potential soil map

Fig. 16.3 (continued) in contrast to arid alluvial fans with low vegetation density in center [*red* to *purple* in Color Plate]). (**C**) Map illustrating a supervised classification of the PCA, focusing on the area indicated by the upper right box in B. Training areas were selected in the field for the supervised classification. Each shade (color in Color Plate) relates to a predicted soil map unit relating to soil class (e.g., loamy-skeletal Typic Haplocalcids) and dominant vegetation (e.g., *Artemisia tridentata* ssp. *wyomingensis* community). (See also Plate 17 in the Colour Plate Section) units, enabling us to delineate extended areas with these distinct soil-vegetation-landscape characteristics (Fig. 16.3, C).

16.3.4 Other Classifications and Analyses

There are many other possible classifications and analyses of Landsat data, with or without ancillary raster data (see Chapter 1, Section 1.4; McBratney et al., 2003). The alternate classifications can be raster (pixel) based or, as discussed in Chapter 30, object oriented.

Where possible, a formal accuracy assessment (Congalton and Green, 1999) of soil classes or properties predicted using Landsat data should be performed using independent field observations (e.g., Nield et al., 2007).

16.4 Conclusions

Landsat scenes are spatially explicit spectral reflectance data that represent real physical properties of environmental covariates on the land surface. Examples mentioned here illustrate the utility of Landsat data for digital soil mapping for both initial soil survey and soil survey updates. While Landsat imagery may be most useful for representing vegetation cover or parent material, these data may also represent land use or land cover, which can be related to soil properties or classes. Landsat bands 1–5 and 7 have a 30-m spatial resolution, which may limit some spatially detailed soil mapping applications. Whereas hyperspectral and ASTER imagery (see Chapter 2, Section 2.2.1) have finer spectral resolution, Landsat data can be as useful in digital soil mapping predictive models (e.g., see Chapter 4, particularly Tables 4.1, 4.2 and 4.3). In general, the wide availability, low expense, and increasing utility of image processing hardware and software should facilitate the utility of Landsat spectral data in digital soil mapping.

References

Amen, A. and Blaszczynski, J. 2001. Integrated Landscape Analysis. U.S. Department of the Interior, Bureau of Land Management, National Science and Technology Center, Denver, CO, pp. 2–20.

Cole, N.J., and Boettinger, J.L. 2007. A pedogenic understanding raster classification methodology for mapping soils, Powder River Basin, Wyoming, USA. In: P. Lagacherie, A.B. McBratney and M. Voltz (eds.), Digital Soil Mapping: An introductory perspective. Developments in Soil Science Vol. 31, Elsevier, Amsterdam, pp. 377–388.

Congalton, R.G. and Green, K. 1999. Assessing the accuracy of remotely sensed data: Principles and practices. CRC Press, Boca Raton, FL.

Jenny, H. 1941. Factors of soil formation. McGraw-Hill, New York.

Jensen, J.R. 2005. Introductory digital image processing. Prentice Hall, Upper Saddle River, NJ.

Kienast-Brown, S. and Boettinger, J.L. 2007. Land cover classification from Landsat imagery for mapping dynamic wet and saline soils. In: P. Lagacherie, A.B. McBratney, and M. Voltz (eds.), Digital Soil Mapping: An introductory perspective. Developments in Soil Science Vol. 31, Elsevier, Amsterdam, pp. 235–244.

Leica Geosystems. 2003. ERDAS Imagine V8.6. Leica Geosystems, Atlanta, GA

Leica Geosystems. 2005. ERDAS Field Guide. Leica Geosystems, Norcross, GA.

Lillesand, T.M. and Kiefer, R.W. 2000. Remote sensing and image interpretation. John Wiley and Sons, New York.

McBratney, A.B., Mendonça Santos, M.L. and Minasny, B. 2003. On digital soil mapping. Geoderma 117: 3–52.

NASA (National Aeronautics and Space Administration). 2006. The Landsat program. http://landsat.gsfc.nasa.gov/about/ (last verified 4 July 2007).

Nield, S.J., Boettinger, J.L. and Ramsey, R.D. 2007. Digitally Mapping Gypsic and Natric Soil Areas Using Landsat ETM Data**.** Soil Science Society of America Journal 71(1): 245–252.

Rouse, J.W. Jr., Hass, R.H., Schell, J.A. and Deering, D.W. 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351, NASA, Washington, D.C., pp. 309-317.

Saunders, A.M. and Boettinger, J.L. 2007. Incorporating classification trees into a pedogenic understanding raster classification methodology, Green River Basin, Wyoming, USA. In: P. Lagacherie, A.B. McBratney, and M. Voltz (eds.), Digital Soil Mapping: An introductory perspective. Developments in Soil Science Vol. 31, Elsevier, Amsterdam, pp. 389–399.

Soil Survey Division Staff, 1993. Soil Survey Manual. United States Department of Agriculture, Washington, D.C.

USGS (United States Geological Survey). 2006. The Landsat Project Website. (http://landsat.usgs.gov/index.php) (last verified 4 July 2007).

Zeng, X., Dickinson, R.E., Walker, A., Shaikh, M., DeFries, R.S. and Qi, J. 2000. Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling. Journal of Applied Meteorology 39(6): 826–839.

Chapter 17
From a Large to a Small Scale Soil Map: Top-Down Against Bottom-Up Approaches

Application to the Aisne Soil Map (France)

F. Carré, H.I. Reuter, J. Daroussin and O. Scheurer

Abstract This paper compares two approaches for upscaling the Aisne (a 7,536 km^2 French department) soil database from the initial 1:25,000 nominal scale to the 1:250,000 target scale. Soil features are represented at the nominal scale, whereas pedolandscapes, which are a combination of soil-forming factors and soil variables, are required at the target scale. Because the initial soil database does not contain soil forming factor information, data on pedogenesis have to be added to the initial database. Based on the assumption that most of lithographic layers are horizontal in the area, only landform attributes are chosen to represent the soil-forming factors.

Two different approaches are used to map the final pedolandscapes. The first one, called the bottom-up approach consists of classifying the soil and the landform attributes together for defining taxonomic units, which then undergo generalisation of their contours to result in pedolandscape mapping units.

The second approach, called a top-down approach, consists of classifying and then mapping the landform units in order to delineate the pedolandscapes. In this paper, we focus only on the pedolandscape delineation for the target scale. The results of the two methodologies are compared to contours manually drafted by soil surveyors. The final discussion analyses the impact of taking the very detailed soil database in the Digital Soil Mapping process into account, and to give advice for digital soil mapping with limited input data.

17.1 Introduction

The French National Program "Inventaire Gestion et Conservation des Sols" (IGCS) aims to provide Digital Soil Maps at a regional scale which is considered to be 1:250,000. These regional soil databases are composed of pedolandscape mapping

F. Carré
European Commission, Directorate General Joint Research Centre, Land Management & Natural Hazards Unit, TP 280, 21020 Ispra (VA), Italy
e-mail: Florence.Carre@jrc.it

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

units as geographic entities that are legible at the scale of 1:250,000. Each pedolandscape unit can be defined as an association of one or more Soil Typological Units legible at the scale of 1:100,000. The concept of pedolandscape can be defined as a combination of soil-forming factors and soil variables. It is also known in the literature as: landscape (Astle et al., 1969), polycombinational soil areal (Fridland, 1972), pedological province (Smeck et al., 1983), mesociation (Hole, 1978), small natural region (Favrot, 1981), soil system (Brabant, 1989), soilscape (Lagacherie et al., 2001).

The purpose of the study is to provide a pedolandscape database of the Aisne French Department at the scale of 1:250,000 using an initial soil database at a scale of 1:25,000 (see Section 1.3 about legacy data). As the final target scale is smaller than the initial one, the main core of the study can be related to upscaling (Bierkens et al., 2000), but because the final target feature (pedolandscape) is different to the initial one (combination of soil features), additional information has to be added to the initial database. Therefore this methodology differs from theoretical upscaling issues.

This paper only focuses on the delineation processes of the pedolandscapes (see also Chapter 30). The objective is to quantify the information brought by soil features for mapping pedolandscapes, versus an approach based only on auxiliary information (see Chapter 31). This allows to compare two approaches: one with an initial soil dataset (the bottom up approach), and another one with no prior soil information (the top-down approach) which can be associated with “Digital Soil Mapping with limited data”. We first present the two approaches and discuss the results in relation to soil surveyor’s expertise.

17.2 Materials

17.2.1 The Soil Database

The initial soil database holds 5 agronomic soil properties: carbonate rate, hydromorphic rate, texture, parent material when reached before 120 cm, nature and rate of gravel and stones, and depth to a textural change. The scale of representation is 1:25,000 (Soil database of the Agriculture Chamber of the Aisne Department).

The database contains many spatial allocation errors since all variables were digitised independently from each other, leading to a severe amount of sliver polygons. These sliver polygons increase the number of polygons (by an estimated 50%) and the size of the database. There are 336,951 polygons in the soil database which represent 14,091 different taxonomic units. The qualitative variables were transformed into Boolean variables when there was no logical coding rule, otherwise they were transformed into quantitative variables.

17.2.2 The Auxiliary Data

Because lithological layers within the study area are horizontal, the terrain morphology correlates with the lithology. We derived altitude, slope, a Wetness Index

(Gessler et al., 1995) and curvature (plan, profile and transversal curvatures) from a 50 resolution Digital Elevation Model (DEM) (BD ALTI®; IGN, 1998). Because, altitude and the Wetness Index explain 80% of the variability of the landform (based on PCA of the landform attributes – results not shown), only these two auxiliary attributes are considered. Altitude ranges from 36 m to 295 m, and the dimensionless Wetness Index from 0 to 20.

17.3 The Methodology

17.3.1 The Bottom-Up Approach (BU)

The BU approach is similar to Lagacherie et al. (2001), where some reference areas are chosen as being representative of the whole area and the remaining area is then classified according to the reference areas. The approach goes through two steps (Fig. 17.1a): a taxonomic aggregation, where the reference areas are built, and a geometric generalisation, to make the units legible at the 1:250,000 scale.

For the taxonomic aggregation, soil features were combined to landform attributes thus forming the taxonomic units. From these data, we chose some reference units for representing the whole taxonomic variability in order to form the pedolandscape units. The reference units were chosen as being the largest and most frequent

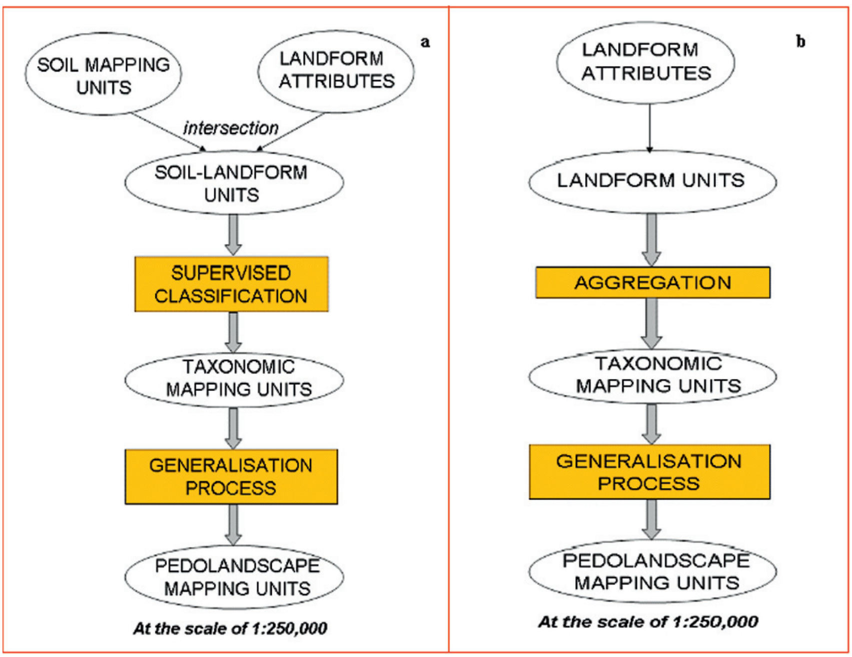


Fig. 17.1 Comparison of the bottom-up (**a**) and top-down (**b**) approaches

Table 17.1 Overview of the results of the different steps

Results of the different steps	Bottom-up approach	Top-down approach
Combination of the data		
- in number of units	- 41,305 units	- infinite (integer*real values)
- in number of polygons	- 336,951 polygons	- not applicable (raster grid)
After the aggregation		
- in number of units	- 256 units	- 25 units
- in number polygons	- 247,577 polygons	- 2,928 polygons
After generalisation		
- in number of units	- 31 units	- 25 units
- in number of polygons	- 820 polygons	- 651 polygons

units (in terms of polygons) based on the distribution histogram of the taxonomic units. They are 256 reference units (Table 17.1). The rest of the units (the smallest and less frequent units) were taxonomically aggregated to the reference units using a supervised classification algorithm (Carré and Girard, 2002). For the classification, landform attributes were weighted so that landform and soil had equal impacts on the classification.

This taxonomic aggregation process was applied to reduce the number of units (256 units) based on their attributes and the target scale: the smaller the target scale the higher the reduction. Despite this reduction, the number of polygons remained high (247,577 instead of 336,951 original polygons) and there remained a strong contrast between neighbouring polygons. This was due to the sliver polygons and to the large scale of the initial soil database. This prevented the neighbouring polygons from being simply merged in order to eliminate polygons that were not cartographically meaningful at the scale of 1:250,000 (smaller than 1.56 km^2 – Boulaine, 1980). Thus, a more sophisticated geometric generalisation was necessary by which small polygons were aggregated to their most taxonomically similar neighbour (Dobos et al., 2005) and simplified. This polygon generalisation was constrained by:

– geometry (area size): each polygon with an area less than 1.56 km^2 had to be removed;
– topology (neighbouring): polygon removal was done by merging with a neighbouring one;
– taxonomy (similarity of neighbours): choice of the best neighbour for merging was based on similarity of taxonomy (the most similar neighbour was selected for merging).

The process was iterative until all units were meaningful at the scale of 1:250,000.

17.3.2 The Top-Down Approach (TD)

The TD approach is based on the assumption that there is a discontinuity in the pedolandscape model when moving from large to small scale: at the 1:25,000 scale the pedogenesis is best modelled by agronomic variables whereas at the 1:250,000 scale it is best modelled by landform and lithology. As in the previous method, this approach also contains two stages (Fig. 17.1b): the first one is a taxonomic aggregation based on the combination of altitude and wetness index grids, and the second one is a generalisation.

Based on the distribution histogram of the values, the altitude was transformed into 12 classes and the Wetness Index into 4 classes. The combination of these classes resulted in the taxonomic units. This was followed by the same generalisation process as in the previous BU approach (small units were merged with their most similar neighbours) in order to get all mapping units legible at the scale of 1:250,000. There was no supervised classification as for the BU approach; this process is thus less sophisticated.

17.4 Results and Discussion

The results of the TD and BU approaches are presented and each approach is compared to a reference. The reference is a manually created 1:250,000 soil map drafted by the authors of the 1:25,000 soil map. Finally, we discuss the value of soil expertise for Digital Soil Mapping over automatic delineation.

17.4.1 Comparison of the Results Between the Two Approaches

Table 17.1 shows that the generalisation process is the main step that differentiates the two procedures: it allows a reduction of 88% ($256 - 31 = 225$ units) of the taxonomic information obtained after aggregating the units coming from the BU approach, whereas there is no reduction of taxonomic information with the TD approach.

This is mainly due to the very large scale of the original soil database which presents high taxonomic variability. As a consequence, taxonomic distances can be high between two neighbouring polygons obtained through the BU approach and threshold values for aggregating these must be high. This may result in aggregating polygons despite their strong differences from a pedological point of view. The two approaches are completely different; therefore the authors expected the results to also differ (Fig. 17.2).

The BU approach creates mapping units that are less segmented than the TD approach which, by definition, gives greater emphasis to the colluvions and alluvions (two pedogenetic phenomena that are strongly landform-dependent). In the depressions, the BU approach gives more detailed results because of parent material information contained in the soil database, contrary to the plateau. In the Northern area, the BU approach shows a big unit (in yellow in Fig. 17.2a) which represents a change in soil texture. This is very important (for example) for soil erosion estimation but it doesn't appear with the TD approach (Fig. 17.2b). As

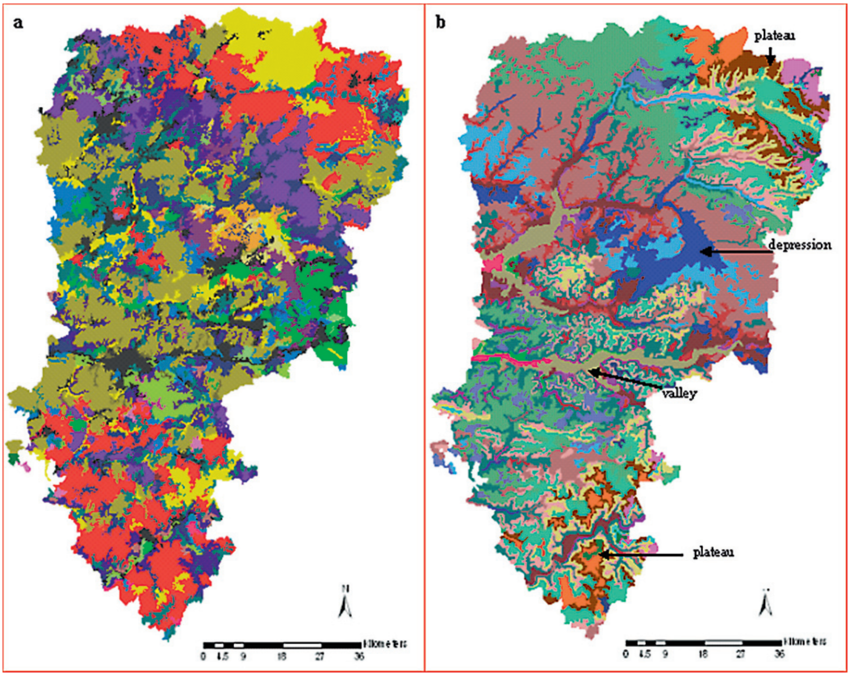


Fig. 17.2 Maps resulting from the bottom-up (**a**) and the top-down (**b**) approaches (See also Plate 18 in the Colour Plate Section)

a consequence, the BU approach, that contains soil information (related also to parent material) allows for greater emphasis on pedogenetic processes due to parent material (where lithologic layers are not horizontal), whereas the TD approach emphasises (by construction) the landform-induced pedogenesis.

17.4.2 Comparison of the Results Between the Automatic Approaches and the Manual Approach

The two manually drafted maps at the scale of 1:250,000 (in the South and East parts) were mainly based on two soil maps at the scale of 1:100,000 derived from the initial soil map by the same authors (Jamagne, 1967; Roque, unpublished; Guerin et al., unpublished). The final upscaling at the 1:250,000 scale was drafted from these soil maps by an experienced local soil surveyor (Fig. 17.3). Thus, we assume that the results are comparable to our BU approach.

Visually, the Southern part seems more detailed in the top-down than in the bottom-up approach. The opposite remark stands for the East part. The manually drafted map seems to be more segmented (*c.f.* black contours in Fig. 17.3a,b) in the East part than in the South one. For a comparison between the automatic approaches

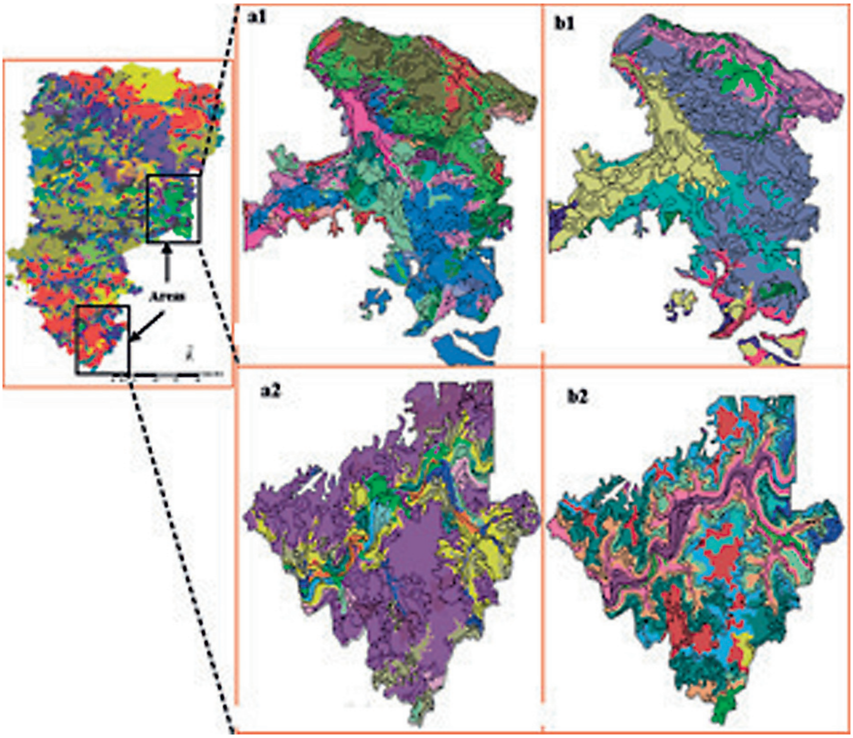


Fig. 17.3 Comparison between contours of the manually drafted 1:250,000 map (in *black*) to the results of the bottom-up (**a**) and top-down (**b**) approaches for the Eastern (1) and Southern (2) parts (See also Plate 19 in the Colour Plate Section)

and the manually drafted map, we take the last as the reference. We are aware of the weighted Kappa coefficient (Fleiss and Cohen, 1973) but since we focus only on delineation (not on taxonomy), we prefer to use different indices (Table 17.2) which are:

(1) the ratio between the number of polygons for the manual and the "automatic" maps (Segmentation Index),
(2) the average number of "automatic" polygons partially contained in each manual reference polygon (Inclusion Index),
(3) the average number of manual reference polygons partially contained in each "automatic" polygon (Dispersion Index),
(4) the ratio of Inclusion Index/Dispersion Index (the closer to 1 the better).

In the Eastern part, the Segmentation Index shows that for the BU approach there is an average of about two polygons per reference polygon (Table 17.2). In contrast, we observe that there are approximately three polygons of the TD approach per reference polygon (Table 17.2). The Inclusion Index shows that an average of four polygons of the BU approach is observed within one reference polygon (Table 17.2).

Table 17.2 Indices for the comparison of the manual reference map to the "automatic" maps

Reference East	Bottom-up approach	Top-down approach
Segmentation Index	210/114 (1.84)	210/66 (3.18)
Inclusion Index	4.45	2.1
Dispersion Index	7.72	10.28
Inclusion/Dispersion	0.58	0.20
Reference South		
Segmentation Index	127/83 (1.53)	127/80 (1.58)
Inclusion Index	4.94	3.92
Dispersion Index	5.88	6.20
Inclusion/Dispersion	0.84	0.63

In contrast an average of two polygons is observed for the TD approach. The Dispersion Index shows that eight reference polygons in average are observed within one BU approach polygon (Table 17.2), whereas ten reference polygons are observed within one TD approach polygon (Table 17.2). The Inclusion/Dispersion index gives an estimate on the polygon matching between the automatic approaches and the manual reference. For the Eastern part, this ratio of the BU approach is closer to 1 (Table 17.2) compared to the TD approach. For that part, the bottom-up approach is clearly the best method to use. However, for the Southern part, both methods are comparable (results are almost the same). This is due to the tabular relief of the South part, where altitude is strongly correlated to lithology. In that case, the TD approach can be recommended for digital soil mapping with limited data since DEM is the most relevant factor of the pedogenesis.

17.5 Discussion

In this study, the results are validated against a reference map. This may be done only if the reference map itself is considered valid. However, the map itself has not yet been validated. Therefore, our results have to be considered with caution.

Furthermore, because the contours of the reference map were delineated on the basis of the initial soil map (as for the BU approach), we need to explain why the Inclusion/Dispersion index is not so close to 1 (perfect concordance between BU and reference map contours). This can be explained by:

- auxiliary data not being used in the same manner;
- the fact that other auxiliary data (such as lithology and vegetation) must be taken into account in the aggregation process;
- the spatial allocation errors (sliver polygons) of the soil database which do not appear in the paper map at 1/25.000 scale (the reference map was created from the paper map);
- the generalisation process which is very sensitive to taxonomic distances. Both algorithms (generalisation process and calculation of taxonomic distances) should be then evaluated.

In that paper we showed a BU and a TP approach for soil data information stored in polygon form. A limitation is certainly that more advanced algorithms like fractals or wavelets (see McBratney, 1998) have not been evaluated. Further on we did not rely on point information as shown by Heuvelink and Edzer (1999), which could further elaborate about change in support, effects of aggregation, and interpolation (Bierkens et al., 2000). However, we were able to acquire knowledge, in which situations the BU and TP algorithms allowed for estimation under the special premises of Digital Soil Mapping.

17.6 Conclusions

This study aimed to test two different approaches for automatically delineating soil mapping units at the 1:250,000 scale. The first delineation was derived by generalising an existing, very detailed, large scale soil map (1:25,000) (an ideal situation since usually such data does not exist for such a large extent) using also auxiliary data (Bottom-Up approach). The second (Top-Down approach) was based on auxiliary data segmentation (common procedure used in DSM with limited data) – in this case, only the DEM was used. The results were compared to a reference map drafted manually from the original large scale soil map.

Comparison of the two approaches allows us to conclude that the Top-Down approach is recommended for digital soil mapping with limited data (i.e. in the absence of soil data). Nevertheless this implies the availability of auxiliary data representative of the pedogenesis and of good quality to be used as inputs. In our study, it was difficult to conclude on the efficiency of the Bottom-Up approach as the soil data we used had severe quality deficiencies. There again the Top-Down approach can be seen as a good alternative to the heavy means needed to improve an existing soil database before applying a Bottom-Up approach. The present work would benefit from being applied in a situation where good quality, large scale soil data is available. It would allow comparing the performance of the two methods independently from the differences in the quality of the input data.

Acknowledgments This work was financed by the French Ministry of Agriculture through the IGCS project. We would like to acknowledge the Chamber of Agriculture of the Aisne Department, the owner of the original soil database. They kindly gave us access to the original soil database and to the Digital Elevation Model. The INFOSOL Unit of INRA (National Institute of Agronomy Research) has given support to this project by delivering the maps at scale 1:100,000 and by approving our grant demand to the French Ministry of Agriculture.

References

Astle, W.L., Webster, R., Lawrance, C.J., 1969. Land classification for management planning in the Luangwa valley of Zambia. Journal of Applied Ecology 6, 143–169.

Bierkens, M.F.P., Finke, P.A., de Willigen, P., 2000. Upscaling and downscaling methods for environmental research. In: M.F.P. Bierkens, P.A. Finke, P. de Willigen (Eds.), Developments in Plant and Soil Sciences, Vol. 88. Kluwer Academic Publishers, Doredrecht.

Boulaine, 1980. Pédologie appliquée, Coll. Sciences agronomiques. Masson, Paris.
Brabant, P., 1989. La cartographie des sols dans les régions tropicales: Une procédure à 5 niveaux coordonnés. Science du sol 27, 369–394.
Carré, F., Girard, M.C., 2002. Quantitative mapping of soil types based on regression-kriging of taxonomic distances with landform and land-cover attributes. Geoderma 110, 241–263.
Dobos, E., Daroussin, J., Montanarella, L., 2005. A SRTM based-procedure to delineate SOTER Terrain Units on 1:1M and 1:5M scales. European Commission Report, EUR 21571, 55p.
Favrot, J.C., 1981. Pour une approche raisonnée du drainage agricole en France: la méthode des secteurs de référence. C.R. Acad. Agric., 716–723 séance du 6 mai 1981.
Fleiss, J.L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement 33, 613–619.
Fridland, V.M., 1972. In: D.H. Yaalon (Ed.), Pattern of the Soil Cover. Keterpress Enterprises, Jerusalem, translated from Russian by N. Kaner.
Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modeling and spatial prediction of soil attributes. International Journal of Geographical Information Science, 9, 421–432.
Guerin, B., Hardy, R., Duval, O. (unpublished). Carte pédologique à 1/100000 de Laon.
Heuvelink, G.B.M., Edzer J.P., 1999. Spatial Aggregation and soil process modelling. Geoderma 89, 47–65.
Hole, F.D., 1978. An approach to landscape analysis with emphasis on soils. Geoderma 21, 1–23.
IGN, 1998. Descriptif technique BD ALTI. Consulted the 11/10/2005. URL: http://www.ign.fr/telechargement/MPro/produit/BD_ALTI/DT_BDALTI.pdf
Jamagne, M., 1967. Bases et techniques d'une cartographie des sols. Annales agronomiques. Vol 18. Hors série.
Lagacherie, P., Robbez-Masson, J.M., Nguyen-The, N., Barthès, J.P., 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma 101, 105–118.
McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems 50(1), 51–62.
Roque, J. (unpublished). Carte pédologique à 1/100 000 de Meaux.
Smeck, N.E., Runge, E.C.A., McKintosh, E.E., 1983. Dynamics and Genetic Modelling of Soil Systems. Pedogenesis and Soil Taxonomy I: Concepts and Interactions. Elsevier, Amsterdam.

Chapter 18
An Approach to Removing Uncertainties in Nominal Environmental Covariates and Soil Class Maps

T. Behrens, K. Schmidt and T. Scholten

Abstract In this chapter we present an automated approach to correct the delineation of nominal soil and environmental datasets based on auxiliary metric attributes, aiming to enhance positional accuracy. The detection of uncertainties is based on different spatial and non-spatial approaches. The methodological framework mainly consists of nearest neighbour approaches and comprises supervised feature selection, different ensemble classification techniques, as well as spatial and non-spatial smoothing and generalization approaches. The method is described and applied to an artificial dataset as well as a 1:50 000 German soil map and a 1:1 000 000 geological map of the Republic of Niger.

18.1 Introduction

In many situations of applied digital soil mapping we have to handle spatial datasets of varying provenance, age, scale, resolution, mapping scheme, and aggregation level resulting in different sources of errors (Robinson et al., 1984; Lagacherie and Holmes, 1997; Heuvelink, 1998; Bishop et al., 2006) (see also Chapters 6 and 25). In predictive data mining approaches (Behrens et al., 2005; Behrens and Scholten 2006a) existing soil data is extrapolated on the basis of auxiliary environmental datasets (McBratney et al., 2003). Hence, the prediction accuracy can be weak (i) if the soil data and or (ii) the auxiliary datasets contain errors. For example when using a small-scale geological maps (>1:100 000) as predictor datasets for medium or large scale digital soil maps (<1:50 000) the delineation is propagated through the analysis and can be found in the prediction results, assuming there is a significant relation. Thus, in general, maps of smaller scales should not be used to compile maps of larger scales (Robinson et al., 1984) (see also Chapter 17). As "in practice even the best-drawn maps are not perfect" (Burrough

T. Behrens
Department of Physical Geography, Institute of Geography, University of Tübingen, Tübingen, Germany
e-mail: thorsten.behrens@uni-tuebingen.de

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

and McDonnell, 1998) positional inaccuracies can be found in most classically surveyed soil maps for many well known reasons. In this case, the soil map as the training dataset contains noise – which again weakens prediction accuracy (Brighton and Mellish, 2002). Hence, it is important to provide solutions to automatically correct existing datasets. One step towards better predictions – or in some cases to allow predictions at all, when the data is sparse and at small scale – is to provide techniques that correct the boundaries of nominal datasets on the basis of auxiliary datasets of higher resolutions and/or scales. Demonstrated on different artificial and real datasets this Chapter presents automated approaches to adjust the boundaries of nominal datasets based on terrain attributes.

18.2 Rationale

The correction of the delineation in nominal datasets as introduced in this paper is based on two major steps: first, the detection and removal of inaccuracies and second, the prediction of new class values for all incorrect pixels. The detection of positional inaccuracies can be achieved in an unsupervised or a supervised fashion, based on simple band removal approaches or on outliers in terrain attributes found within each class-area. Concerning the outlier based approach digital terrain analysis plays a crucial role. Hence a large library of terrain attributes (Behrens, 2003; Behrens et al., 2005) is used. The prediction of a new class value for an uncertain pixel is based on a spatial and non-spatial nearest neighbour data mining framework, comprising feature selection, ensemble classifications as well as spatial and non spatial smoothing and generalization approaches to provide stable and reliable results.

The whole approach is applied iteratively, as the position of the boundaries and thus the outliers changes after each run. The system is stopped if no further significant or plausible changes occur.

18.3 Test Sites

18.3.1 Artificial Datasets

An artificial DEM of a hemisphere (radius = 40 pixels) set on top of a plain surface (100 by 100 pixels) was used as a test bed for the framework developed here. The corresponding artificial nominal environmental dataset consists of two mapping units: first, an inner, irregular shaped ellipsoid which overlaps large parts of the hemisphere and minor parts outside the hemisphere and second a surrounding mapping unit mainly covering the plain surface surrounding the hemisphere (Fig. 18.1). The aim is to correct the ellipsoid mapping unit in such way that it covers the hemisphere completely and is removed from the surrounding plain surface.

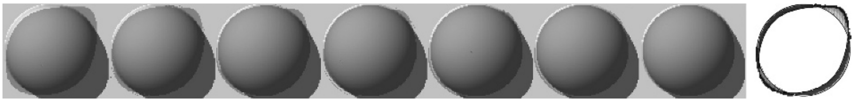


Fig. 18.1 Iterative correction of positional inaccuracies of an ellipsoid mapping unit to an artificial hemisphere DEM. The *left* image is the original ellipsoid followed by images showing the results after 2, 4, 6, 8, 10 and 12 iterations. The image in the *right* shows the location of the corrected pixels

18.3.2 1:50 000 Soil Map of Central Hesse, Germany

The 1:50 000 soil map of central Hesse, Germany shows parts of the Vogelsberg, Europe's biggest shield volcano with a relief of 170–750m asl. The soilscape is mainly characterized by Cambisols, often influenced by loess components (HLUG, 2002). The soil map was rasterized to a resolution of 20 m.

18.3.3 1:1 000 000 Geological Map of the Republic of Niger

To provide an example for countries with sparse datasets, we analyzed the 1:1 000 000 geological map of the Republic of Niger in Western Africa. The map is based on the work of Greigert (1961) and disseminated digitally as part of the "Atlas of Natural and Agronomic Resources of Niger and Benin" (Herrmann et al., 1999). Shuttle Radar Topography Mission (SRTM) data with a resolution of 90 m was used to derive terrain attributes (Section 18.4.1).

18.4 Methods

The technical key steps of the proposed methodology are briefly described in the following:

Beyond the correction of uncertainties the major goal of the methodological framework applied is to provide stable results. In the first step after digital terrain analysis (Section 18.4.1) and dataset creation, the data are analyzed to remove noisy and irrelevant features (attributes) (Section 18.4.2). After reducing dimensionality, the core algorithms, that is, removing uncertain pixels form the dataset, are applied (Section 18.4.3). Afterwards, to speed up computation time we use a stratified random sampling (Section 18.4.5.1) over the resulting dataset followed by Wilson editing to remove noise (Section 18.4.5.2). To allocate each uncertain pixel we apply a simple kNN-classifier (Section 18.4.5.3) to assign the most probable class out of the spatially adjacent neighbours (Section 18.4.4). Finally, to keep the system stable and to avoid blurry borders local spatial noise removing is applied (Section 18.4.6). Additionally, the methods described above are embedded in an ensemble prediction approach (Section 18.4.7) – again to provide accurate results.

18.4.1 Digital Terrain Analysis

As there is a strong dependence between terrain attributes and soil properties (McBratney et al., 2003; Behrens et al., 2005), a large pool of continuous geomorphometric terrain attributes is used in digital soil mapping. Based on a terrain-analysis framework (Behrens, 2003), the following 25 terrain attributes were calculated: flow accumulation, relative hillslope position, elevation above channel, distance to channel, average slope, steepest slope, aspect, profile curvature, planform curvature, mean curvature, maximum curvature, minimum curvature, relative profile curvature, relative planform curvature, topographic roughness, relative richness, waxing and waning slopes, solar insolation, compound topographic index, USLE LS-factor, landform evolution, relative mass balance, stream power index, surface area ratio, and surface volume. For details see Behrens (2003), Behrens et al. (2005), and Behrens and Scholten (2006b).

SRTM data was used to remove uncertainties in the geological map of the Republic of Niger. Additional terrain attributes had to be calculated based on Monte Carlo simulations to derive more natural spatial flow patterns and geomorphometric positions (i.e. flow accumulation, relative hillslope position, and elevation above channel). This was essential for boundary adjustments, due to the error component of the SRTM data which is relatively large in this area (visual interpretation), the scale of the geological map, and the width of the valleys. A comparison

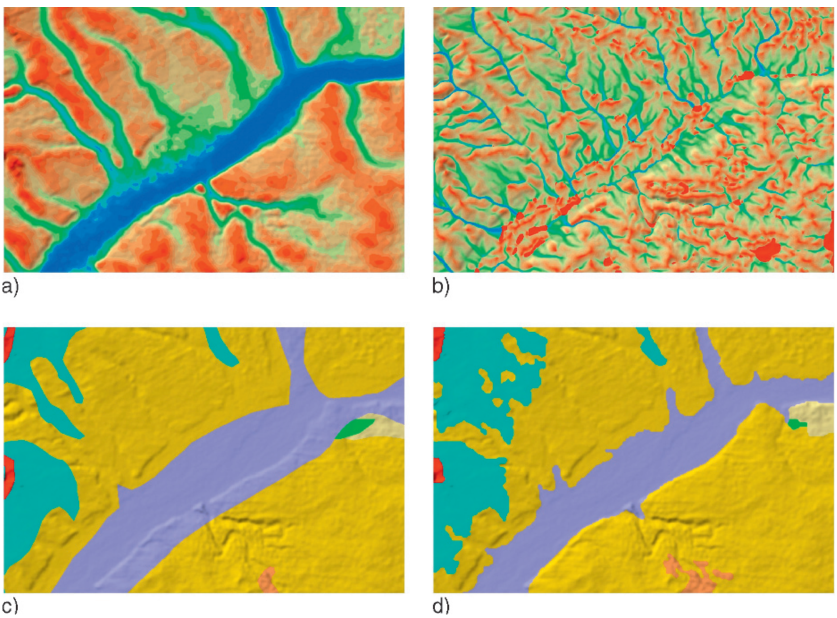


Fig. 18.2 Comparison of flow-accumulation based on a Monte Carlo simulation using a single-flow approach (**a**) and a multiple-flow approach (**b**) for a section of the geological map of the Republic of Niger (**c** = original, **d** = corrected). (See also Plate 20 in the Colour Plate Section)

of a standard multiple-flow algorithm to calculate contributing area (Dietrich and Montgomery, 1998) and a Monte Carlo approach based on the D8 single-flow algorithm (Jenson and Domingue, 1988) is shown in Fig. 18.2a,b. It can be seen that the spatial distribution of the resulting flow accumulations based on the different approaches differs especially in the valley bottoms, where the Monte Carlo based approach produces much more plausible results and models the valley bottom according to the draped hillshade.

18.4.2 Feature Selection

K-nearest neighbour classifiers as instance-based learners are sensitive to correlated as well as irrelevant and noisy features. Thus, feature selection techniques need to be applied to achieve accurate predictions. The feature selection algorithm used in this study is the well known Relief-F approach (Kira and Rendell, 1992; Kononenko, 1994; Liu and Motoda, 1998).

For every class combination and every randomly selected instance (i.e. vector containing terrain attributes) in a dataset the difference between the feature values of the nearest hit (i.e. the shortest instance to the same class) and the nearest miss (i.e. the shortest instance to the neighbouring class) are calculated and summed up over all selected instances in a weight vector. Thus each feature has a weight indicating its potential to differentiate between the classes in a dataset. In this study we used the mean weight as the lower limit to remove features and 50 randomly selected instances per class.

18.4.3 Removing Uncertainties

18.4.3.1 Band Removal

As the probability of noise is generally higher at the polygon boundaries than within a polygon, a simple spatial denoising approach is to remove all pixels at the class boundaries. This idea is based on the concept of “error bands” as introduced by Perkal (1966). As the width of the band can not easily be predicted and is irregular in most cases we use a buffer width of one pixel inside each class.

18.4.3.2 Outlier Detection

Based on the terrain attributes derived (Section 18.4.1) we developed a non-spatial denoising approach which is the initial idea behind this study. Therefore, each class-area of the nominal dataset to be corrected is analyzed separately in terms of outliers within the frequency distribution of each terrain attribute. If a threshold is reached, that is if the majority of all terrain attributes is outside twice the standard deviation, the corresponding pixel is marked as uncertain. Thus, this process is data driven and in contrast to band denoising it is not fixed to the boundary between two polygons.

18.4.4 Spatial Neighborhood Search

To determine the most likely soil class for each uncertain pixel a local spatial neighbourhood search is applied. The advantage of a local search procedure is that the universe of potential classes to be assigned to an uncertain pixel is reduced resulting in predictions that are generally more accurate and stable.

The search for adjacent soil classes is based on the moving window technique. If a pixel contains no soil class information, its neighborhood is analyzed initially on the basis of a three-by-three pixel neighbourhood. If no adjacent soil class is found within this kernel, the neighborhood size is automatically enlarged until at least two adjacent soil classes are found.

18.4.5 Instance Selection and Classification

18.4.5.1 Random Sub-Sampling

Instance selection, or sub-sampling (Liu and Motoda, 2001) aims to remove redundant information from datasets as well as to speed up learning and/or prediction time while preserving prediction accuracy (Schmidt et al., 2008). This becomes important for large datasets with thousands of training samples and for computationally expensive algorithms like *k-nearest neighbour*.

In this study we use stratified random sampling. Kohonen et al. (1995) recommend a disproportional approach for supervised classification applications, where an equal amount of observations or instances is selected for each class, even when the *a-priori* probabilities differ strongly.

The sample size for each class in this study is 50. As this process is embedded in the ensemble prediction approach (Section 18.4.7) random sampling is the subagging (Breiman, 1996; Bühlmann and Yu, 2002) part of the procedure.

18.4.5.2 Dataset Editing

Wilson editing (Wilson, 1972) is a competitive supervised denoising technique (Zeidat et al., 2005) with the goal of obtaining more accurate classifiers. This is achieved by removing all vectors in a dataset that have been misclassified by a k-nearest neighbour classifier (Section 18.4.5.3), leading to smoother class boundaries in the feature space and better subsequent classification results using a k-nearest neighbour classifier. In our case this is done separately for every class combination found within the neighborhood search to get optimized and spatially dependent classification results, which is an advantage in heterogeneous soilscapes.

18.4.5.3 Supervised Classification

The k-nearest neighbour classifier (Fix and Hodges, 1951) labels an unknown instance with the class label of the majority of its k-nearest neighbours in terms of

Euclidean distance in feature space. We use a three-nearest neighbour classifier in this study.

18.4.6 Generalization Using Spatial Noise Removal

To avoid blurry borders and small isolated areas a spatial noise removal is applied after each iteration. The noise removal approach replaces all areas comprising less than five pixels followed by a three-by-three pixel majority filter. An optimized size for areas to be removed can not be determined a-priori and has to be tested iteratively. We suggest a default value of less than 5 pixels (which is less than the half amount of pixels in a local 3*3 three-by-three neighbourhood), as it produces only a weak smoothing effect. As this process avoids fuzzy transition zones due to smoothing it is comparable to the soil scientist's approach of generalization when mapping soil classes.

18.4.7 Ensemble Prediction

Ensemble approaches, i.e. combinations of multiple predictions based on changes in the training dataset, are very popular and powerful, as they increase prediction accuracy (Breiman, 1996). For k-nearest neighbour classifiers, feature subset selection approaches are recommended by a number of authors (e.g., Bay, 1999; Akkus and Güvenir, 1996) and are competitive with boosted (Freund and Schapire, 1996) decision trees (Bay, 1999). The application of instance-based ensembles like bagging or subagging (Breiman, 1996; Andonova et al., 2002; Bühlmann and Yu, 2002) is reported to work on small random samples (Alpaydin, 1997; Hamamoto et al., 1997). Additionally, ensemble approaches are more robust against irrelevant and correlated features (Bay, 1999; Skurichina and Duin, 2001). In this chapter we apply a combination of both random feature subsets (Ho, 1998) as well as small sample subagging (Section 18.4.5.1), which is comparable to decision forest approaches (Ho, 1998; Breiman, 2001).

18.5 Results and Discussion

The aim of applying the proposed approach on an artificial dataset was to visualize the results for an easy-to-interpret example. As shown in Fig. 18.1 the method works as expected. Based on the band removal approach 12 iterations were needed to fit the nominal ellipsoid to the hemisphere (circle) based on 5 terrain attributes (slope, compound topographic index, mean, profile and horizontal curvatures). Using the outlier detection and a low threshold for outlier removal, only one iteration is needed to achieve the same results. Yet in this case band removal offers the opportunity to analyze the correction process and thus the location of spatial uncertainties over the iterations.

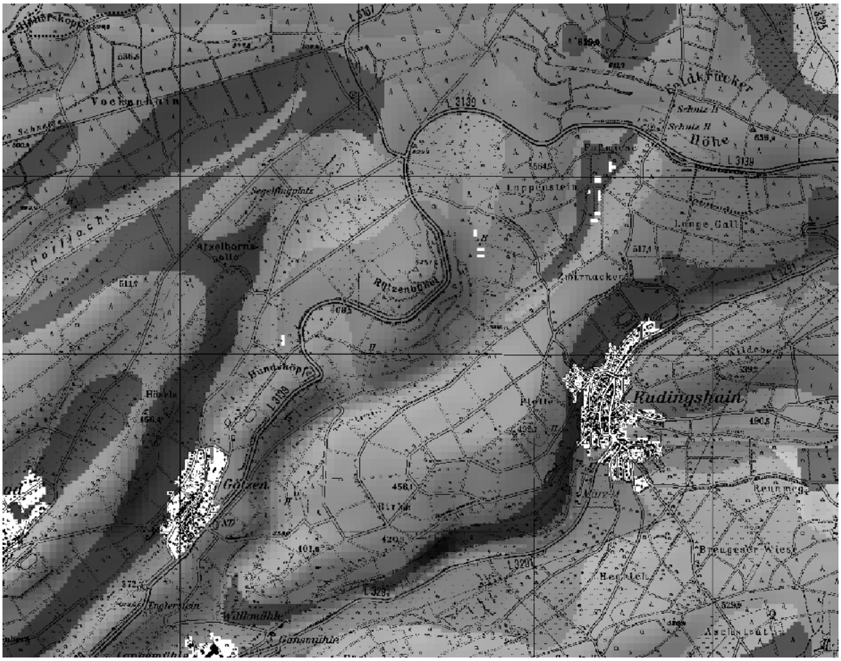


Fig. 18.3 Section of the original 1:50 000 soil map (Central Hesse, Germany) draped over a DEM

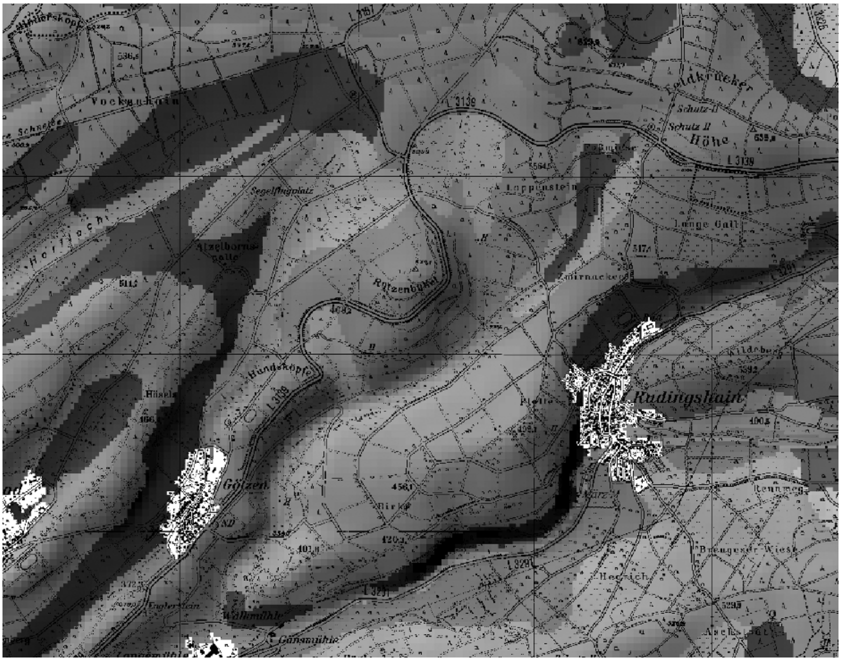


Fig. 18.4 Corrected 1:50 000 soil map (Central Hesse, Germany) draped over a DEM

The correction of the 1:50 000 soil map of Central Hesse, Germany is based on the outlier detection approach (Section 18.4.3.2). A first field survey revealed four iterations to be sufficient to correct the soil map. At first sight, only minor changes can be found between the original soil map (Fig. 18.3) and the corrected soil map (Fig. 18.4). Yet, some boundaries show differences of 80 m. Generally, using a finer resolution more iterations would be required. Depending on the resolution and the scale of the input datasets different resolutions need to be tested to achieve optimized results.

The section of the 1:1 000 000 geological map of the Republic of Niger as shown in Fig. 18.2c demonstrates the problem of small-scale nominal maps in relation to more precise, in terms of positional accuracy, data like SRTM. For the mapped floodplain in the centre of the section, 33% of the area is located outside the floodplain as shown by the SRTM DEM. Thus, automated-data-mining-based boundary adjustments become critical due to large amounts of noise. Yet, 12 iterations based on the outlier approach were sufficient to correct the map (Fig. 18.2c,d). Attempts with standard terrain attributes only, were not successful, whereas a combination with the Monte Carlo based approaches returned reasonable results. In the case of some of the other mapping units which do not show a strong relation to relief, we recommend not using the adjusted polygons in the final map.

Based on the three datasets tested during the development of the approach presented the technique works as expected. Yet, it is not possible to make assumptions about the quality of the boundary adjustment a priori as it depends on the errors and interactions of the datasets used. Generally, three approaches to test the accuracy of the method seem reasonable: (i) direct field validations, (ii) indirect comparisons between digital soil maps based on the original and on the corrected datasets, and (iii) expert interpretations based on map overlays using hillshades or 3D visualisations. Concerning the 1:50 000 soil map of Hesse all 3 approaches are currently tested and compared for different landscapes.

Concerning different methodological techniques (iterations, removal techniques) the general tendency is that more iterations are needed if the uncertainties are high or the resolution of the terrain attributes is relatively fine compared to the scale of the map to correct. Further research is needed on criteria when to stop the iterative process to provide fully automated applications. Band removal needs more iterations than outlier based removal. Additionally, the risk to change boundaries not related to relief is higher. Finally, for geomorphologic settings like wide valleys as found in the geological map of Niger, plausible results can only be achieved when carefully selected special terrain attributes are used in the correction approach (see also Chapters 1 and 28).

18.6 Conclusions

The approach introduced in this study helps to enhance positional accuracy of nominal soil and environmental datasets. This is important:

– in terms of error propagation if datasets of different scales and resolutions are used together in *scorpan* (McBratney et al., 2003) predictions,
– to produce better prediction results due to an enhancement of the soil map to predict and/or the environmental covariates used for prediction,
– and in terms of quality control of conventionally produced soil maps.

The second step of the procedure – the prediction of the most probable class – can also be used as a post-processing tool after predicting single soil units which often results in overlapping areas and gaps (Behrens et al., 2005).

Future research is needed to find optimized settings for the different model parameters like the feature selection threshold, the removal approach, the subsample size, the number of neighbours used for classification, the settings for spatial smoothing, and the iterations. Yet, as using the default values returns promising and stable results the major model parameters are the number of iterations and the algorithm to remove noise. The principal advantage of the outlier-based removal compared to the error-band removal is that pixels at class boundaries that are not related to the environmental covariates used, are not per se regarded as uncertain, which preserves these boundaries in their original shape. Hence, outlier-based removal is recommended.

The method, proposed and described here for the first time, may become an important pre- and post-processing tool in digital soil mapping. In countries with sparse and coarse soil information it might help to refine maps of soil and its environmental covariates.

Acknowledgments The authors would like to gratefully acknowledge the critical and helpful comments and suggestions received from Alex McBratney on this manuscript. Partial funding for this research was provided by the Collaborative Research Centre 299 of the German Research Foundation. We would like to thank the Hessian State Office for Environment and Geology (HLUG) for providing data.

References

Andonova, S., Elisseeff, A., Evgeniou, T., Pontil, M., 2002. A simple algorithm for learning stable machines. In: F. van Harmelen (Ed.), Proceedings of the Fifteenth Eureopean Conference on Artificial Intelligence. IOS Press, Amsterdam. pp. 513–517.

Akkus, A., Güvenir, H.A., 1996. K nearest neighbor classification on feature projections. In: L. Saitta (Ed.), Machine Learning: Proceedings of the Thirteenth International Conference. Morgan Kaufmann, San Francisco, CA. pp. 12–19.

Alpaydin, E., 1997. Voting over multiple condensed nearest neighbors. Artificial Intelligence Review 11, 115–132.

Bay, S.D., 1999. Nearest neighbor classification from multiple feature subsets. Intelligent Data Analysis 3, 191–209.

Behrens, T., 2003. Digitale Reliefanalyse als Basis von Boden-Landschafts-Modellen – am Beispiel der Modellierung periglaziärer Lagen. Boden und Landschaft 42, 189p. Giessen.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. Journal of Plant Nutrition and Soil Science 168, 21–33.

Behrens, T., Scholten, T., 2006a. Digital soil mapping in Germany – a review. Journal of Plant Nutrition and Soil Science 169, 434–443.

Behrens, T., Scholten, T., 2006b. Chapter 25. A comparison of data-mining techniques in predictive soil mapping. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science, Vol. 31. Elsevier, Amsterdam. pp. 353–364.

Bishop, T.F.A., Minasny, B., McBratney, A.B., 2006. Uncertainty analysis for soil-terrain models. International Journal of Geographical Information Science 20, 117–134.

Breiman, L., 1996. Bagging predictors. Machine Learning 24, 123–140.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Brighton, H., Mellish, C., 2002. Advances in instance selection for instance-based learning algorithms. Data Mining and Knowledge Discovery 6, 153–172.

Bühlmann, P., Yu, B., 2002. Analyzing bagging. Annals of Statistics 30, 927–961.

Burrough, P.A., McDonnell, R.A., 1998. Principles of Geographical Information Systems. Oxford University Press, New York.

Dietrich, W.E., Montgomery, D.R., 1998. SHALSTAB: a digital terrain model for mapping shallow landslide potential, NCASI (National Council of the Paper Industry for Air and Stream Improvement), Corvallis OR. Technical Report, 29p.

Fix, E., Hodges, J.L., 1951. Discriminatory analysis – nonparametric discrimination: consistency properties. Technical Report Project 21-49-004, Report Number 4, USAF School of Aviation Medicine, Randolf Field, Texas.

Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm, In: L. Saitta (Ed.), Machine Learning: Proceedings of the Thirteenth International Conference. Morgan Kaufmann, San Francisco, CA, pp. 148–156.

Greigert, J., 1961. République du Niger. Carte géologique de reconnaissance du Bassin des Iullemeden 1:1 Mio. BRGM, Niamey, Niger.

Hamamoto, Y., Uchimura, S., Tomita, S., 1997. A bootstrap technique for nearest neighbor classifier design. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 73–79.

Herrmann, L., Vennemann, K., Stahr, K., von Oppen, M., 1999. Atlas of Natural and Agronomic Resources of Niger and Benin. http://www.uni-hohenheim.de/~atlas308/startpages/page2/english/content/title_en.htm.

Heuvelink, G.B.M., 1998, Error Propagation in Environmental Modelling with GIS. Taylor & Francis, London, 144p.

HLUG (Hessische Landesamt für Umwelt und Geologie), 2002. Erläuterungen zur Bodenkarte von Hessen 1:50000. HLUG, 575p.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 832–844.

Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for Geographic Information System analysis. Photogrammetric Engineering and Remote Sensing 54, 1593–1600.

Kira, K., Rendell, L., 1992. A practical approach to feature selection. In D.H. Sleeman, P. Edwards (Eds.), Proceedings of the Ninth International Conference on Machine Learning. Morgan Kaufmann, San Francisco, CA, pp. 249–256.

Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J., 1995. The Learning Vector Quantization Program Package, Version 3.1. http://www.cis.hut.fi/research/lvq_pak/.

Kononenko, I., 1994. Estimating attributes: analysis and extensions of RELIEF. In: L. De Raedt, F. Bergadano, (Eds.), European Conference on Machine Learning. Springer, Heidelberg, pp. 171–182.

Lagacherie, P., Holmes, S., 1997. Addressing geographical data errors in a classification tree soil unit prediction. International Journal of Geographical Information Science 11, 183–198.

Liu, H., Motoda, H., 1998. Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic Publishers, Norwell, MA, 244p.

Liu, H., Motoda, H., 2001. Instance Selection and Construction for Data Mining. Kluwer Academic Publishers, Norwell, MA, 448p.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Perkal, J., 1966. An Attempt at Objective Generalization. Geodézia es Kartográfia, 7(2), 130–142.

Robinson, A.H., Sale, R.D., Morrison, J.L., Muehrcke, P.C., 1984. Elements of Cartography, 5th edition. John Wiley & Sons, New York.

Schmidt, K., Behrens, T., Scholten, T., 2008. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. Geoderma (accepted).

Skurichina, M., Duin, R.P.W., 2001. Bagging and the random subspace method for redundant feature spaces. In: J. Kittler, F. Roli (Eds.), Multiple Classifier Systems, Proceedings Second International Workshop MCS 2001 (Cambridge, UK, July), Lecture Notes in Computer Science, Vol. 2096. Springer Verlag, Berlin, pp. 1–10.

Wilson, D.L., 1972. Asymptotic properties of nearest neighbour rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics 2, 408–420.

Zeidat, N., Wang, S., Eick, C., 2005. Dataset Editing Techniques: A Comparative Study: Department of Computer Science, University of Houston. http://www.cse.unr.edu/∼bebis/cs679/PaperPresentations/ZWE05.pdf.

Chapter 19
Digital Soil Mapping Using Logistic Regression on Terrain Parameters for Several Ecological Regions in Southern Brazil

E. Giasson, S.R. Figueiredo, C.G. Tornquist and R.T. Clarke

Abstract As the relationship between soils and landscape within the context of soil formation is well known, predictive relationships between soils and soil formation factors can be established by regression techniques, relating soil and terrain attributes to occurrence of soil classes. This study proposes the production of maps using logistic regression on soil and terrain information from a pilot area to reproduce the original map and predict soil distribution in other similar landscapes in three study areas (Ibibubá Municipality, Sentinela do Sul Municipality, and Arroio Portão Watershed) in map scales from 1:30,000 to 1:50,000 and located in three ecological regions in Southern Brazil (Planalto, Encosta da Serra do Sudeste, and Depressão Central, respectively). By using logistic regressions for digital soil mapping, the method predicts the occurrence of soil units based on reference soil maps (produced by conventional methods), and on several parameters derived from a USGS SDTS-SRTM DEM, namely slope gradient, profile curvature, planar curvature, curvature, flow direction, flow accumulation, flow length, Stream Power Index (SPI), and Topographic Wetness Index (TWI). Results show that parameters such as elevation, curvature, SPI, TWI, and distance to streams are more frequently selected as parameters for predicting the occurrence of soil classes, with overall percent correct from 61% to 71%, and Kappa Index from 36% to 54% when the maps produced are compared with the original soil maps with a simplified legend (which simulate the production of soil maps with smaller scales that the original soil map). The prediction of soil map units using logistic regressions generated reliable soil maps, and the method appears to deserve more research effort, given the reliability and low cost of the resulting information.

E. Giasson
Department of Soils of the Federal University of Rio Grande do Sul (UFRGS). Av. Bento Gonçalves, 7712, Porto Alegre, RS, Brazil, 90.001-970
e-mail: giasson@ufrgs.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

19.1 Introduction

A cost-effective approach to traditional large-scale soil surveys would be to map soils of representative areas within homogeneous regions and use the soil-landscape relationships to predict soil distribution on non-surveyed areas. This approach is similar to the reference area method (Lagacherie et al., 2001), which is based on the hypothesis that it is possible to sample a reference area including most of the soil classes of a region. Based on this reference area, the prediction of soil distribution on other areas may be facilitated if the landscape is modeled by digital terrain analysis (Hengl and Rossiter, 2003) and if relationships between soils and landscape are modeled. Recently, several approaches to make digital soil maps based on GIS have been used (Moore et al., 2001; Bell et al., 1994; McKenzie and Ryan, 1999; Odeh et al., 1994; Zhu and Band, 1994; Lark, 1999; Gessler et al., 2000; Campling et al., 2002).

Multiple logistic regression has been used successfully in soil science and many other related fields (Camplig et al., 2002; Gurdak, 2006; Mueller et al., 2005; Ohlmacher and Davis, 2003; Lai et al., 2006; Wang et al., 2007), see also Chapters 9, 17, 19, 20, and 25 in this book. As soil map units are categorical variables, multiple logistic regressions may be suitable for predicting occurrence of soil classes from landscape variables, with the advantage of providing estimates of probabilities of occurrence of soil class map units. Although previous works used logistic regression to estimate the occurrence of specific soil characteristics instead of soil taxonomic classes or mapping units, they suggest that logistic regressions may have potential for producing soil maps from terrain parameters based on relationships between these parameters and soil occurrence. The purpose of the study reported here was to test a method of extrapolating landscape parameters by testing how well multiple logistic regressions can reproduce a soil map of a reference area.

19.2 Material and Methods

The study was conducted in three ecological regions of Rio Grande do Sul State, Brazil. The first study area was in the municipality of Ibirubá (720 km^2), where original soil survey at scale 1:30,000 (Santos et al., 1970) identified deep and well drained Oxisols in the nearly flat lands at higher elevation, Entisols and Molisols on hill-slopes, and poorly drained soils in flat and low lands close to streams. The original soil classification of this and following surveys was updated to Soil Taxonomy (United States Department of Agriculture, 1998).

The second study area was the Arroio Portão watershed (225 km^2), with Endoaquents and Endoaqults in flat lands, Udarents on hill-slopes, and Hapludults in gently sloping areas, as identified by Klamt and Schneider (1992) in a 1:50,000 soil survey.

The third study area was the municipality of Sentinela do Sul (253 km^2). A soil survey of the municipality at scale 1:50,000 (Klamt et al., 1996) identified map units composed mainly of combinations (undifferentiated group, complex, or association) of soil taxonomic class units dominated by Endoaquents in flat lands in the alluvial plains, by Inceptisols and Ultisols in gently sloping lands, by Inceptisols and by Udarents in strongly sloping lands, and Udarents in very steep lands (Giasson et al., 2006).

To improve the capacity to reproduce the original soil maps, predictions were done both using the original soil map legend and using a simplified legend. The legends were simplified by grouping similar mapping units based on their higher taxonomic categorical level. This was intended to verify whether a simplified categorical legend, which would usually correspond to a cartographic scale reduction, could be better predicted. In both situations, the same procedures were used for estimating multiple logistic regressions and evaluating map accuracy.

The digital elevation model (DEM) was the 3 arc sec or 92 m (USGS SDTS – SRTM) (Rabus et al., 2003). As in Chapter 32, the DEM was used, directly or as a component, to calculate nine other soil predictor variables: slope gradient, profile curvature, planar curvature, curvature, flow direction, flow accumulation, flow length, Stream Power Index (SPI), and Topographic Wetness Index (TWI) (Wolock and McCabe, 1995). Each of these landform parameters was selected for test as explanatory variable because they were expected to represent changes on soil-forming factors and, therefore, were believed to contain information on the occurrence of soil mapping units.

Logistic regressions were used to determine the relationships between these explanatory variables and soil distribution. Data sampling for training points consisted of random map observations (approximately one observation for each 3.5 ha) consisting of the digital elevation model (DEM), DEM-derived parameters, and soil classes as classified by traditional soil survey. Data sampled in ArcView 3.2 environment (ESRI,1999) was exported and analyzed statistically using Minitab version 11 (Minitab Inc., 1996). A step-by-step procedure was used to obtain the best fit set of logistic regressions, starting with a larger number of variables and excluding the variables considered less related to variations on the response variable. Sets of best fit regressions were selected based on criteria as goodness to fit tests (Pearson and deviance), log-likelihood, odds ratio, and Z test (Hosmer and Lemeshow, 1989). These equations were organized as Avenue scripts in ArcView 3.2 environment, assigning a probability value to each pixel and creating maps representing probability of occurrence of soil mapping units. Estimated soil class maps were elaborated by assigning to every single pixel the denomination of the soil mapping unit that had the larger probability of occurrence for that pixel.

The accuracy of the estimated soil maps was determined by using error matrices (Congalton, 1991) comparing pixel by pixel if the soil map unit as estimated agreed with the soil map unit determined by the original soil survey. Based on this error matrices, several map accuracy indicators and the overall accuracy and the Cohen's Kappa index (Cohen, 1960) were calculated.

19.3 Results and Discussion

The final maps produced can be compared with original maps (Fig. 19.1). The variables more frequently selected and therefore taken as predictors of occurrence of soil mapping units were elevation, distance to streams, TWI, curvature, and slope

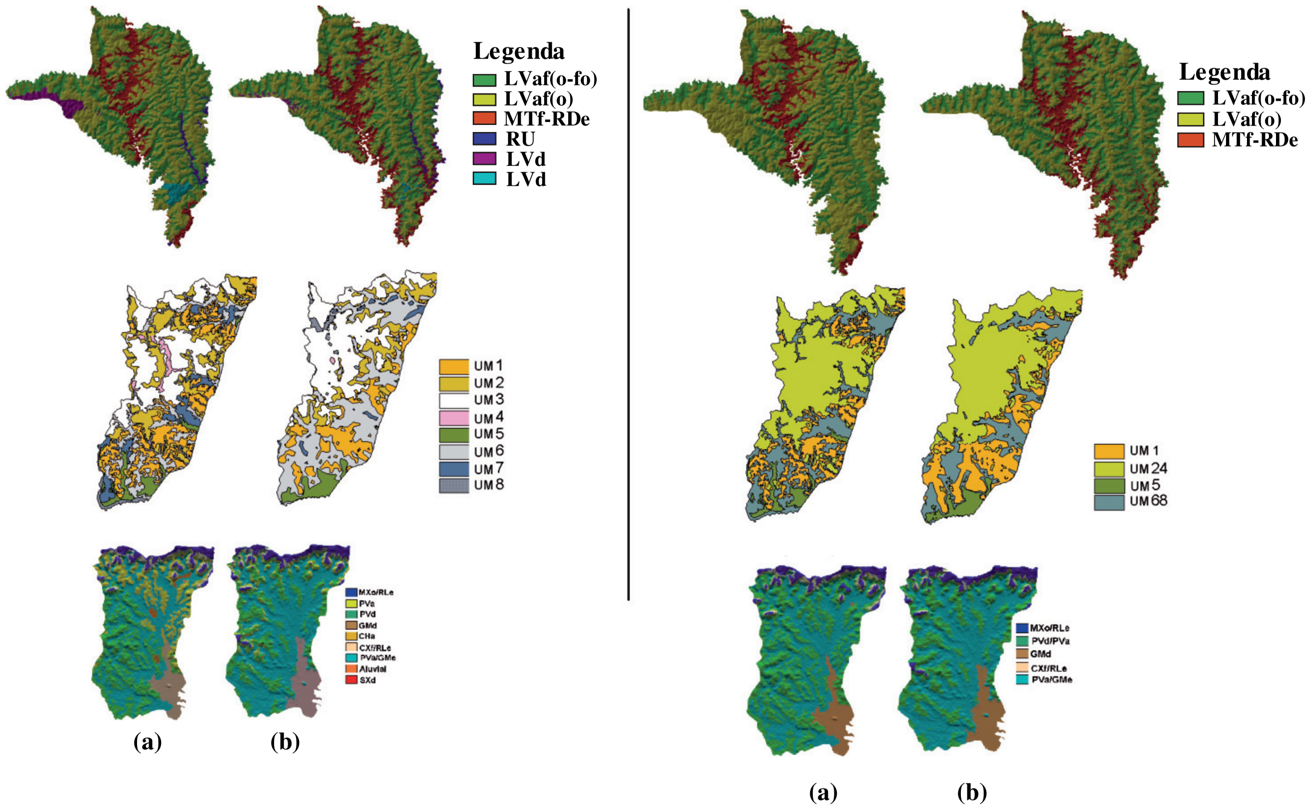


Fig. 19.1 Original soil maps (**a**) and respective estimated soil maps (**b**) using the original soil survey legend (left side of plate) and taxonomic simplifications of the original soil survey legend (right side of plate) for Ibirubá Municipality (up of plate), Sentinela do Sul Municipality (middle of plate), and Arroio Portão Watershed (bottom ○ (See also Plate 21 in the Colour Plate Section)

Table 19.1 Variables selected as predictors of occurrence of soil mapping units ($p<0.001$) for the three study areas in Rio Grande do Sul, Brazil

	Study area		
	Ibirubá	Arroio Portão	Sentinela do Sul
Predictor	elevation distance to streams TWI curvature SPI flow length slope	elevation distance to streams curvature slope	elevation distance to streams TWI

TWI = Topographic Wetness Index and SPI = Stream Power Index

(Table 19.1), which are variables related to water accumulation and water dynamics. These sets of best predictors are related to long-term known relationships between soil-forming factors, landform, and soil distribution, which relate soil distribution to erosion processes in steep well-drained areas and with water dynamics, such as water table depth (variable elevation) and water movement and accumulation in poorly drained areas.

For evaluating the reproducibility of the original soil map, the agreement between these maps and the newly generated maps was evaluated using error matrices, and its accuracy quantified in Table 19.2.

Attempts to classify the entire landscape never achieved better results than 59% overall accuracy (Kappa Index = 42%) when using the original legend. Overall accuracy and Kappa Index were considered unsatisfactory in both cases, although they are in the same magnitude that values found by Hengl and Rossiter (2003). The indicators with lower precision found for Sentinela do Sul study area can be explained by a presence of complexes as soil mapping units, and a large variation on parent material characteristics, not included in this study because of lack of adequate information (Klamt et al., 1996).

Given the characteristics of the SRTM DEM, its precision may not have reproduced small variations of elevation and its resolution may not have showed the actual terrain variations at short distance, so that these characteristics may have contributed to the low resolution of the maps produced. The use of finer resolution DEM (as used in Chapter 28 and 31) could improve the resolution and precision of estimated maps.

Table 19.2 Map accuracy indicators for the three study areas in Rio Grande do Sul, Brazil

	Study area			
Type of legend	Indicator	Ibirubá	Arroio Portão	Sentinela do Sul
Original	Overall Accuracy (%)	58	59	48
	Kappa Index (%)	37	42	36
Simplified	Overall Accuracy (%)	61	68	71
	Kappa Index (%)	38	50	54

Maps estimated using the simplified legend (Fig. 19.1) had an overall accuracy of 61% to 71% (Kappa Index = 38% to 54%) (Table 19.2), which is a mean increase of 23% in the overall accuracy and of 25% in the Kappa Index in relation to when using the original legend. The increases in overall accuracy and Kappa Index were larger in Sentinela do Sul (48% and 50%, respectively), followed by Arroio Portão watershed (15% and 21%, respectively), and by the Ibirubá region (5% for both indicators). Larger increases in overall accuracy and Kappa Index seem to be related to the areas where the changes in soil distribution are more closely related to topography, since Sentinela do Sul has greater variability in relief and since in Ibirubá the relief pattern is more uniform and changes in slope are smaller, being variations in soil type probably more related to changes in parent material constitution.

Although the use of a simplified legend causes the soil map to lose precision (more soil classes are included in a map unit), it increases the accuracy of the digital soil map, i.e., the ability to reproduce a reference soil map, either using an original field survey or a map with simplified legend. Therefore, one must be able to choose between precision and accuracy for selecting the appropriate procedure for its objectives.

19.4 Conclusion

Maps generated using this procedure may be adequate for extrapolating soil distribution information to areas where no comprehensive soil map is available, but where reference soil maps representing soil diversity and distribution in such regions do exist.

Given the importance of soil surveys as sources of information for land use planning and management, countries such as Brazil where soil surveys are most of available at small scales (1:750,000) can benefit from using such digital soil mapping procedures. In Rio Grande do Sul State, the southern state in Brazil, most ecological regions have at least one larger scale soil survey, such as those used in this study. The acceptance of these digital soil maps as useful sources of information may refine the available soil distribution information.

For these study areas there were no comprehensive comparison of the precision of digital soil maps with the precision of traditionally produced soil maps using ground truth procedures.

Currently, more research effort is needed to evaluate costs, benefits, precision, and accuracy of digital soil maps, which may in future come to have precision as good as, or better than, other soil maps.

At this time, digital soil mapping is a growing and promising field of theoretical and applied research, capable of produce applicable knowledge and technology for soil mapping.

References

Bell, J.C., Cutler, C.A., Thompson, J.A. 1994. Soil-terrain modeling for site-specific agricultural management. In Site Specific Management for Agricultural Systems, P. C. Robert, R.H. Rust, W.E. Larson, Eds. ASA-CSSA-SSSA, Madison, WI, pp. 209–228.

Campling, P., Gobin, A., Feyen, J. 2002. Logistic modeling to spatially predict the probability of soil drainage classes. Soil Science Society of America journal 66:1390-1401.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Congalton, R.G. 1991. A review of assessing the accuracy of classification of remotely sensed data. Remote Sensing of Environment 37, 35–46.

ESRI. 1999. Environmental Systems Research Institute. ArcView 3.2. Redland, California.

Gessler, P.E. Chadwick, O.A., Chamron, F., Holmes, K. Althouse, L. 2000. Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Science Society of America Journal 64, 2046–2056.

Giasson, E., Clarke, R.T., Inda Junior, A.V., Merten, G.H., Tornquist, C.G. 2006. Digital soil mapping using multiple logistic regression on terrain paramenters in Southern Brazil. Sci Agric 63, 262–268.

Gurdak, J.J. 2006. Using logistic regression to assess regional ground-water vulnerability; High Plains aquifer. 5th National Water-Quality Monitoring Council meeting, May 7–11, 2006, San Jose, CA.

Hengl, T., Rossiter, D.G. 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Science Society of America Journal 67, 1810–1822.

Hosmer, D.W., Lemeshow, S., 1989. Applied logistic regression. New York: John Wiley & Sons, 307p.

Klamt, E., Schneider, P. 1992. Solos da Bacia do Arroio Portão: características, distribuição e aptidão de uso. 115p. Porto Alegre: Departamento de Solos, UFRGS.

Klamt, E., Schneider, P., Kämpf, N., Giasson, E., Bastos, C.A.B., Lima, V.S., Grehs, S.A., Hesseln, N.E. 1996. Levantamento de reconhecimento de alta intensidade dos solos do município de Sentinela do Sul, RS. Porto Alegre: Departamento de Solos, UFRGS.

Lark, R.M. 1999. Soil-landform relationships at within-field scales: na investigation using continuous classification. Geoderma 92, 141–165.

Lagacherie, P., Robbez-Masson, J., Nguyen-The, N., Barths, J., 2001. Mapping of reference are a representativity using a mathematical soilscape distance. Geoderma 101, 105–118.

Lai, S.Y., Chang, W.J., Lin, P.S. 2006. Logistic regression model for evaluating soil liquefaction probability using CPT data. Journal of Geotechnical and Geoenvironmental Engineering 132(6), 694–704.

MINITAB Inc., 1996. Minitab Inc. Minitab User's Guide. Release 11 for Windows. Minitab Inc.

Moore, R.B., Grayson, A.R., Ladson, A.R. 2001. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 3:3–30.

McKenzie, N.J., Ryan, P.J. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.

Mueller, T.G., Cetin, H., Fleming, R.A., Dillon, C.R., Karathanasis, A.D. and Shearer, S.A. 2005. Erosion probability maps: calibrating precision agriculture data with soil surveys using logistic regression. Journal of Soil and Water 60(6), 462–467

Odeh, I.O.A., Gessler, P.E., McKenzie, N.J., McBratney, A.B. 1994. Using attributes derived from digital elevation models for spatial prediction of soil properties. In Resource Technology '94, Melbourne, Australia, pp. 451–463.

Ohlmacher, G.C., Davis, J.C. 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. Engineering Geology 69, 331–343.

Rabus, B., Eineder, M., Roth, A., Bamler, R. 2003. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. ISPRS. Journal of Photogrammetry and Remote Sensing 57, 241–262.

Santos, M.C.L., Klamt, E., Kampf, N., Abrão, P.U.R., Azolim, M.A. 1970. Levantamento eutilização dos solos do município de Ibirubá. Porto Alegre: INCRA – Sec. Agricultura-RS-DS-FA-UFRGS.

United States Department of Agriculture, Natural Resources Conservation Service. 1998. Keys to Soil Taxonomy, Eighth Edition. Soil Survey Staff.

Wolock, D.M., McCabe, G.J. 1995. Comparison of single and multiple flow-direction algorithms for computing topographic parameters in TOPMODEL: Water Resources Research 31, 1315–1324.

Wang, H.B., Sassa, K., Xu, W.Y. 2007. Assessment of landslide susceptibility using multivariate logistic regression: A case study in southern Japan. Environmental and Engineering Geoscience 13(2), 183–192.

Zhu, A.X., Band, L.E. 1994. A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing 20, 408–418.

Chapter 20
Purposive Sampling for Digital Soil Mapping for Areas with Limited Data

A. Xing Zhu, Lin Yang, Baolin Li, Chengzhi Qin, Edward English, James E. Burt and Chenghu Zhou

Abstract Digital soil mapping requires two basic pieces of information: spatial information on the environmental conditions which co-vary with the soil conditions and the information on relationship between the set of environment covariates and soil conditions. The former falls into the category of GIS/remote sensing analysis. The latter is often obtained through extensive field sampling. Extensive field sampling is very labor intensive and costly. It is particularly problematic for areas with limited data. This chapter explores a purposive sampling approach to improve the efficiency of field sampling for digital soil mapping. We believe that unique soil conditions (soil types or soil properties) can be associated with unique combination (configuration) of environmental conditions. We used the fuzzy *c*-means classification to identify these unique combinations and their spatial locations. Field sampling efforts were then allocated to investigate the soil at the typical locations of these combinations for establishing the relationships between soil conditions and environmental conditions. The established relationships were then used to map the spatial distribution of soil conditions. A case study in China using this approach showed that this approach was effective for digital soil mapping with limited data.

20.1 Introduction

Digital soil mapping often takes a predictive approach based on the classic concept that soil is a function of its environment factors (Jenny, 1941, 1980; Hudson, 1992). Recently, McBratney et al., (2003) developed this concept further and formulated the *scorpan* model which can be expressed as following equations:

$$\begin{aligned} S_c &= f(s, c, o, r, p, a, n) \\ S_a &= f(s, c, o, r, p, a, n) \end{aligned} \tag{20.1}$$

A.X. Zhu
State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China; Department of Geography, University of Wisconsin, Madison, Wisconsin, USA
e-mail: axing@lreis.ac.cn, azhu@wisc.edu

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

where S_c is soil classes and S_a is soil attributes (properties). The *s* in Equation (20.1) refers to existing soil information, *c* is climate condition at the site, *o* means organism (vegetation, fauna or human activities), *r* is local topography (such as elevation, slope gradient, profile curvature, contour curvature, and topographic wetness index), *p* is parent materials, *a* means age, *n* is space and can be perceived as spatial topology or spatial relationship. These factors are referred to as environment covariates. Information on many of these covariates can be derived using GIS and remote sensing techniques. For example, information on topography is typically derived through digital terrain analysis (Evans, 1972, 1998; Wilson and Gallant, 2000; Shary et al., 2002, Section 28.2.2) while information on vegetation is often derived from analyzing remotely sensed imagery (Lillesand and Kiefer, 2000; Section 16.2; Table 2.1).

The *f* in Equation (20.1) refers to the relationship between soil conditions (either soil class or soil properties) and the covariates (Section 1.2). The relationship is of two kinds: one is in the spatial domain and the other is in the parameter domain. The spatial domain relationship explores the spatial auto-correlation of the soil attribute of interest and/or the cross-correlation between the soil attribute and environment covariates. Examples of this type of relationship are "Soil Type A occurs down slope from Soil Type B"; "Soil Type C occurs on slide-slopes below narrow ridges". Another example is the statement that the difference in attribute between two points is a function of the distance between the two points in space (the semivariogram concept). Geostatistical techniques are often used for extracting the latter form of spatial relationship (Burgess and Webster, 1980a,b; Lark and Papritz, 2003; Li et al., 2004; Li et al., 2005; McBratney and Webster, 1983; Odeh et al., 1995; Walvoort and de Gruijter, 2001; Webster, 1991; Webster and Burgess, 1980). However, these methods are typically "data hungry", thus are of limited application for areas with limited data.

The parameter domain relationship refers to the relationships between soil conditions and the first five covariates in the *scorpan* model. An example of such relationship is the relationship between soil properties and terrain variables. Many methods have been developed to extract or determine these relationships (see McBratney et al., 2000; McBratney et al., 2003 for overview). These techniques can be grouped into four major types based on data sources: (1) methods for obtaining knowledge on relationship from local soil scientists (Zhu, 1999; Zhu et al., 2001); (2) methods for establishing relationships from field samples (Bell et al., 1992; Bell et al., 1994; Gessler et al., 2000; Zhu, 2000); (3) methods from discovering relationships from existing soil maps (Bui et al., 1999; Moran and Bui, 2002, Qi and Zhu, 2003; also see Chapter 25); and (4) methods for extracting relationships from typical pedons (typical cases) (Qi et al., 2006; Shi et al., 2004). For areas with limited data, most likely the required data sets for these techniques would not be available, thus these methods may not be applicable as well.

This paper describes a purposive sampling approach for defining the parameter domain relationships. The objective of this approach is to direct field investigation efforts to locations which are expected to capture the spatial pattern of soil variation effectively, thus reduce field sampling efforts and make digital soil mapping for areas with limited data is not only possible but also more effective. Section 20.2 describes the approach, which is followed by a case study in China. Section 20.4 will provide summaries and conclusions.

20.2 Methods

20.2.1 The Basic Idea

The basic idea in this approach is that we only need to sample the locations where the soils are typical of the soil categories to be mapped and soil variation from these typical locations can be approximated by membership gradation under fuzzy logic. Under this notion, soil-landscape relationships between the soils and the environmental conditions at these locations can be used to approximate the soil-landscape model of the area. In this way we can minimize the extent and the amount of field investigation efforts and makes it possible to use digital soil mapping approaches to map soils in areas with limited data.

The key problem in this pursue is to distinguish locations where the soils are the typical instances of the classes to be mapped from other locations where soils are somewhere between types without extensive sampling. To address this question, we assume that the typical instances of soil classes correspond to the unique configurations (combinations) of environment conditions. Thus, the problem now becomes that of finding unique configurations of environment conditions. GIS/RS techniques are used to characterize the soil environmental conditions and fuzzy classification techniques are used to identify the unique combinations (or environment classes) that exist in the environmental data set. Fuzzy maps of the derived environment classes are then used to determine locations where the soils are typical. Soil-landscape model can then be developed for the area by linking these typical instances of soils to the environmental conditions.

20.2.2 The Approach

Our purposive sampling method consists of four major steps: (1) development of soil environment database; (2) identification of environmental configuration (combination) patterns; (3) field-investigation; (4) development of soil-landscape model.

Soil-environment database development. Environment covariates related to soil conditions are first identified. A GIS database on these environment covariates are then generated given that the source data is available and GIS data layers can be created for each covariate. Typical environment covariates includes elevation, slope gradient, slope aspect, surface profile curvature, surface contour curvature, topographic wetness index, parent materials (often approximated by geology data layer) and vegetation information. The specific list of data layers to be used for a given area depends on the pedogenesis and data availability of the area.

Identification of environment configuration patterns. We employed a fuzzy c-means classifier (FCM) to identify patterns of environment configuration. FCM is a classifier which first optimally partitions a dataset (such as the environmental database) into a given set of classes and then computes the membership of each object in each of the classes (Bezdek et al., 1984). It identifies the centroids of classes by minimizing the fuzzy partition error as given in Equation 20.1 (Bezdek et al., 1984):

$$J_m(U, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \left\| y_k - v_i \right\|^2 \mathrm{A} \tag{20.2}$$

where y is the data; c is the number of clusters in y; m is a weighting exponent for fuzziness; U is a fuzzy c-partition of Y; v is a vector of cluster centres; A is a weighting matrix; n is the number of objects in set y; u_{ik} is the membership of the kth object (y_k) belonging to the ith cluster. J_m, the fuzzy partition error, can be described as a weighted measure of the squared distance between pixels and class centroids, and so is a measure of the total squared errors as minimized with respect to each cluster (Ahn et al., 1999; Ross, 1995). J_m decreases as the clustering improves (meaning that pixels tend to be overall closer to their representative centroids).

In most cases, one does not know the number of classes that best describe the structure in the data set. To judge the effectiveness of the clustering results generated using the above fuzzy c-means algorithm, two cluster validity measures (partition coefficient (F) and entropy (H)) are defined as (Bezdek et al., 1984):

$$F_c(\hat{u}) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\hat{u}_{ik})^2 / n \tag{20.3}$$

$$H_c(\hat{u}) = -\sum_{k=1}^{n} \sum_{i=1}^{c} (\hat{u}_{ik} \log_{\mathrm{a}}(\hat{u}_{ik}))/n \tag{20.4}$$

Partition coefficient F will take the values of $1/c$ to 1, while entropy H ranges from zero to $\log_{\mathrm{a}}(c)$ (Ahn et al., 1999). F measures the amount of overlap between clusters, and is inversely proportional to the overall average overlap between pairs of fuzzy sets (Ahn et al., 1999). H, conversely, is a scalar measure of the amount of fuzziness in a given fuzzy partition U (Bezdek, 1981). The best fuzzy c-partition, e.g. the number of classes that best describe the structure in the data set, is thus the c-partition which realizes the highest $F_c(\hat{u})$ and the lowest $H_c(\hat{u})$ (Ward et al., 1992). Note that both H and F will reach maxima and minima at the same points, and in this sense they are essentially equivalent (Bezdek, 1981).

It is often the case that F increases and H decreases as the number of classes decreases. To determine if a fuzzy clustering can be considered optimal, i.e. the number of clusters optimally describes the structure in the dataset, one should examine the improvement in entropy or partition coefficient over adjacent clusterings

(English, 2001). If there is a significant improvement, one can consider the current clustering is a better partition of the dataset.

Field-investigation. Once the optimal clustering of the environmental data set (the environment configuration pattern) is determined, membership maps for clusters can be produced. Spatial locations of environment clusters and areas of environment transition can be identified on these maps. For each membership map, the locations of its cluster centres (typical instances) are at those locations where membership values are very high (>0.8 on a 0∼1 scale). Thus, field investigation efforts should be mostly allocated to these areas.

Development of soil-landscape model. By investigating the status or property of the soils at the locations of typical instances of environmental clusters, one can quickly establish the relationships between soils and its environmental conditions. The so-developed soil-landscape model can then be used in predictive soil mapping using approaches such as SoLIM (Zhu et al., 2001).

20.3 Case Study

In this paper we present the result of a case study using the purposive sampling approach. The case study was conducted over a watershed in Heilongjiang Province, China, where there are no local field soil surveyors, no existing soil maps, no field soil samples. Digital data on environmental conditions are scarce.

20.3.1 Study Area

The study area is a 60 km^2 watershed located in Heshan Farm, Nenjiang county, Heilongjiang province, China (Fig 20.1). The elevation ranges from 270 to 370 m. Most slope is under 4°. The native vegetation is meadow, but the area has been cultivated as cropland for the past 40 years. Crops in the watershed and over the general area are generally limited to soybean and wheat. The soils in the area are formed from the deposits of loamy loess. The parent materials for the area are the same in the whole area except in the valley bottom which mostly occupied by fluvial deposits.

The area is considered data poor. The only soil map available is the 1:1,000,000 national soil map. For the general area, there are no field soil surveyors from whom soil-landscape knowledge might be extracted. There are no field soil samples of any kind for the watershed and the general area. Digital data on environmental conditions is scarce. However, there is a topographic map at 1:10 000 for the area.

20.3.2 Soil Classes and Environmental Data

The Chinese soil taxonomy was chosen for this study (Chinese Soil Taxonomy Research Group, 2001). This taxonomy currently is at the soil subgroup level and lower classification units (such as soil series) are under development. In this case study we employed subgroup as the basic soil unit to test our idea.

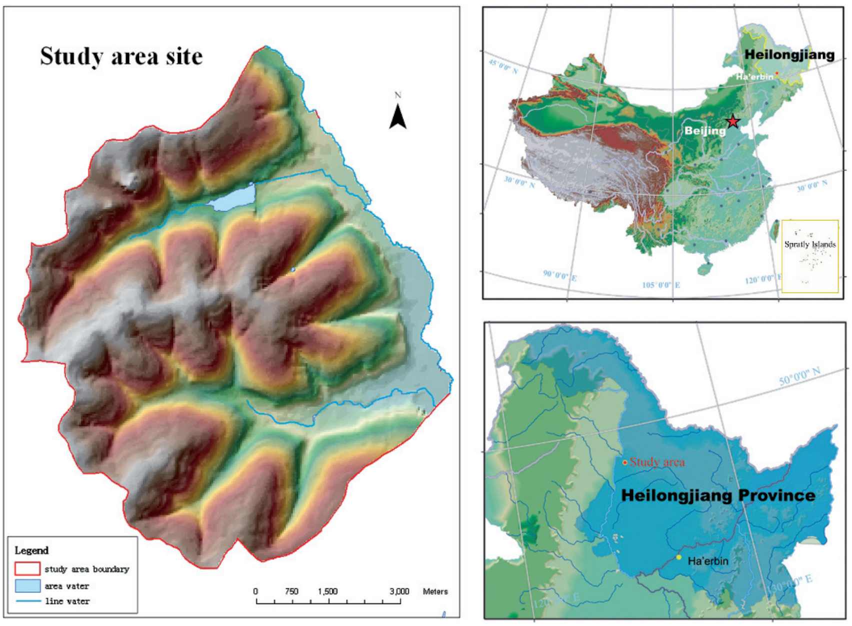


Fig. 20.1 Location of study site: Heshan Farm, Nenjiang county, Heilongjiang Province, China (See also Plate 22 in the Colour Plate Section)

As indicated in 20.2.2.1, at watershed scale the most important environment factors which co-vary with soil are usually bedrock geology, topographical characteristic and vegetation. Bedrock geology of the whole study area is almost the same. The vegetation of the area is heavily altered by human cultivation. The human cultivation over the area is quite similar. Thus, we did not include information data layer for our analysis. Topographic variables are expected to be important in reflecting local soil conditions. Thus, the follow five topographic variables (elevation, slope gradient, contour curvature, profile curvature and topographic wetness index) were used in this study to characterize the environment conditions. We did not use slope aspect as the relief in the area are so gentle we do not expect the difference in soil among different slope aspects to be significant. Information on the five terrain variables were derived from a 10 m resolution DEM which was created for this project from the 1:10 000 topographic map of the area. Figure 20.2 shows slope gradient, profile curvature, contour curvature and topographic wetness index for the area.

20.3.3 Identification of Unique Environment Combinations

Environmental data layers were preprocessed before FCM classification was performed. The processing was to remove outliers and to standardize the ranges of the input data layers. The outliers were those data values which are low in frequency

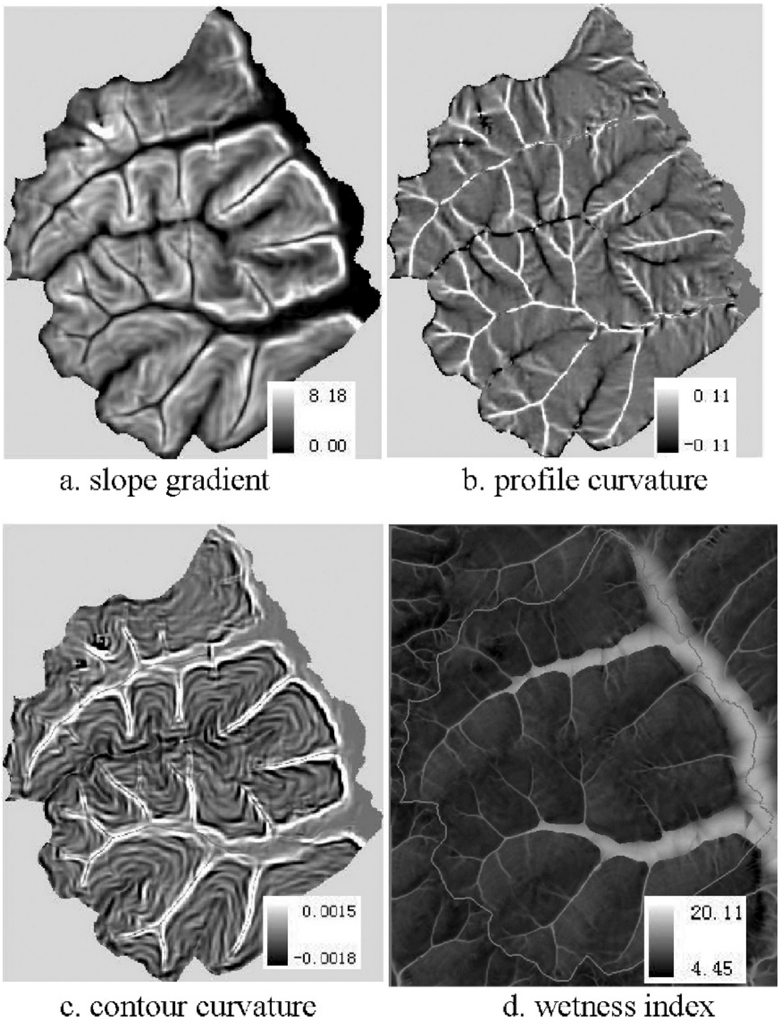


Fig. 20.2 Environmental data of study area: (**a**) slope gradient, (**b**) profile curvature, (**c**) contour curvature, (**d**) wetness index

and at the extremes of data distribution. These outliers were mostly errors introduced during the creation of the digital elevation model and will have a strong impact on the classification results. Thus, they need to be removed. The new values assigned to these outlier pixels are those next to these extremes so that integrity of the data is preserved.

Most of the data layers will have different data ranges. For example, the data range for the slope gradient is certainly different from that of the contour curvature. If the ranges are not standardized, the variables (factors) will be weighted differently in the classification process. In this case, we assume that these variables have equal level of co-variation with soil, thus we need to treat them equally in the classification. Thus, we standardized data layers to the same numerical range. Elevation, slope and wetness index were stretched to $0 \sim 100$. Profile curvature and contour curvature were stretched to $–50 \sim 50$, keeping the 0 value unchanged.

A set of FCM classification were performed across five different fuzziness exponents ($m = 1.5$, 1.75, 2.0, 2.25, and 2.5). For each run (per m) the number of clusters examined ranges from 2 to 20. By examining the improvements in partition coefficient (F) and entropy (H) for all five sets, we chose 13 as the optimal cluster number of our data set. We argue that there are 13 clusters (13 unique combinations of environmental conditions) within the data.

20.3.4 Investigation of Soil Types at Locations of Environmental Combinations

A total of 23 field observations were made in the field to investigate the soil at these class centres (typical locations of these environmental combinations). The field observations were guided by the membership values. For each class, the observation sites were at locations where the membership values for the class are higher than 0.85. Two field points were selected for each class and soil types at each point were identified by a soil taxonomy expert. If the soil types of the two points were different, a third point was then selected.

20.3.5 Establishment of Soil-Landscape Model

Field observations were first linked with each class to establish the association between the environmental classes and soil types observed in the field. Table 20.1

Table 20.1 Association between environmental clusters and soil types

FCM classes	
Environmental combination	Soil types (subgroup)
Class 1	Mollic Bori-Udic Cambosols
Class 2	Pachic Stagni-Udic Isohumosols
Class 3	Mollic Bori-Udic Cambosols
Class 4	Typic Hapli-Udic Isohumosols
Class 5	Pachic Stagni-Udic Isohumosols
Class 6	Fibric Histic-Stagnic Gleyosols, Typic Haplic-Stagnic Gleyosols
Class 7	Typic Hapli-Udic Isohumosols
Class 8	Typic Bori-Udic Cambosols
Class 9	Typic Hapli-Udic Isohumosols
Class 10	Typic Hapli-Udic Isohumosols
Class 11	Typic Hapli-Udic Isohumosols
Class 12	Lithic Udi-Orthic Primosols
Class 13	Typic Bori-Udic Cambosols

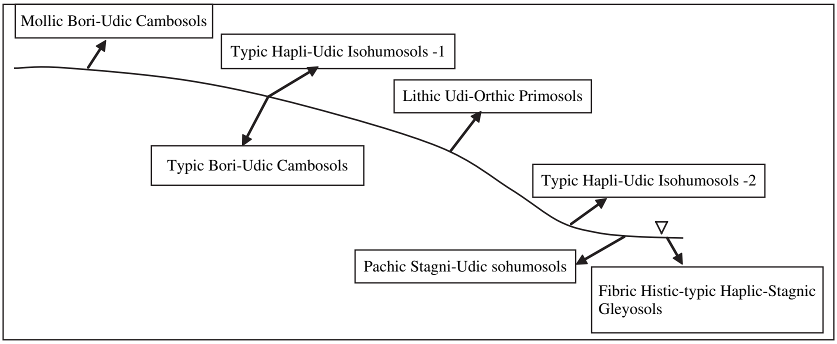


Fig. 20.3 General catenary sequences of soil types in the study area

shows such association for the study area. It is clear that there is a good association between the environmental classes and the soil types in the area. However, the association is not one-to-one. A soil type can occur under different unique environmental conditions. For example, Mollic Bori-Udic Cambosols is relating to two different environmental combinations (class 1 and class 3). It might also be possible that two soil types may occur under one environmental class. This typically occurs over areas where the environmental variables used are insufficient to differentiate the environmental conditions under which each of the soil types occupies. We encountered this problem for Class 6 which covers the central floodplain of the valley where two soil types: Fibric Histic Stagnic Gleyosols and Typic Haplic Stagnic Gleyosols intermittently distributed. Due to their intermittent nature we put these two as an association in the soil mapping process.

The second stage is to combine the classes to form a soil landscape model for the area. For example, both class 1 and 3 were identified as Mollic Bori-Udic Cambosols. By examining their spatial relationships we found them located next to each other so we combined the two into one. While there were five environmental combinations relating to Typic Hapli-Udic Isohumosols (Table 20.1). Class 4, 7, 9, and 11 were adjacent in space, thus they were combined into one. Class 10 was not contiguous with those four classes and treated as a different instance of Typic Hapli-Udic Isohumosols. Thus, Typic Hapli-Udic Isohumosols has two instances: (1) the combination of class 4, 7, 9, and 11 is one instance and referred to as Typic Hapli-Udic Isohumosols-1; and (2) Class 10 represents another and referred to as Typic Hapli-Udic Isohumosols-2. After this process of combining classes, spatial catenary sequence of soil types over the area was generated (Fig. 20.3).

20.3.6 Evaluation of the Soil-Landscape Model

To assess the validity of the soil-landscape model constructed using the purposive sampling approach, the soil-landscape model was used under the SoLIM framework to generate soil maps for the area. The SoLIM approach is a knowledge-based approach to digital soil mapping. It combines the knowledge on soil-environment relationships (soil-landscape model) with the environmental conditions to infer the spatial distribution of soils (Zhu, 1997; Zhu, 1999; Zhu et al., 2001). Case studies have demonstrated that the SoLIM approach to soil mapping is successful (Zhu et al., 2001). However, the quality of the soil maps from SoLIM largely depends on the quality of the soil-landscape model. Thus, the SoLIM approach provides us with the opportunity to examine the quality of the soil-landscape model constructed with the use of FCM. The details on operation of SoLIM approach are beyond the discussion of this paper and can be found in the user manual and tutorial guide of SoLIM available at *http://solim.geography.wisc.edu/software/index.htm*.

The harden soil map derived from SoLIM based on the soil-landscape model derived is shown in Fig. 20.4. The soil map shows a catenary sequence of the soils in the area: This pattern matches field observation of catenary sequence in the area well.

To validate this soil map, a second set of 45 field sites were investigated. This set is referred to as the validation data set. Regular and transecting sampling strategies were employed to collect this validation data set. Soil type at each field site was identified at the subgroup level by the same soil taxonomy expert.

The field observed soil type at these sites was compared with the soil type obtained from the inferred soil map. Soil subgroups from the inferred soil map match field observed soil subgroups at 34 of the 45 sites, which accounts for about 76% of accuracy. In addition, the accuracy of transacting points was 80%, which indicated that the hardened soil map could capture local variation of soil information as well as the overall soil spatial variation. Unfortunately, there is no large scale soil map in our study area for comparison. Given that the accuracy of most 1:24,000 scale

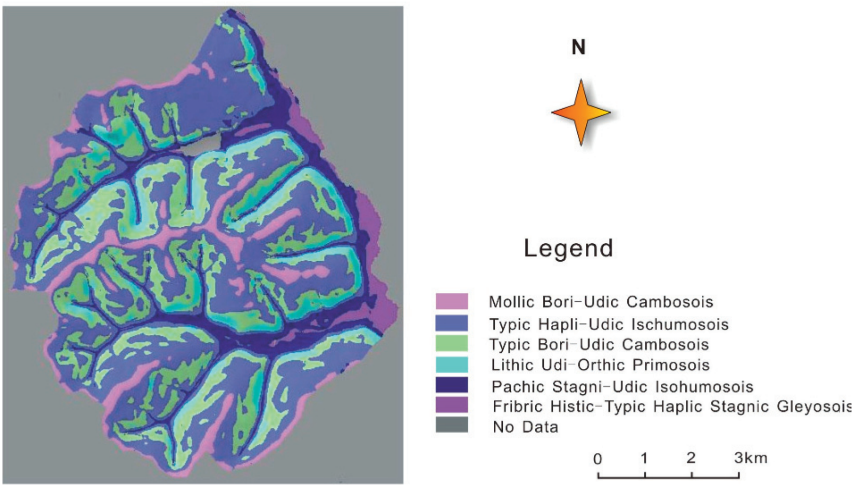


Fig. 20.4 Soil map produced from SoLIM using the soil-landscape model constructed using the FCM-based method (See also Plate 23 in the Colour Plate Section)

soil maps produced in U.S. is about 60%, 76% accuracy is acceptable for an initial soil mapping with limited data. Therefore, we believe that the soil-landscape model developed through the purposive sampling methodology is of good quality.

20.4 Conclusions

This paper presents a purposive sampling approach to assist the development of knowledge of relationships between soil and its environmental conditions. The method employed a fuzzy *c*-means classification to identify unique combinations of environmental conditions and to discern locations of these unique combinations. The results (the unique combinations and the spatial locations of these unique combinations) were then used to direct field investigation efforts and to improve the efficiency of acquisition of knowledge on soil-environment relationships.

Through a soil mapping case study it was found that the approach was effective in developing the knowledge of the soil-environmental relationships. The approach was able to reduce the amount of field observations and the acquired knowledge of the relationships was of good quality.

Acknowledgments The research reported here was supported by the Chinese Academy of Sciences through its International Partnership Project "Human Activities and Ecosystem Changes" (CXTD-Z2005-1) and its "One-Hundred Talents" program to A-Xing Zhu. Assistance in soil classification provided by Mr. Guohua Du from the Institute of Soil Science, Chinese Academy of Sciences, is greatly appreciated.

References

Ahn, C.W., Baumgardner, M.F., Biehl, L.L., 1999. Delineation of soil variability using geostatistics and fuzzy clustering analyses of hyperspectral data. Soil Science Society of America Journal 63, 142–150.

Bell, J.C., Cunningham, R.L., Havens, M.W., 1992. Calibration and validation of a soil-landscape model for predicting soil drainage class. Soil Science Society of America Journal 56, 1860–1866.

Bell, J.C., Cunningham, R.L., Havens, M.W., 1994. Soil drainage probability mapping using a soil-landscape model. Soil Science Society of America Journal 58, 464–470.

Bui, E.N., Loughhead, A., Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. Australian Journal of Soil Research 37, 495–508.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences 10, 191–203.

Burgess, T.M. Webster, R., 1980a. Optimal interpolation and isarithmic mapping of soil properties: I. The semivariogram and punctual kriging. Journal of Soil Science 31, 315–331.

Burgess, T.M. Webster, R., 1980b. Optimal interpolation and isarithmic mapping of soil properties: II. Block kriging. Journal of Soil Science 31, 333–341.

Chinese Soil Taxonomy Research Group, 2001. Keys to Chinese Soil Taxonomy (3rd edition). University of Science and Techonology of China Press, Hefei.

English, E.M., 2001. Assisting Knowledge-Based Inferential Soil Mapping: The Application of Fuzzy c-means Clustering to Expose Environmental Niches. M.Sc. Thesis, University of Wisconsin-Madison, Madison, Wisconsin, U.S.A., 80pp.

Evans, I.S., 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. In: Chorley, R.J. (Ed.), Spatial Analysis in Geomorphology, Methuen, London, pp. 17–90.

Evans, I.S., 1998. What do terrain statistics really mean? In: Lane S.N., Richards, K.S., Chandler, H. (Eds.), Land Monitoring Modelling and Analysis, John Wiley & Sons, Chichester, pp. 119–138.

Gessler, P.E., Chadwick, O.A., Chamran, F., Althouse, L., Holmes, K., 2000. Modeling soil-landscape and ecosystems properties using terrain attributes. Soil Science Society of America Journal 64, 2046–2056.

Hudson, B.D., 1992. The soil survey as paradigm-based science. Soil Science Society of America Journal 56, 836–841.

Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-Hill, New York.

Jenny, H., 1980. The Soil Resource: Origin and Behaviour, Springer-Verlag, New York.

Lark, R.M., Papritz, A., 2003. Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115, 245–260.

Li, W., Burt, J.E., Zhu, A.X., Zhang, C., Feyen, J., 2004. Two-dimensional Markov chain for simulating spatial distribution of soil types, Soil Science Society of America Journal 68, 1479–1490.

Li, W., Zhang,C., Burt, J.E., Zhu, A.X., 2005. A Markov chain-based probability vector approach for modeling spatial uncertainties of soil classes. Soil Science Society of America Journal, 69, 1931–1942.

Lillesand, T.M., Kiefer, R.W., 2000. Remote Sensing and Image Interpretation, John Wiley, New York.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97, 293–327.

McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

McBratney, A.B., Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy. Journal of Soil Science 34, 137–162.

Moran, C.J., Bui, E.N., 2002. Spatial data mining for enhanced soil map modeling**.** International Journal of Geographical Information Science 16, 533–549.

Odeh, I.O.A., McBratney, A.B., Chittleboruough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokringing and regression-kriging. Geoderma 67, 215–225.

Qi, F., Zhu, A.X., 2003. Knowledge discovery from soil maps using inductive learning, International Journal of Geographical information Science 17, 771–795.

Qi, F., Zhu, A.X., Harrower, M., Burt, J.E., 2006. Fuzzy soil mapping based on prototype category theory. Geoderma 136, 774–787.

Ross, T.J., 1995. Fuzzy logic with Engineering Applications. McGraw Hill, New York.

Shary, P.A., Sharayab, L.S., Mitusov, A.V., 2002. Fundamental quantitative methods of land surface analysis. Geoderma 107, 1–43.

Shi, X., Zhu, A.X., Burt, J.E., Qi, F., Simonson, D., 2004. A cased-based reasoning approach to fuzzy soil mapping. Soil Science Society of America Journal 68, 885–894.

Walvoort, D.J.J., de Gruijter, J.J., 2001. Compositional kriging: a spatial interpolation method for compositional data. Mathematical Geology 33, 951–966.

Ward, A.W., Ward, W.T., McBratney, A.B., de Gruijter, J.J., 1992. MacFUZZY: A program for data analysis by generalized fuzzy k-means on the Macintosh. CSIRO Australia Division of Soils Divisional Report 116, Glen Osmond, South Australia.

Webster, R., 1991. Local disjunctive kriging of soil properties with change of support. Journal of Soil Science 42, 301–318.

Webster, R., Burgess, T.M., 1980. Optimal interpolation and isarithmic mapping of soil properties: III Changing drift and universal kriging. Journal of Soil Science 31, 505–524.

Wilson, J.P., Gallant, J.C., (Eds.), 2000. Terrain Analysis: Principles and Applications, John Wiley, New York.

Zhu, A.X., 1997. A similarity model for representing soil spatial information. Geoderma 77, 217–242.

Zhu, A.X., 1999. A personal construct-based knowledge acquisition process for natural resource mapping using GIS. International Journal of Geographic Information Science 13, 119–141.

Zhu, A.X., 2000. Mapping soil landscape as spatial continua: the neural network approach. Water Resources Research 36, 663–677.

Zhu, A.X., Hudson, B., Burt, J.E., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal 65, 1463–1472.

Chapter 21
Assessment of Land Degradation Using NASA GIMMS: A Case Study in Kenya

D.L. Dent and Z.G. Bai

Abstract Direct assessment of land degradation globally is constrained by limited spatial data – soil data in particular (see also Chapter 7). As a proxy, biomass has been adopted as an integrated measure of productivity; its deviance from the norm may indicate land degradation or improvement. Biomass can be assessed by remote sensing of the normalized difference vegetation index (NDVI); norms may be established according to climate, soil, terrain and land use. As a pilot for a *Global Assessment of Land Degradation and Improvement*, spatial patterns and temporal trends of green biomass across Kenya were analysed using 23 years of fortnightly NOAA-AVHRR NDVI data and CRU TS 2.1 station-observed monthly precipitation. Trends of various biomass indicators and climate variables were determined by regression at annual intervals and mapped to depict spatial changes. In Kenya over the period of 1981–2003, biomass increased over about 80% of the land area and decreased over 20%. Most of the decrease has been across the more-productive areas – cropland in the high-rainfall zones. To assess whether this trend represents land degradation or declining rainfall, we calculated rain-use efficiency, the ratio between green biomass (NDVI) and rainfall. Combined trends of biomass and rain-use efficiency may be a more robust indicator of land degradation in areas where productivity is limited by rainfall. Thus defined, degrading areas occupy 17% of the country: most extensively in the drylands around Lake Turkana and the marginal cropland in Eastern Province.

21.1 Introduction

Land degradation is a global development and environmental issue (UNCED, 1992; UNEP, 2007) but there is no authoritative, global measure. The only harmonized assessment, the *Global Assessment of Human-induced Soil Degradation* (Oldeman

D.L. Dent
ISRIC – World Soil Information, Wageningen, The Netherlands
e-mail: david.dent@wur.nl

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

et al., 1991), is a map of perceptions of the kinds and degree of degradation – not a measure of degradation. Its expert assessments have proven to be inconsistent, not reproducible, and not well correlative with other policy relevant measures such as crop yields (Sonneveld and Dent, 2007). Land degradation and perceptions have moved on; there is pressing need for a quantitative, reproducible assessment to support policy development for food and water security, environmental integrity, and economic development. This is now under way within the FAO/UNEP program Land Degradation Assessment in Drylands (LADA) to identify: 1) the state and trends of land degradation, 2) *black spots* suffering extreme constraints or at severe risk, *bright spots* where degradation has been arrested or reversed.

Biomass is an integrated measure of productivity; its deviation from the local norm may be taken as a measure of land degradation or improvement. Global measurements can be derived from satellite data, in particular the normalized difference vegetation index (NDVI – the difference between reflected near-infrared and visible wavebands, divided by the sum of these two wavebands). NDVI has a strong linear relationship with the fraction of photosynthetically active radiation absorbed by the plant (Asrar et al., 1984; Sellers et al., 1997); many studies have shown strong correlation between NDVI and vegetation cover (e.g. Purevdoj et al., 1998) and above-ground net primary productivity (Paruelo et al., 1997). It has been applied in studies of land degradation from the field scale (1:10 000) to regional and global scale (1:1 million to 1:5 million) (e.g. Tucker et al., 1991; Bastin et al., 1995; Stoms and Hargrove, 2000; Wessels et al., 2004; Singh et al., 2006). Local norms may be established by stratifying the land area according to climate, soils and terrain, and land use/vegetation; deviation may then be calculated locally and combined globally to allow universal comparisons, this is further discussed in Chapter 22.

As a pilot for the global LADA program, we analyse the trend of NDVI indicators in Kenya over a 23-year period (1981–2003) alongside climatic and land use data for the same period and the KENSOTER digital soil and terrain database. More than 80% of Kenya is dryland. The pressure of burgeoning population without compensating investment in soil and water conservation threatens irreversible land degradation, loss of rural livelihoods, and water supplies to urban areas, hydro-power and irrigation schemes.

21.2 Data and Analysis

The Global Inventory Modelling and Mapping Studies (GIMMS) dataset comprises radiometer (AVHRR) data collected by National Oceanic and Atmospheric Administration (NOAA) satellites, generalised to fortnightly images of 8-km spatial resolution, corrected for calibration, view geometry, volcanic aerosols, and other effects not related to actual vegetation change (Tucker et al., 2004). The accuracy of GIMMS is proven to be suitable for a global assessment and it is compatible with MODIS and SPOT data (Tucker et al., 2005; Brown et al., 2006).

We used GIMMS data from July 1981 to December 2003, along with monthly rainfall from the CRU TS 2.1 dataset (Mitchell and Jones, 2005), digital soil and terrain data from the KENSOTER database (Batjes and Gicheru, 2004) and contemporary information on land cover (FAO, 2005, Fig. 21.1). ArcGIS Spatial Analyst and ERDAS IMAGINE were used to calculate various biomass indicators: NDVI minimum, maximum, maximum-minimum, mean, sum, standard deviation (SD) and coefficient of variation (CV) as well as climate variables. Annual NDVI indicators were derived for each pixel, their temporal trends were determined by linear

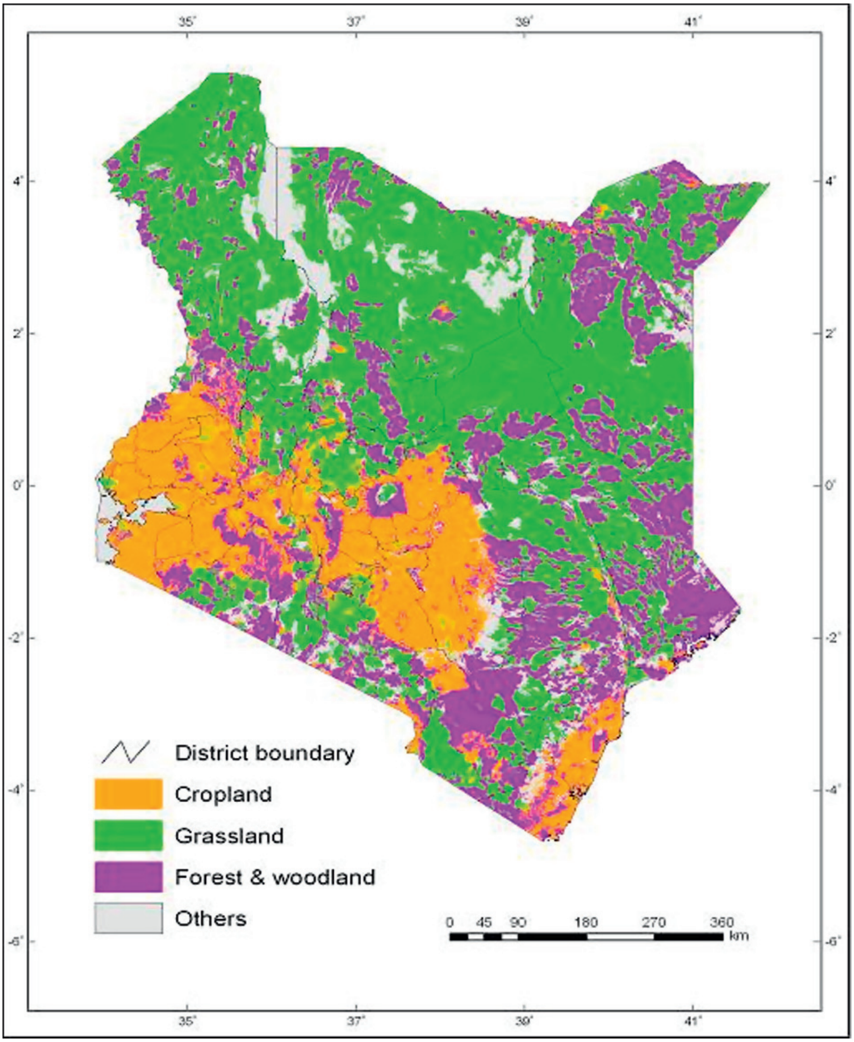


Fig. 21.1 Kenya, dominant land use types (FAO 2005) (See also Plate 24 in the Colour Plate Section)

regression (significance level = 0.05) and mapped to depict spatial changes. A negative slope of linear regression indicates a decline of green biomass; and positive, an increase – except for SD and CV which indicate trends in variability.

21.3 Results

21.3.1 NDVI Indicators

The values of the NDVI indicators and their temporal trends for each pixel, determined by the slope of the linear regression equation are summarised in Table 21.1.

Minimum NDVI: The lowest value that occurs in any one year (annual) – which is usually at the end of the dry season. Variation in minimum NDVI may serve as a baseline for other parameters.

Maximum NDVI: Represents the maximum green biomass. The large spatial variations reflect the diverse landscape and climate.

Maximum-minimum NDVI: The difference between annual maximum and minimum NDVI reflects annual biomass productivity for areas with just one growing season but may not be meaningful for areas with bimodal rainfall, such as Kenya.

Sum or aggregated NDVI: The sum of fortnightly NDVI values for the year, most nearly aggregates annual biomass productivity.

Standard deviation (SD): NDVI standard deviation is the root mean square (RMS) deviation of the NDVI time series values (annual) from their arithmetic mean. It is a measure of statistical dispersion, measuring the spread of NDVI values.

Coefficient of variation (CV): CV images were generated by computing for each pixel the standard deviation of the set of individual NDVI values and dividing this by the mean (M) of these values. This represents the dispersion of NDVI values relative to the mean value. A positive change in the value of a pixel-level CV over time relates to increased dispersion of values, not increased NDVI; similarly, a negative CV dispersion – which is the case over nearly the whole country – means decreasing dispersion of NDVI around mean values, not decreasing NDVI. The trends in CV may reflect land cover change.

21.3.2 Spatial Patterns, Biomass and Rainfall

In Kenya, the trend in biomass over 23 years increased over 80% of the country but decreased over 20%. Most of the decrease was in the better-watered areas (Fig. 21.2a,b), and coincides mostly with cropland, especially, with the expansion of cropland into dryer, marginal areas. For the country as a whole, the 23-year trend was upwards (Fig. 21.3).

Table 21.1 Statistics of NDVI indicators

NDVI indicators	NDVI values (range)			Pixels (%)		% NDVI change/year			Δ NDVI/year		
	min	max	mean	+	–	+	–	mean	+	–	mean
Minimum	0	0.782	0.213	62.1	37.9	1.089	0.759	0.392	0.0023	0.0018	0.0007
Maximum	0.002	0.997	0.497	79.2	20.8	0.240	0.100	0.170	0.0012	0.0006	0.0008
Max-Min	0.0001	0.667	0.298	78.1	21.9	1.914	0.741	1.329	0.0057	0.0026	0.0038
Mean	0.001	0.846	0.329	73.0	27.0	0.979	0.377	0.615	0.0024	0.0011	0.0014
Sum	0.013	10.154	3.946	79.0	21.0	0.993	0.378	0.700	0.0271	0.0116	0.0188
SD	0.048	0.158	0.096	74.0	26.0	1.901	0.797	1.197	0.0015	0.0008	0.0009
CV	0.16	0.668	0.301	67.0	33.0	1.291	1.074	0.507	0.0058	0.0047	0.0023

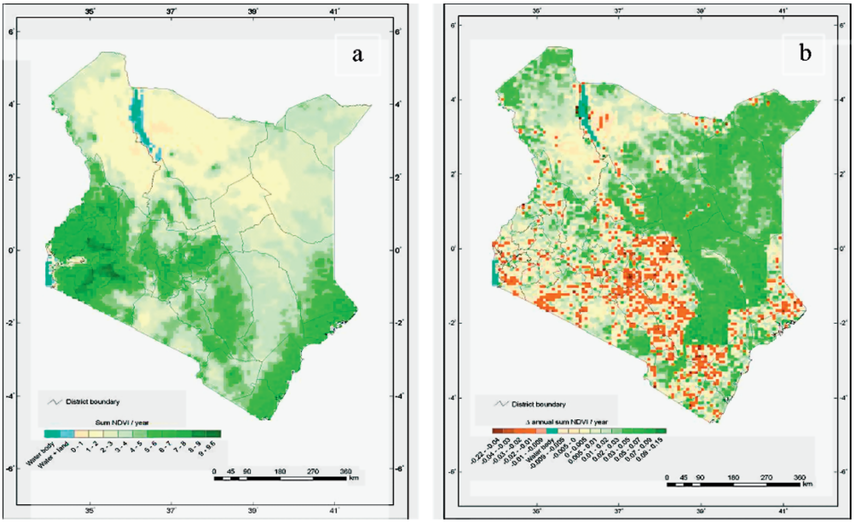


Fig. 21.2 Spatial pattern (**a**) and temporal trend (**b**) of biomass 1981–2003 (See also Plate 25 in the Colour Plate Section)

But productivity depends on rain as well as soil and land use. Mean biomass essentially reflects the mean annual rainfall (Fig. 21.4a) *which has fluctuated significantly*, both spatially and cyclically over the period (Fig. 21.4b). Rainfall increased over about 80% of the country and decreased over 20% (Fig. 21.5).

The overall trend of rainfall is up, so is the overall trend of biomass although the correlation for Kenya as a whole is only moderate ($r = 0.53$).

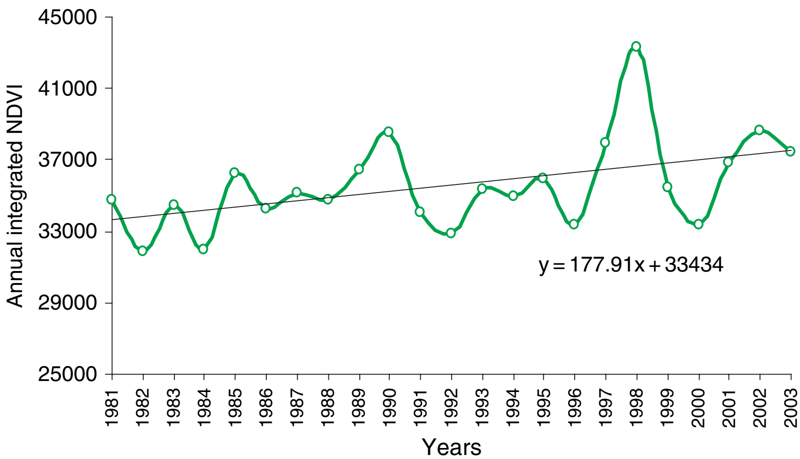


Fig. 21.3 Spatially aggregated annual NDVI 1981–2003

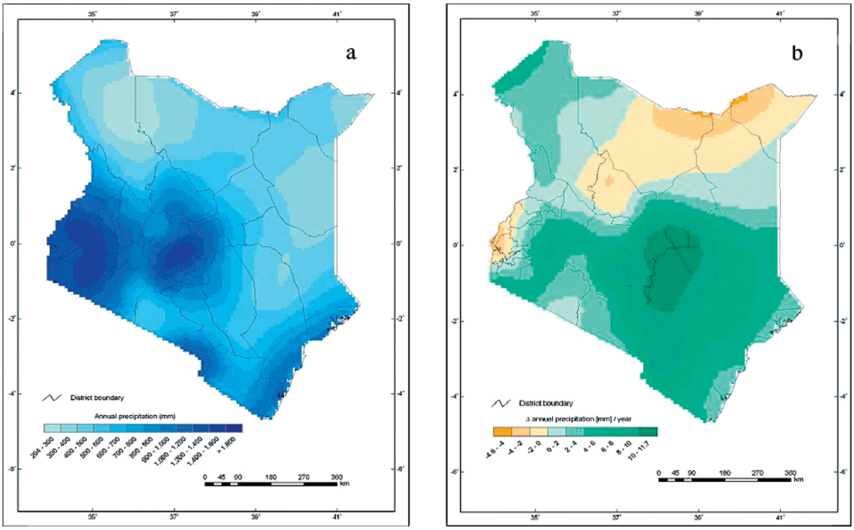


Fig. 21.4 Spatial pattern (**a**) and temporal trend (**b**) of annual rainfall 1980–2002 (See also Plate 26 in the Colour Plate Section)

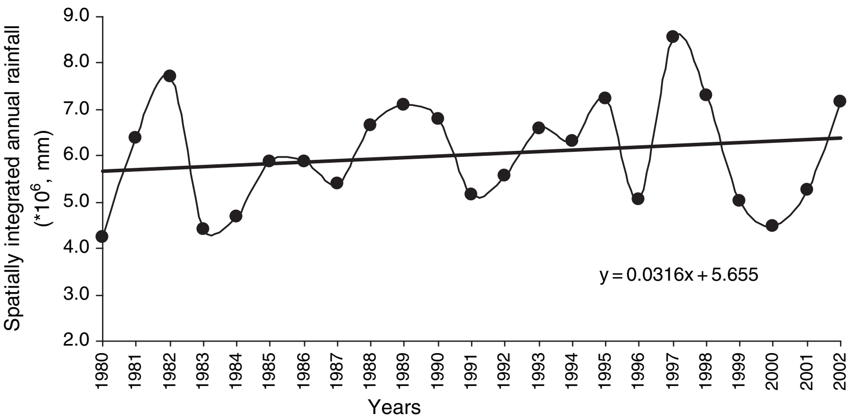


Fig. 21.5 Spatially aggregated annual rainfall 1980–2002

21.3.3 Spatial Patterns of Rain-Use Efficiency

A reduction in biomass does not necessarily mean land degradation. Biomass fluctuates according to variation in rainfall, stage of growth, and changes in land use – which may or may not be related to the land degradation. Rain-use efficiency (RUE), the ratio of net primary productivity to rainfall counters this problem by expressing production per unit of rain. RUE is lower in degraded drylands than in equivalent non-degraded areas (Le Houérou, 1984, 1989) – so deviation from the normal value of RUE and biomass may indicate land degradation or improvement better than biomass alone.

For North China, Bai et al. (2005) demonstrated that values for rain-use efficiency calculated from NDVI, *which are easy to obtain*, were comparable with those calculated from measurements of net primary productivity, which values are not easy to obtain. For Kenya, rain-use efficiency was calculated as the ratio between annual aggregated NDVI and station-observed annual rainfall on a yearly time-step. Figure 21.6 shows the trend of RUE.

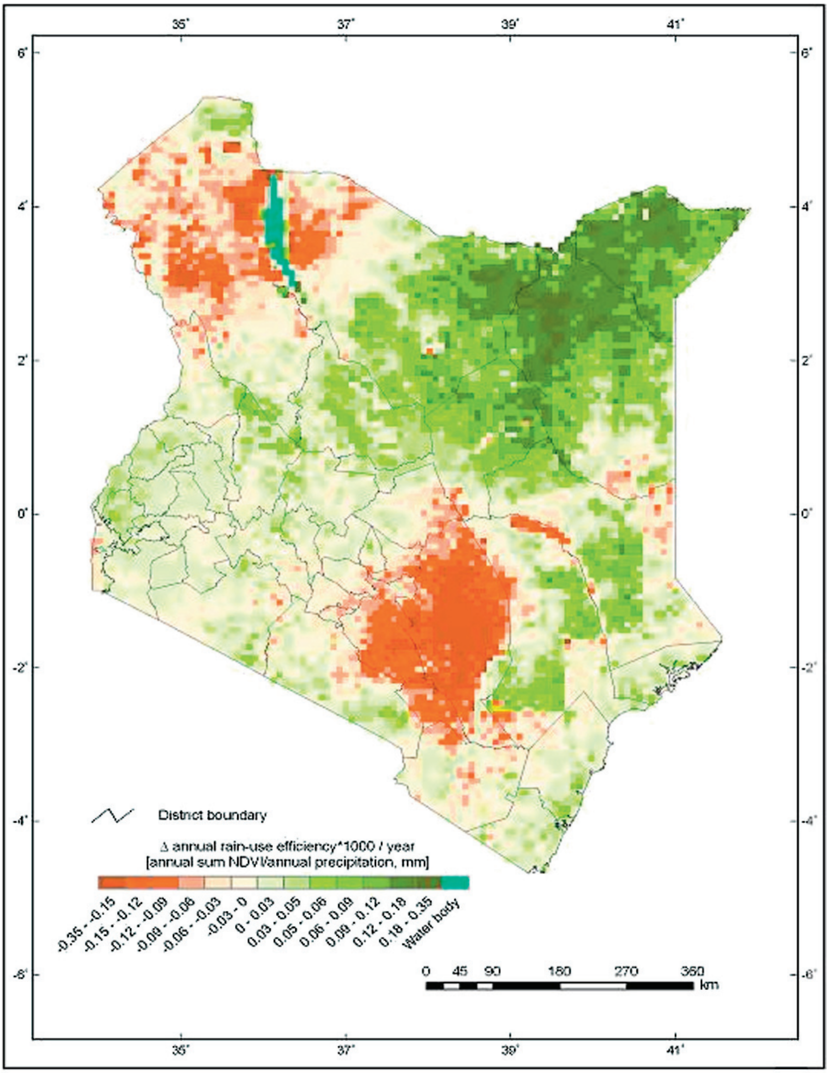


Fig. 21.6 Trend of rain-use efficiency 1981–2002 (See also Plate 27 in the Colour Plate Section)

21.3.4 Analysis of Degrading Land

Potential areas of land degradation were identified as those areas with both declining net primary productivity and declining rain-use efficiency; these areas occupy 17% of the country. Comparison of degradation with land use shows:

Black spots. These include the drylands around Lake Turkana, and a swath of cropland in Eastern Province from Meru south to Machakos including land taken into cultivation during the study period (Fig. 21.7). In the drylands, production has declined from a low base. Degradation of croplands represents decline in an areas of much higher potential.

Relationship with soil type. There is no obvious relation with land degradation and individual soil attributes: e.g. soil organic carbon class 0–0.5%, 0.5–1%, 1–3% and >3%, occupy 20, 46, 26 and 8% of the degrading area, respectively. In respect of terrain, more than 70% of the degrading land is flat and gently undulating, rather than steeply sloping (Table 21.2). There is a weak relationship between degradation and texture-contrast soils; much of degrading land has coarse-textured topsoil over clayey subsoil. It appears that degradation is influenced more by management (i.e. cultivation) than soils and terrain *per se*. A more rigorous analysis is needed to tease out association between land degradation and soil attributes making use of a consistent digital soil map of key soil attributes (see also Chapter 1).

The seamless, quantitative index of land degradation enables statistical correlations with other explanatory variables, such as socio-economic data, provided that these are also geo-located.

21.4 Conclusions

Remote sensing of biomass indicators can indicate *black spots* of land degradation. Interpretation is not straightforward; the various NDVI patterns must be followed up by fieldwork to establish the actual conditions on the ground. For drylands, combination of the biomass trend with rain-use efficiency trend is a more robust indicator of land degradation than crude biomass. For areas not dominated by drought or dry spells, other combinations of indicators, such as length of growing season, may be more appropriate.

Data from a defined, recent period enable us to distinguish between the legacy of historical land degradation and degradation that is taking place now. Much of the historical land degradation is irreversible.

The consistent, spatial data for biomass also enable a statistical examination of other data for which we do not have continuous spatial coverage – such as spot measurements of soil attributes, and socio-economic data that may reveal the drivers of land degradation. Provided these other data are geo-located, they can be scaled against the continuous biomass trends.

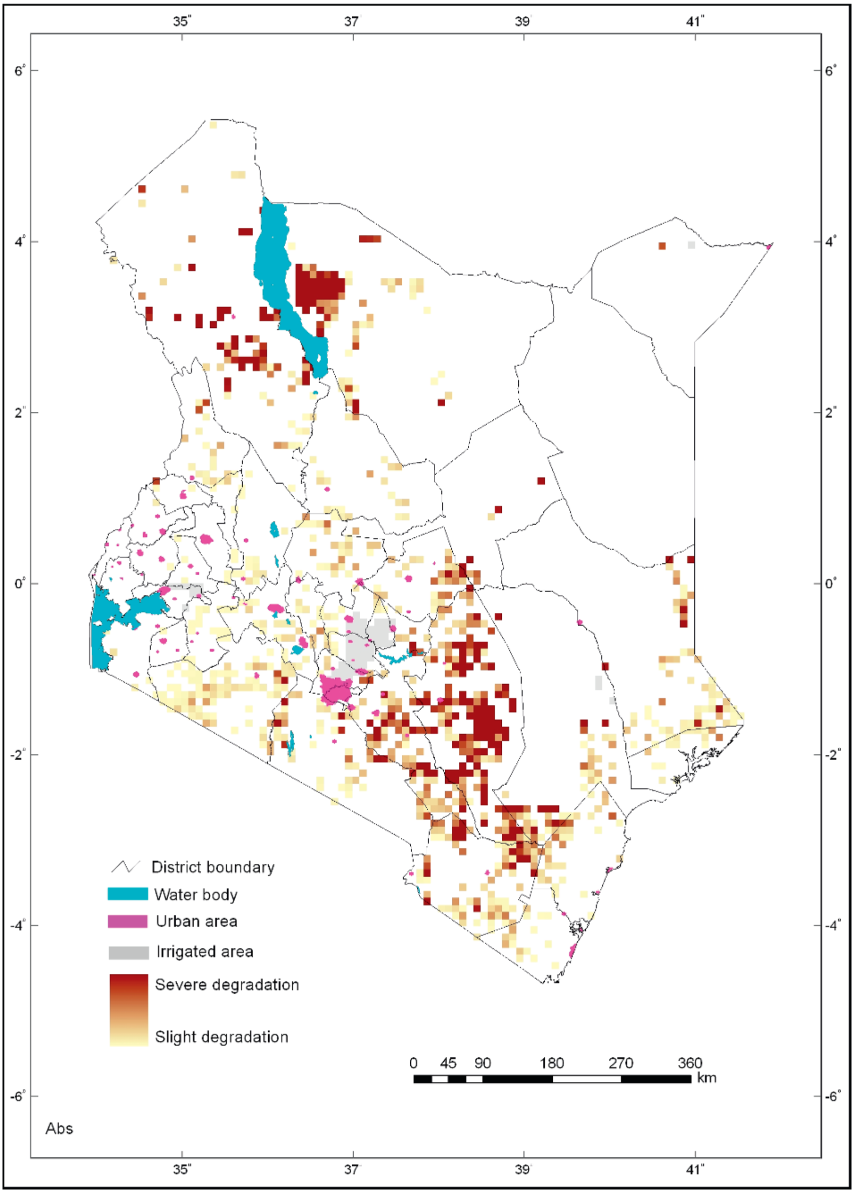


Fig. 21.7 Kenya: black spots of land degradation between 1981 and 2003 (See also Plate 28 in the Colour Plate Section)

Table 21.2 Slope percentage in the degrading land area

Terrain class	Slope (%)	%
W (flat, wet)	0–0.5	3.6
F (flat)	0.5–2	48.8
G (gently undulating)	2–5	23.9
U (undulating)	5–10	8.0
R (rolling)	10–15	2.5
S (moderately steep)	15–30	7.9
T (steep)	30–45	4.9
V (very steep)	45–60	0.4

Acknowledgments We thank C.J. Tucker, J.E. Pinzon and M.E. Brown for access to the GIMMS datasets and M. Salmon for providing the CRU TS 2.1 climate observation data. We are indebted to M.E. Schaepman for concept development and to ISRIC colleagues P. Tempel and J. Huting, for the assistance in data handling. This work is part of the FAO program Land Degradation Assessment in Drylands.

References

Asrar, G.M., Fuchs, M., Kanemasu, E.T., Hatfield, J.L., 1984. Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat. Agronomy Journal 87, 300–306.

Bai, Z.G., Dent, D.L., Schaepmam, M.E., 2005. Quantitative global assessment of land degradation and improvement: pilot study in North China. ISRIC Report 2005/06, Wageningen.

Bastin, G.N., Pickup, G., Pearce, G., 1995. Utility of AVHRR data for land degradation assessment – a case study. International Journal of Remote Sensing 16, 651–672.

Batjes, N.H. and Gicheru, P., 2004. Soil data derived from SOTER for studies of carbon stocks and change in Kenya (ver. 1.0). Report 2004/01, ISRIC – World Soil Information, Wageningen.

Brown, M.E., Pinzon, J.E., Didan, K., Morisette J.T. and Tucker, C.J. 2006. Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-Vegetation, SeaWIFS, MODIS and LandSAT ETM+, IEEE Trans. Geoscience and Remote Sensing 44, 1787–1793.

FAO, 2005. Agricultural trend analysis using satellite remote sensing data: a case study – Kenya. FAO Land and Plant Nutrition Management (AGLL), Rome.

Le Houérou, H.N., 1989. The Grazing Land Ecosystems of the African Sahel. Springer-Verlag, Berlin.

Le Houérou, H.N., 1984. Rain-use efficiency: a unifying concept in arid-land ecology. Journal of Arid Environments 7, 213–247.

Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high resolution grids. International Journal of Climate 25, 693–712.

Oldeman, L.R., Hakkeling, R.T.A., Sombroek, W.G., 1991. World map of the status of human-induced soil degradation: an explanatory note, second revised edition. ISRIC, Wageningen, and UNEP, Nairobi.

Paruelo, J.M., Epstein, H.E., Lauenroth, W.K., Burke, I.C., 1997. ANPP estimates from NDVI for the central grassland region of the United States. Ecology 78, 953–958.

Purevdoj, Ts., Tateishi, R., Ishiyama, T., Honda, Y., 1998. Relationships between percentage vegetation cover and vegetation indices. International Journal of Remote Sensing 19, 3519–3535.

Sellers, P.J., Dickinson, R.E., Randall, D.A., Betts, A.K., Hall, F.G., Berry, J.A., Collatz, G.J., Denning, A.S., Mooney, H.A., Nobre, C.A., Sato, N., Field, C.B., Henderson-Sellers, A., 1997.

Modeling the exchanges of energy, water and carbon between continents and the atmosphere. Science 275, 502–509.

Singh, D., Meirelles, M.S.P., Costa, G.A., Herlin, I., Berroir, J.P., Silva, E.F., 2006. Environmental degradation analysis using NOAA/AVHRR data. Advances in Space Research 37, 720–727.

Sonneveld, B.G.S. and Dent, D.L., 2007. How good is GLASOD? Journal of Environmental Management doi: 10.1016/j.jenvman.2007.09.008

Stoms, D.M., Hargrove, W.W., 2000. Potential NDVI as a baseline for monitoring ecosystem functioning. International Journal of Remote Sensing 21, 401–407.

Tucker, C.J., Dregne, H.E., Newcomb, W.W., 1991. Expansion and contraction of the Sahara Desert from 1980–1990. Science 253, 299–301.

Tucker, C.J., Pinzon, J.E., Brown, M.E., 2004. Global Inventory Modeling and Mapping Studies (GIMMS) Satellite Drift Corrected and NOAA-16 incorporated Normalized Difference Vegetation Index (NDVI), Monthly 1981–2002. The University of Maryland.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N., 2005. An extended AVHRR 8-km NDVI Dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing 26, 4485–4498.

UNCED, 1992. Managing fragile ecosystems: combating desertification and drought. Agenda 21, Chapter 12. United Nations Conference on Environment and Development http://www.un.org/esa/sustdev/documents/agenda21 (5 August 2005).

UNEP, 2007. Global Environmental Outlook GEO-4, UN Environment Programme, Nairobi.

Wessels, K.J., Prince, S.D., Frost, P.E., Van Zyl, D., 2004. Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1 km AVHRR NDVI time-series. Remote Sensing of Environment 91, 47–67.

Part IV
Digital Soil Mapping – Examples

Chapter 22
Spatial-Temporal Changes in Land Cover, Soil Properties and Carbon Stocks in Rio de Janeiro

A.P.D. Turetta, M.L. Mendonça-Santos, L.H.C. Anjos and R.L.L. Berbara

Abstract The purpose of this study was to evaluate spatial-temporal dynamics of land cover of the Campo Grande and Santa Cruz Administration Regions, both in Rio de Janeiro city. LANDSAT5-TM images from 1984, 1994 and 1999 were used to create land cover maps. A Geographical Information System was used for integrating information into a cohesive and easy to consult cartographic base and database. Matrices were generated by applying Markov's Chains, which allow to describe, model and predict transitions of the land cover. This study examined the relationship between land use change, soil orders and carbon stock in the top 10 cm. It was possible to observe the land cover dynamics, with the conversion of agriculture, anthropogenic area and exposed soil in urban areas, especially in the period 1994–1999. Using secondary data, from soil survey reports, and combine it with the land cover maps in the temporal series, it was possible to observe a potentiality of this approach in soil properties-landscape modeling. The main finding of this study was that land use change is a dynamic process, and the use of soil properties based on secondary data – soil survey reports – can helps environmental planning, but the accuracy depends of the quality and the spatial data distribution. So, its stress that is important to planning the soils surveys for the good data exploitation for future projects.

22.1 Introduction

Inappropriate use of the natural resources in urban and rural areas has been the main cause of environmental degradation. The prediction and to evaluation of the impacts of human activities on the environment is important for environmental management and for the development of procedures to mitigate or prevent negative effects.

The landscape changes because it reflects the dynamic interaction between natural and cultural forces in the environment (Antrop, 2005). Increasing population and

A.P.D. Turetta
Brazilian Agricultural Research Corporation – EMBRAPA Solos, Jardim Botânico, Rio de Janeiro, RJ. Brazil, CEP 22460-000
e-mail: anaturetta@cnps.embrapa.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

urbanization result in a complex process of land use and land cover changes from the local to global scale. This process has profoundly disrupted the structure and function of many ecosystems (Yu and Ng, 2007).

Land use change is a dynamic process and the direction and magnitude of landscape change differ. Land use changes modelling permits to describe processes of landscape changes in quantitative terms and shows alterations tendencies (Mendonça Santos, 1999; Li and Yeh, 2004; Turetta, 2004).

The methodology is based on the use of remote sensing and GIS. GIS-based studies started in the 1990s (McBratney et al., 2003). GIS provides a useful tool to implement the landscape approach with powerful functions and convenient modeling environments (Li and Yeh, 2004 – also discussed in Chapter 4) as digital soil mapping and land use change analysis (Chapter 2 is focused in some technologies for digital soil mapping). It allows land use and transition maps showing changes over time but also transition matrices that show the alterations between the land use classes in an area (Yang and Lo, 2003). Remote sensing data can be used to quantify the type, amount, and location of land use conversion (Fung and LeDrew, 1987; Eastman and Fulk, 1993; Jensen and Cowen, 1999).

Soil parameters can be used as an indicator of land use changes and organic carbon is considered an important indicator (Swift and Woomer, 1999; Murage et al., 2001; Urioste et al., 2006; Manlay et al., 2007). The soil organic carbon pool in the surface soil is sensitive to changes in land use and soil management practice (Tan and Lal, 2005).

The role of climatic variables in soil organic carbon dynamics has been widely recognized at small scales. In general, soil organic carbon pool increases with precipitation and decreases with temperature (Kononova, 1966; Jenny, 1980; Burke et al., 1989). In regions where climatic variation is not distinct, as is the case in our study area, other parameters that are not well correlated with the climate but vary with land use and control organic C levels are important to analyse. These include: texture, drainage class, and slope gradient (Schimel et al., 1994; Tan and Lal, 2005).

The purpose of this study was to evaluate the spatial-temporal dynamics of land coverage of the Campo Grande and Santa Cruz Administration Regions, in Rio de Janeiro city for environmental management and planning.

22.2 Material and Methods

The study area is in Rio de Janeiro city, specifically in Campo Grande and Santa Cruz Administration Regions (Fig. 22.1). This area is approximately 31 000 ha and in the past was responsible for the agricultural production for Rio de Janeiro city. Since 1960, this rural area is urbanized. Our study area is a land cover mosaic, with native covers like forest, mangrove and agricultural and urban areas. For this reason, this area is strategic for research on modelling to predict, test and choose between urban growth scenarios. It is a multidiscipline study that involves remote sensing, socio-economic data and landscape information, as shown in Fig. 22.2.

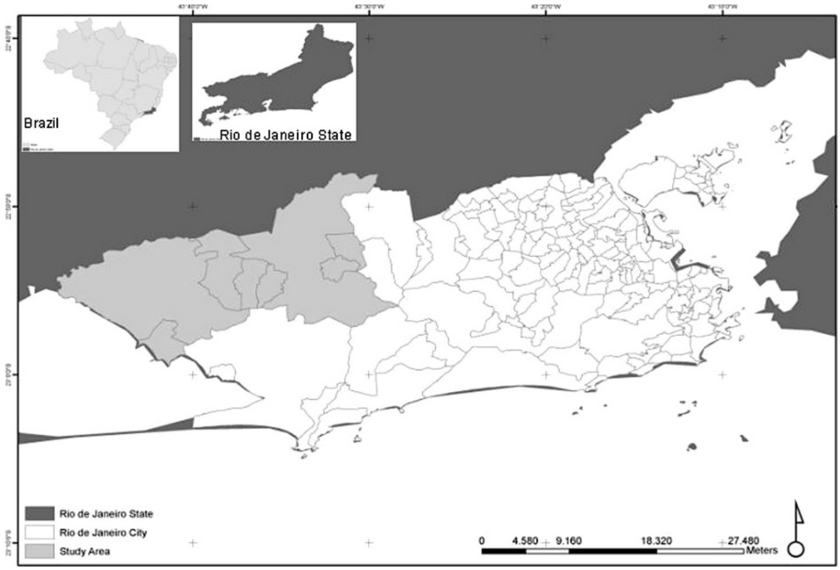


Fig. 22.1 Study area localization. Rio de Janeiro city – RJ, Brazil

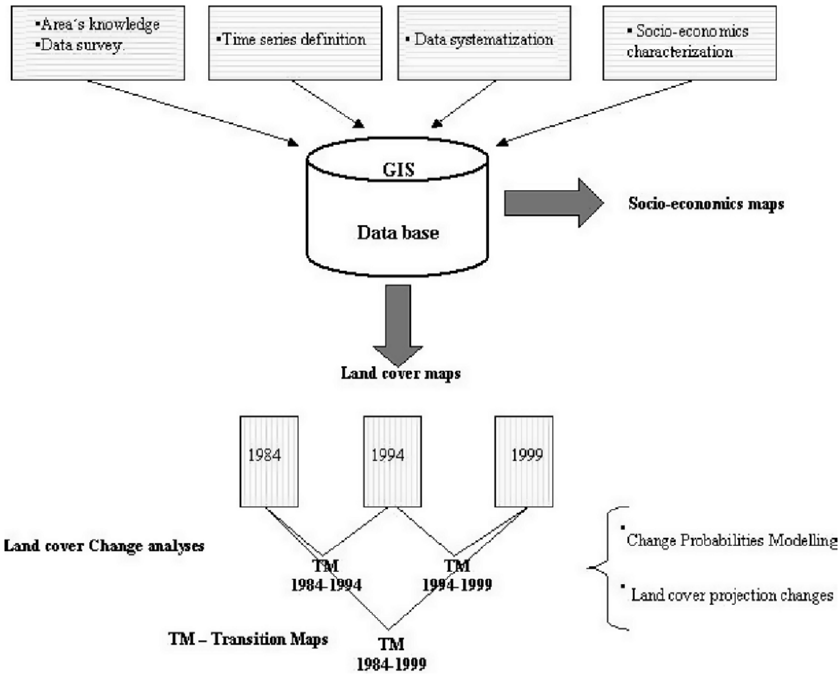


Fig. 22.2 Methodology used in this study

LANDSAT5-TM images of 1984, 1994 and 1999 have been used with 5, 4 and 3 spectral bands in RGB canals, respectively. These images have been classified using the software Spring developed by Spatial Research Institute – INPE (Câmara et al., 1996). With land cover maps, which were generated from the interpretation and classification of satellite images, time change analysis for the land cover have been processed. Transition maps and matrices were generated from algorithms that apply Markov's Chains developed by Mansilla Baca (2002). *MatLab* software was used which allows the analysis of the land cover change in a qualitative and quantitative way over the period of 1984–1999.

In order to analyse the relationship between soils and land cover changes, 21 soil profiles of the study area were selected of National Soil Archives of Embrapa Soils, Rio de Janeiro, Brazil (Santos et al., 2005). The data were from soils surveys conducted between 1958 and 2003. The information considered for these profiles were the soils orders – according to Brazilian Soil Classification System (Santos et al., 2006) and the organic carbon stock in the top 10 cm.

22.3 Results and Discussion

22.3.1 Landscape Dynamics

Eight land cover classes were established: Forest, Shrub, Mangrove, Flooding Area, Agriculture, Exposed Soil, Anthropogenic Areas and Urban Zone (Table 22.1).

The values found between the pairs of images point to a landscape dynamics with conversion of natural vegetation and agriculture to urban use especially between 1994 and 1999. Between 1984 and 1999 there was a great expansion of the urban areas and a decrease in the shrub and agriculture areas (Table 22.2).

The Transition Matrixes presented in Table 22.3 analyses the landscape space-temporal dynamics and identify the transitional and successional land cover states. As synthesis of change analysis results in the considered times, the difference

Table 22.1 Land cover classes

Class	Land cover	Description
1	Forest	Atlantic Forest Fragment.
2	Shrub	Atlantic Forest in regeneration after agriculture use.
3	Mangrove	Low marshy areas, along riverbanks and ocean coastlines in tropical areas.
4	Flooding Area	Wetlands, probably was mangrove before.
5	Agriculture	The most common is a family farm, with main crops: banana, cassava, tomato, pumpkin and many vegetables.
6	Exposed Soil	These are areas for urban expansion.
7	Anthropogenic Areas	No well define use; abandoned areas after agriculture use with urban expansion objective.
8	Urban	Urban use.

Table 22.2 Land cover classes in hectares

Year	Forest	Shrub	Agriculture	Anthrop. area	Urban
	ha				
1984	992	11,780	4,929	3,658	7,440
1994	2,883	4,929	6,169	4,898	8,370
1999	2,201	5,177	3,751	3,999	12,090

Table 22.3 Land cover classes Transition Matrixes (TM)

	1984								
	Class	1	2	3	4	5	6	7	8
1999	1	6.	2.4	0.1	0.1	0.4	0.0	0.2	0.2
	2	1.0	2.5	0.0	0.4	1.2	0.2	1.6	3.1
	3	0.1	0.1	6.0	2.4	0.1	0.0	0.0	1.4
	4	0.1	0.5	0.7	7.5	0.4	0.0	0.4	0.6
	5	0.4	1.6	0.0	0.7	2.6	0.1	1.9	2.6
	6	0.4	2.0	0.1	0.5	1.5	0.3	1.4	4.0
	7	0.4	1.9	0.0	0.5	1.4	0.2	1.5	4.1
	8	0.1	0.6	0.0	0.1	0.5	0.1	0.7	7.9

Class 1: Forest; **Class 2**: Shrub; **Class 3**: Mangrove; **Class 4**: Flooding Area; **Class 5**: Agriculture; **Class 6**: Exposed Soil; **Class 7**: Anthropogenic Areas; **Class 8**: Urban Zone.
* The main diagonal – in gray – represents the land cover class with no class alteration.

between each "State Vector (SV)" is presented in Fig. 22.3. The difference between the State Vectors helps to understand the landscape dynamics. Positive values indicate the increment of an area, and negatives values indicate a decrease. Zero indicates no changes over time for a particular class.

From Fig. 22.3 and Table 22.3, it is possible to summarize the main landscape dynamics: first, a dynamic which characterizes the succession of land cover changes that follows the order: from Forest to Shrub to Anthropogenic Areas/Agriculture. Second, a dynamics characterized by an increase in the urban area. The Fig. 22.4 illustrates this dynamic, witch every land use class contributes to increase the urban class. The wider arrows represents the classes that losses more area to urban use. These classes are agriculture (class 5), exposed soil (class 6) and anthropogenic area (class 7). It characterizes that between 1994 and 1999 was the period of urban expansion in the area. It is possible to distinguish the tendency of this expansion with the conversion of anthropogenic area and exposure soil, both classes characterized by the intention to be converted in urban use (Table 22.1).

22.3.2 Soil Properties

The majority of the soil profiles (8 profiles) are located within the orders of Argilossolos (Ultisols), as well as in the class "Shrub" (6 profiles) (Fig. 22.5).

The organic carbon stock differed between soil orders and different land cover classes, but related well with the geomorphology (Fig. 22.6).

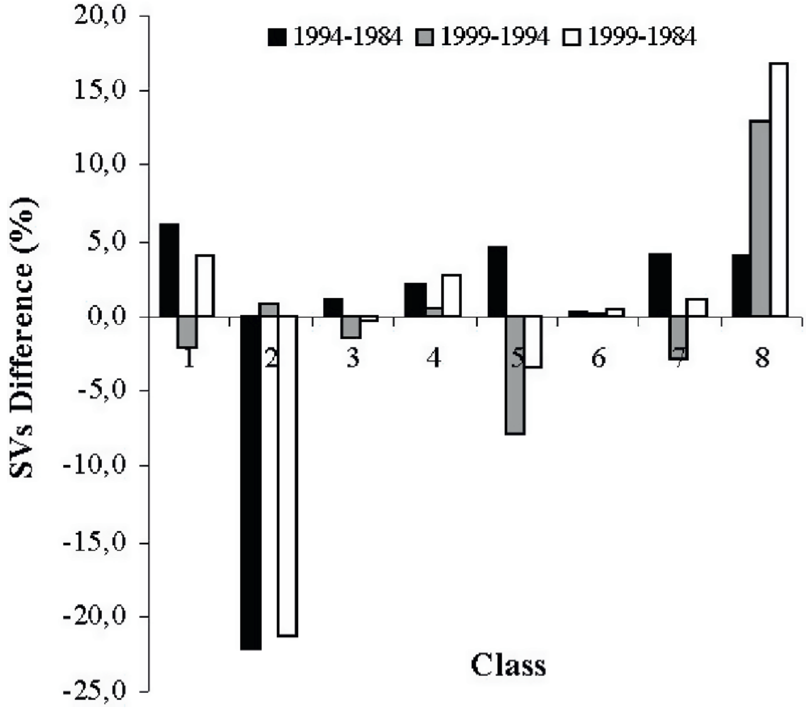


Fig. 22.3 State Vectors difference (%) between the pairs of image **Class 1**: Forest; **Class 2**: Shrub; **Class 3**: Mangrove; **Class 4**: Flooding Area; **Class 5**: Agriculture; **Class 6**: Exposed Soil; **Class 7**: Anthropogenic Area; **Class 8**: Urban Zone

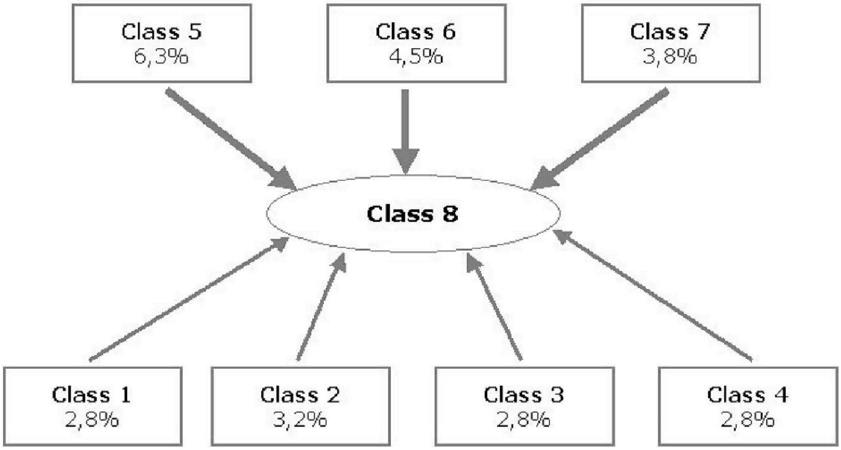


Fig. 22.4 1994–1999 dynamic, with detached to class 8 (Urban zone) **Class 1**: Forest; **Class 2**: Shrub; **Class 3**: Mangrove; **Class 4**: Flooding Area; **Class 5**: Agriculture; **Class 6**: Exposed Soil; **Class 7**: Anthropogenic Area; **Class 8**: Urban Zone

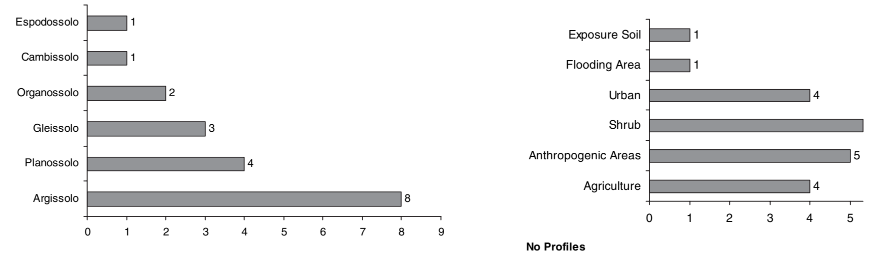


Fig. 22.5 Soils profiles distribution by soil class and land cover Espodossolo = Spodosols; Cambissolo = Inceptisols; Organossolo = Histosols; Gleissolo = Entisols; Planossolo = Alfisols; Argissolo = Ultisols

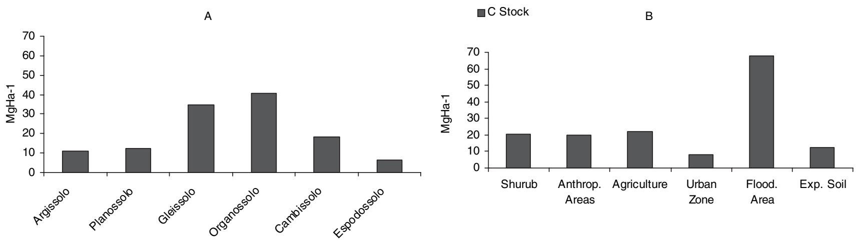


Fig. 22.6 Carbon stocks distribution by soil orders (**A**) and land cover (**B**) Espodossolo = Spodosols; Cambissolo = Inceptisols; Organossolo = Histosols; Gleissolo = Entisols; Planossolo = Alfisols; Argissolo = Ultisols

The "Organossolos" (Histosols) has the highest median value of C stock, followed by "Gleissolos" (Entisols), whereas the lowest values were observed in "Argissolos" (Ultisols) and "Planossolos" (Alfisols) (Fig. 22.7 and Table 22.4). It is not surprising that "Organossolos"and "Gelissolos" contained a high C stock concentration because in this case the soils profiles were located in lowlands, with accumulate organic material, poor drainage and the inhibition of decomposition. Similar results have been found by Tan et al. (2005) with lowest C stock in "Ultisols" soil order and highest in "Entisols".

Although the data are few, it is also possible to observe the influence of geomorphology on C stock levels. The lowlands present the higher median values of C (29, 38 Mgha^{-1}) and the lower number of profiles (7 profiles) while slope presents 14 profiles and the lower median values (21, 86 Mgha^{-1}). Slope and altitude can influence C stock by controlling soil water, soil erosion and geologic deposition processes. Soils profiles GB29 and GB 32, which present the lowest value of C stock in "Gleissolos" (Entisols), are located on steep side slopes, with a favourable decomposition conditions, that is: good drainage.

Similar C stocks were found between "Anthropogenic Areas" and "Shrub" classes, irrespective of the soil orders. The "Urban Zone" showed the lower median and standard deviation suggesting an interruption in organic matter accumulation (Fig. 22.8).

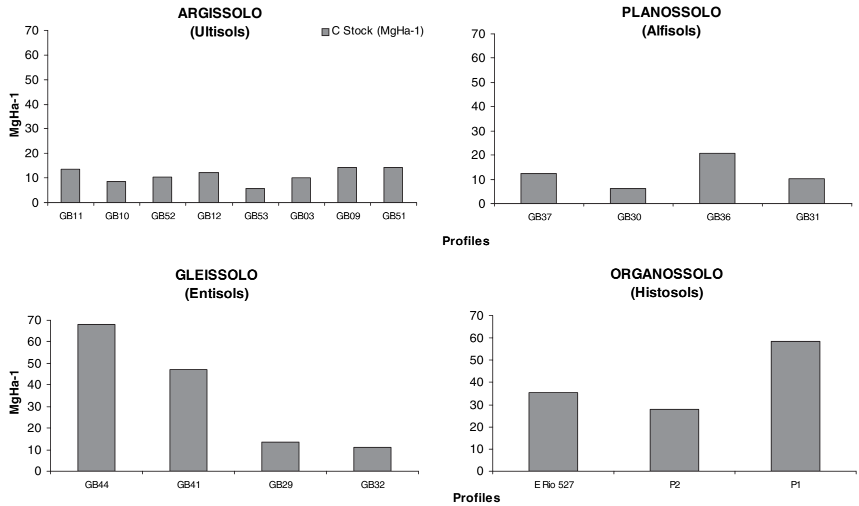


Fig. 22.7 Distribution of the C Stock (0–10 cm depth) in accordance with the soils orders GB11, GB10...

The C stocks in “Organossolos” (Histosols) seems to be related to land cover change, because the highest value found in this soil order – P1 – is under “Shrub” class, whereas the others soil profiles are under “Agricultural” land cover class. This suggests an Organossolo sensitivity to land cover change.

Among four land use systems – Agriculture, Shrub, Anthropogenic areas and Urban Zone – “Argissolos” (Ultisols) contained the lowest C stock.

Many authors have observed the variability of soil properties as a result of soil management (Nascimento et al., 1993; Lepsch et al., 1994; Lutzow et al., 2002). Some basic statistics have been used to show variability. In this study, secondary data soil profiles were used with few samples (21 soil profiles) in a large area (approximately 31 000 ha). Its result in a higher standard variation (17, 46%) with the minimum value of C stocks by 5.92 and the highest by 67.85. The data spatial

Table 22.4 Median and standard deviation of carbon stocks MgHa^{-1} (0–10 cm depth) according soil order and land cover class

Soil Orders				
	Argissolos (Ultisols)	Planossolo (Alfisols) MgHa^{-1}	Gleissolo (Entisols)	Organossolos (Histosols)
Median	11.2	12.4	35.0	40.6
Standard deviation	3.0	6.1	27.35	15.9
Land Cover Class				
	Agriculture	Anthropogenic MgHa^{-1}	Shrub	Urban
Median	21.1	13.5	13.0	6.4
Standard deviation	11.8	14.3	18.7	4.1

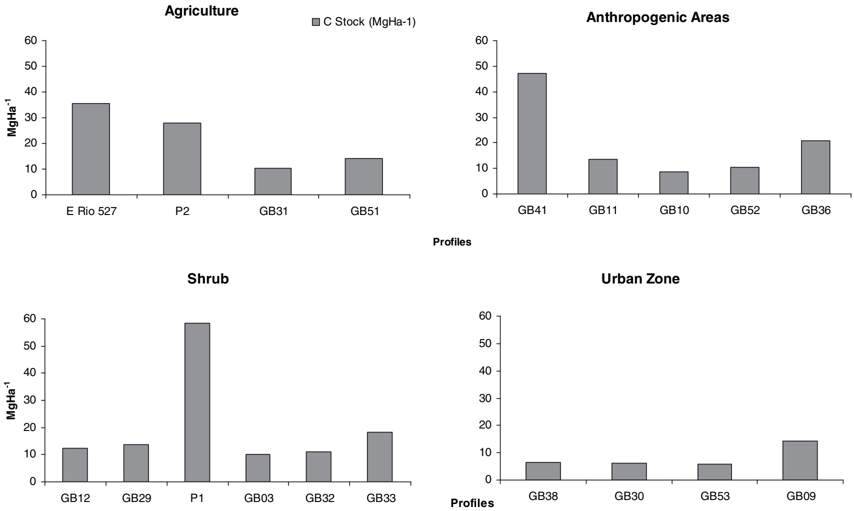


Fig. 22.8 C Stock (0–10 cm depth) distribution under different land covers

distribution was not favourable for using geostatistic tools (similar comments in section 8.1). This question highlights the need to invest in digital soil mapping, considering environmental variables and soil-landscape relationships (the details on digital soil mapping is discussed in Chapter 1).

22.4 Conclusions

The methodological approach was helpful to understanding the landscape dynamics. GIS and modelling techniques allowed the integration of data and information to better characterize landscape dynamics.

The land use change analysis evidenced this sequence of conversion classes: from Forest to Shrub to Anthropogenic Areas/Agriculture and an increment of the urban areas.

The C stock presents variations as in function of soil orders and land cover. The “Organossolos” (Histosols) had the highest C stock independent of the land cover class. The “Argissolos” (Ultisols) and the “Planossolos” (Alfisols) presented the lowest values and also the lowest standard deviation, irrespective of the land cover. The C stock also presented variation by geomorphology, with highest values in lowlands.

“Agriculture”, “Anthropogenic Areas” and “Shrub” had similar C stocks. The “Urban Zone” class had the lower mean, as well as the lower standard deviation, which suggests a relationship between declining values of C stocks with urban land use.

It is important to stress that this work highlighted trends, because it was not possible to establish a correlation between soil orders and soil properties and land cover classes with these dataset. There are few soil profiles in the area and they have been described at different times to the temporal series established in this study (1984–1994–1999). For this kind of study is recommended to use soil data synchronized with the land cover class and temporal series.

However, our results showed that mapping soil properties based on secondary data – soil survey reports – and them cross with land cover dynamic helps to environmental and/or urban planning, because shows the tendency of conversion areas and evidences those with vocations to appropriate uses. This analysis helps differentiate zones in the landscape with different uses and conditions, identifying not only urban and rural settings, but also the interface between the different land cover classes. This approach demonstrates potential to be used in land use policies, because it shows the magnitude of landscape change and its pattern. Its also contribute to build-out modeling to project possible future land use change.

This study reinforces the need for an efficient land use planning, and provides information to support research and planning efforts related to land development and conservation. It represents the first approach integrating satellite imagery with soil properties for studying the landscape dynamic in Rio de Janeiro city.

Acknowledgments This work was supported by National Council for Scientific and Technological Development (CNPq). Special thanks to Rio de Janeiro Rural University (UFRRJ) and Brazilian Agricultural Research Corporation (EMBRAPA Solos).

References

Antrop, M., 2005. Why landscapes of the past are important for the future. Landscape and Urban Planning 70, 21–34.

Burke, I.C., Yonker, C.M., Parton, W.J., Cole, C.V., Flach, K., Schimel, D.S., 1989. Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. Soil Science Society America Journal 53, 800–805.

Câmara G., Souza R.C.M., Freitas U.M., Garrido J. 1996.SPRING: Integrating remote sensing and GIS by object-oriented data modelling. Journal of Computers and Graphics 20(3), 395–403.

Eastman, J. R., Fulk, M. 1993. Long sequence time series evaluation using standardized principal components. Photogrammetric Engineering and Remote Sensing 59(8), 1307–1312.

Fung, T., LeDrew, E. 1987. Application of principal components analysis to change detection. Photogrammetric Engineering and Remote Sensing 53(12), 1649–1658.

Jenny, H., 1980. The Soil Resource: Origin and Behavior. Springer, New York, p. 377.

Jensen, J.R. and Cowen, D.C. 1999. Remote sensing of irban/suburban infrastructure and socio-economic attributes. Photogrammetric Engineering and Remote Sensing 65(5), 611–622.

Kononova, M.M., 1966. Soil Organic Matter, 2nd ed. Pergamon, New York.

Lepsch, I.F., Menk, J.R.F., Oliveira, J.B. de. 1994. Carbon storage and other properties of soils under agriculture and natural vegetation in São Paulo State, Brazil. Soil Use and Management 10, 34–42.

Li, X., Yeh, A.G.O., 2004. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. Landscape Urban Planning 69, 335–354.

Lutzow, M. von, Leifeld, J., Kainz, M., Knabnerkogel, I., Munch, J.C. 2002. Indications for soil organic matter quality in soils under different management. Geoderma 105, 243–258.

Manlay, R.J., Feller, C., Swift, M.J. 2007. Historical evolution of soil organic matter concepts and their relationships with the fertility and sustainability of cropping systems. Agriculture, Ecosystems and Environment, 119, 217–233.
Mansilla Baca, J.F. 2002. Landscape dynamics: analytical methods, classification and simulation models and prognostics similations, by a geoecological view. Ph Thesis. UFRJ, RJ. 185p.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B. 2003. On digital soil mapping. Geoderma 117, 3–52.
Mendonça Santos, M.L. 1999. GIS and Spatio-temporal modelling for the study of alluvial soil and vegetation evolution. Ph Thesis n°1829, École Polytechnique Fédérale de Lausanne, EPFL, Switzerland.
Murage, E.W., Karanja, N.K., Smithson P.C., Woomerb, P.L. 2001. Diagnostic indicators of soil quality in productive and non-productive smallholders' fields of Kenya's Central Highlands. Agriculture, Ecosystems and Environment 79, 1–8.
Nascimento, V.M., Almendros, G., Fernandes, F.M. 1993. Evolution patterns of the soil organic matter in some agricultural systems in the Brazilian "Cerrado" region. European Journal of Soil Biology 29, 177–182.
Santos, H.G. dos, Jacomine, P.K.T., Anjos, L.H.C. dos, Oliveira, V.A. de, Oliveira, J.B. de, Coelho, M.R., Lumbreras, J.F., Cunha, T.J.F. (ed.). Sistema brasileiro de classificação de solos. 2. ed. Rio de Janeiro: Embrapa Solos, 2006. 306 p.
Santos, H.G. dos, Mendonça Santos, M.L., ou Mendonça-Santos, M.L. ou Santos, Oliveira, R. P. de, Lumbreras, J.F., Silva, E.F. da. Banco de dados de solos do estado do Rio de Janeiro, com vistas à modelagem quantitativa de estoque de carbono nos solos In: Congresso Brasileiro de Ciência do Solo, 30 2005, Recife-PE. Anais do XXX Congresso Brasileiro de Ciência do Solo – CD. Sociedade Brasileira de Ciência do Solo.
Schimel, D.S., Braswell, B.H., Holland, E.A., McKeown, R., Ojima, D.S., Painter, T.H., Parton, W.J., Townsend, A.R., 1994. Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. Global Biogeochemistry 8, 279–293.
Swift, M.J., Woomer, P. 1999. Organic matter and sustainability of agricultural systems. In: Mulongo, K., Mercer, R. (ed.). Soil organic matter dynamics ans sustainability of tropical agriculture. London: John Wiley, p. 3–18.
Swift, R.S. 2001. Sequestration of carbon by soil. Soil Science 166, 858–871.
Tan, Z., Lal, R. 2005. Carbon sequestration potential estimates with changes in land use and tillage practice in Ohio, USA. Agriculture, Ecosystems and Environment 111, 140–152.
Turetta, A.P.D., 2004. Evolução espaço-temporal da cobertura do solo das regiões administrativas de Campo Grande e Santa Cruz – uma ferramenta para o planejamento ambiental. Ph. Thesis. Universidade Federal Rural do Rio de Janeiro, RJ, 158p.
Urioste, A.M., Hevia, G.G., Hepper, E.N., Anton, L.E., Bono, A.A., Buschiazzo, D.E. 2006. Cultivation effects on the distribution of organic carbon, total nitrogen and phosphorus in soils of the semiarid region of Argentinian Pampas. Geoderma 136, 621–630.
Yang, X., Lo, C.P., 2003. Modelling urban growth and landscape changes in the Atlanta metropolitan area. International Journal of Geographical Information Science 17, 463–488.
Yu, X.J., Ng, C.N., 2007. Spatial and temporal dynamics of urban sprawl along two urban–rural transects: A case study of Guangzhou, China. Landscape and Urban Planning 79, 96–109.
Zhu, A.X., 1997. A similarity model for representing soil spatial information. Geoderma 77, 217–242.

Chapter 23
Broad-Scale Soil Monitoring Through a Nationwide Soil-Testing Database

B. Lemercier, D. Arrouays, S. Follain, N.P.A. Saby, C. Schvartz and C. Walter

Abstract Spatial variability of soil properties strongly influenced by human activity is not well documented by most soil surveys. Soil tests performed at farmers' request represent a large capital of soil information. In France, the results of a large part of these soil tests are continuously gathered in a unique database, the national soil testing database (named BDAT). The aims of the project were to analyse the evolution of soil features within discrete entities over successive time periods and to test the potential of the BDAT for soil dynamic monitoring. Two illustrations are shown: spatial variability of soil pH at national scale, and evolution of soil phosphorus content at regional scale. A validation by census data on agricultural systems was also tested. Taking into account sampling and statistical bias, databases such as the BDAT appear to be relevant tools for soil properties monitoring and can be helpful for digital soil mapping.

23.1 Introduction

Soil properties like pH, bulk density, organic matter, available phosphorus, exchangeable cations, cation exchange capacity or trace elements are strongly influenced by human activities, in particular agricultural practices e.g. fertilization, amendment supply, soil tillage and crop rotation.

In classical pedological surveys, soil mapping units are generally considered to be homogenous in space and constant across time and do not provide an acute description of the spatial variability of dynamic soil properties. Indeed sampling constraints constitute a bottleneck which is difficult to overcome when spatial and temporal variabilities interact over wide areas. Consequently operational monitoring networks with a sufficient hindsight are few. However, several studies attempted to analyse temporal changes of soil characteristics at small scale: in England and Wales, Bellamy et al. (2005) have studied the evolution of organic carbon status in

B. Lemercier
UMR Sol-Agro et hydrosystème-Spatialisation INRA/Agrocampus Rennes, 65 rue de Saint-Brieuc, CS 84215, F-35042 Rennes Cedex, France
e-mail: Blandine.Lemercier@agrocampus-rennes.fr

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

soils over 25 years, Baxter et al. (2006) have performed a spatiotemporal analysis of topsoil nutrient status and pH also across England and Wales. In Belgium, Lettens et al. (2005) have estimated changes in soil organic carbon stocks between 1960 and 2000 for 289 landscape units, and in New Zealand, Wheeler et al. (2004) have analyzed temporal trends in some soil test data over a 14-years period. Soil testing databases represent a large potential source of information on spatial and temporal variability of soil features.

Results from more than one million of soil tests performed in France since 1990 at farmers' request were stored in a unique database (Saby et al., 2004) named the national soil testing database (BDAT). By aggregating these results within discrete spatial entities, maps of statistical indicators were produced at national and regional scales. Aiming to share these data as specified in Section 24.2, the results are available on a web cartographic server (http://bdat.orleans.inra.fr).

Nationwide representations of statistical parameters derived from the BDAT (pH, total carbonate content, organic carbon content and textural class) obviously discriminate wide pedological domains (Walter et al., 1997; Saby et al., 2004; Lemercier et al., 2006), despite large differences in land use historical evolution and present agricultural systems. The interest of the BDAT for the spatial analysis of soil features variability over large distances is thereby clearly established (Walter et al., 1997). Through continuous new data assimilation, data is now available for large temporal duration and evolution of soil properties may be studied.

The aims of the project were to analyse the evolution of soil features statistics over successive time periods within discrete spatial entities and to test the potential of soil test databases for dynamic soil monitoring. Working hypotheses for the statistical comparison of datasets from different periods in order to detect significant temporal trends were: (i) samples are randomly distributed within the spatial entities and (ii) sampling is independent between the compared periods.

23.2 Material and Methods

23.2.1 The Database

The data were obtained from cultivated topsoil samples and were provided by commercial soil testing laboratories approved by the French Ministry of Agriculture. This agreement ensures the use of normalized analytical methods, necessary to gather data from 34 different laboratories in a unique database. The sampling strategy was unsupervised and this may induce some bias difficult to estimate (Schvartz et al., 1997): only cultivated land is represented, farmers' technical motivations for choosing fields to be sampled and frequency of soil testing are unknown.

Assimilation of new data in the database involves a systematic validation procedure in three steps: (i) computing validation and harmonisation of units, (ii) analytical validation (methods, samples from cultivated topsoils, and estimate of a minimum dataset of soil properties) and (iii) geographical validation (validity of location). Currently the rejection rate was about 16% and the results concerning 1,119,000 topsoil samples from cultivated fields collected between 1990 and 2004 have been validated. A list of 39 raw analytical measurements or derived indicators is available but unequally filled in. An average of 10.5 criteria per sample is completed. Exchangeable potassium, exchangeable magnesium, pH in water and organic carbon are the most systematically accessible parameters with respectively 98%, 97%, 96% and 95% of occurrence.

23.2.2 Space and Time Aggregation

The precise spatial position of the samples was unknown. For statistical analysis and spatial representation data were integrated within discrete spatial entities, the administrative cantons (French mean cantonal area is $140\,km^2$). Non-parametric statistics were estimated for each canton having a minimum of ten samples. This geographical scale appears as a satisfactory compromise between maximization of the number of samples and minimization of parameters variability within a spatial entity. Other databases of interest like agricultural censuses data are built with the same integration level permitting data intersection. Indicators of results accuracy are provided through maps of sampling density and maps of interquartile range per canton.

In order to get a sufficient dataset of samples per canton and to minimize the risk of having two results from the same field, Schvartz et al. (1997) found that working within 5-years periods was relevant. In this study, two periods are available: 1990–1994 and 1995–1999.

A regional database on the same scheme than the national one described above has been established over an extended period in Brittany: results from roughly 300,000 additional samples analysed between 1980 and 2003 are available in this region and therefore 4 periods with sufficient data can be considered for Brittany: 1980–1985, 1990–1994, 1995–1999 and 2000–2003.

23.2.3 Soil Extractable Phosphorus Temporal Evolution

Whole France and two regions Brittany (in north-west France) and Nord-Pas-de-Calais (in north France) were studied. The regions were selected due to sampling constraints (a sufficient samples number is required to compare periods) and assumption that even though soil phosphorus contents are high for both regions, trends in evolution differ.

For Brittany, a cumulative phosphorus balance between 1980 and 2003 related to agricultural activities was calculated per canton. Data from French national agricultural censuses (animal numbers by category and exports by crops) performed in 1978, 1988 and 2000, references from a French group (CORPEN[1])

[1] CORPEN: Comité d'ORientation pour des Pratiques agricoles respectueuses del'ENvironnement (Steering Committee for Environmentally Friendly Farming Practices)

studying agricultural practices, and statistical data on departmental deliveries of phosphorus mineral fertilizers were used.

23.3 Results

23.3.1 Quantification of Spatial Variability of pH Values for Topsoil Horizons

Spatial aggregation of pH results highlights large-range spatial structures. These spatial structures remain stable from one period to another (data not shown). For instance, the map of pH median values (Fig. 23.1) reveals marked regional trends. Wide domains of high or low pH are identified and could be related to geological substratum. Soils of central France, north-west, north-east and south-west France developed on crystalline bedrocks are mainly acids whereas soils formed from limestone materials of the north-east France, the Mediterranean area and the central-west zone are alkaline.

The intra-cantonal variability appears clearly correlated with the median value: the more the median pH is close to neutrality, the more the dispersion within the

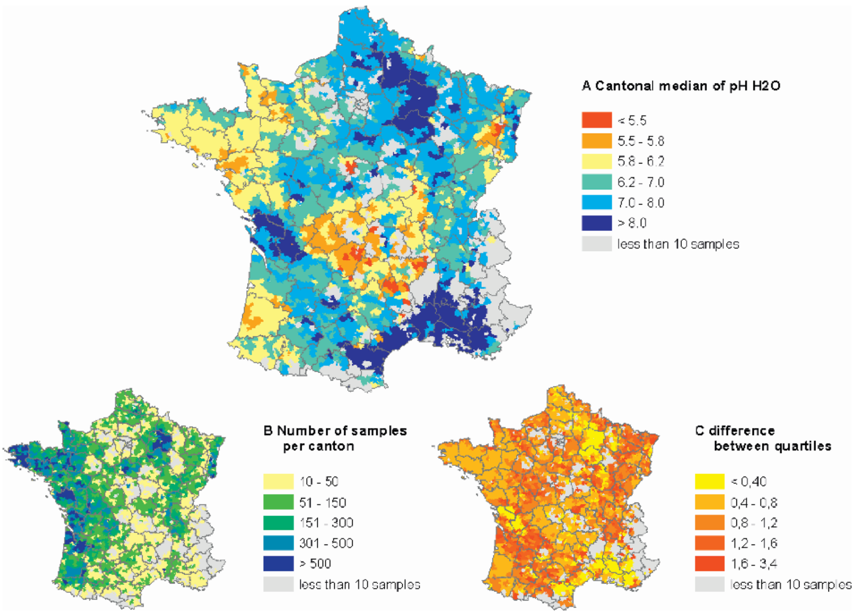


Fig. 23.1 Maps of cantonal statistics of pH in water of cultivated topsoil for the period 1995–1999: **(A)** cantonal median value; **(B)** number of samples per canton; **(C)** inter-quartile value (See also Plate 29 in the Colour Plate Section)

canton is large (Fig. 23.1C). This can be interpreted as the result of soil heterogeneity within the canton or as the effect of variable liming practices.

23.3.2 Detecting Trends in Evolution of Soil Extractable Phosphorus

At national scale, temporal trends of soil parameters are hidden by the heterogeneity of local situations and therefore we need to work at the regional scale to detect evolutions and to assess their signification by statistical means. In addition, human factors determining the value and the evolution of soil parameters are most often managed at this level.

Soil extractable phosphorus content is significantly higher for the regions Nord-Pas-de-Calais and Brittany than for whole France (Table 23.1). Analytical methods being different according to the region, comparing contents is not relevant. Nevertheless, the comparison of the evolution of extractable phosphorus cantonal statistics according to geographical area shows that extractable phosphorus content increased in France, decreased in Nord-Pas-de Calais and increased in Brittany in a greater extent than for whole France. In Nord-Pas-de-Calais, median value of extractable phosphorus content decreased significantly for 31% of the cantons and increased for 5% of them. At the opposite, this parameter increased significantly for 42% of the cantons and decrease for 7% of them in Brittany. This phosphorus increase in Brittany is even more significant when the data from 1980 to 1985 are compared with the data from 1995 to 1999: cantonal median value of extractable phosphorus content increased by 20% and this evolution was significant for 64% of the cantons.

23.3.3 Validation Through Census Data on Agricultural Systems

By linking the soil testing database to an independent source of information on agricultural practices, we attempted to validate the detected temporal trends. A correlation exists between soil extractable phosphorus content evolution and cumulative balance of phosphorus due to agricultural activity in Brittany (Fig. 23.2). The phosphorus enrichment increased proportionally with the cumulative balance, mainly for classes of balances higher than 1200 kgP ha^{-1}.

23.4 Discussion

For most of the soil parameters stored in the BDAT, large-range spatial structures very stable over time were observed. This fact underlines the robustness of the methodology used. Thus such legacy soil data easily available at low cost provides a global view at small scale of the spatial variability of pedological parameters. In several regions of France, this database is presently the unique available soil

Table 23.1 Statistics of extractable phosphorus content per canton and evolution between 1990–1994 and 1995–1999, for whole France and for two regions (Brittany and Nord-Pas de Calais). Two analytical methods were used: Dyer (NF X 31–160) and Joret-Hébert (NF X 31–161)

	P analytical methods	1995–99 P content (mgP kg^{-1}) and relative evolution between 1990–1994 and 1995–1999			Percentage of cantons with a significant evolution*		Regression slope of the medians (with O intercept)
		First quartile	Median	Third quartile	Increase	Decrease	
France	Dyer	68 (+11%)	97 (+5%)	138 (+6%)	24%	10%	1.01
	Joret-Hébert	51 (+11%)	72 (+4%)	102 (−2%)	22%	15%	0.95
Nord-Pas-de-Calais	Joret-Hébert	100 (−10%)	124 (−7%)	143 (−9%)	5%	31%	0.90
Brittany	Dyer	120 (+3%)	155 (+8%)	191 (+11%)	42%	7%	1.08
		1995–1999 P content (mgP kg^{-1}) and relative evolution between 1980–1985 and 1995–1999					
Brittany	Dyer	120 (+13%)	155 (+20%)	191 (+20%)	64%	6%	1.17

*Wilcoxon non-parametric test, one sided, $p = 0.05$

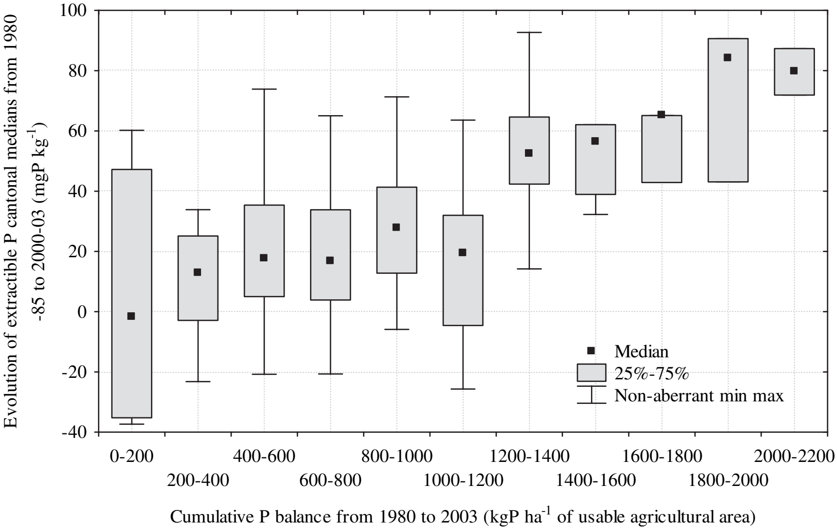


Fig. 23.2 Evolution of extractable phosphorus cantonal median values of cultivated topsoil from 1980–1985 to 2000–2003 related to classes of cantonal cumulative phosphorus balances from 1980 to 2003

information, except for the 1:1,000,000 scale European soil map. Thereby, this kind of database could be of interest for regions or countries with limited soil data infrastructures.

With regard to temporal trends, the database allowed highlighting significant evolution at regional level and can be used for medium to long term soil monitoring. However, only well marked changes can be detected, as temporal variations are usually smaller than spatial ones and therefore more difficult to detect in a statistically significant manner. As an illustration, in Brittany, the interdecile of cantonal medians was 112 mg kg^{-1} within the period 1995–1999, whereas the mean variation of median cantonal P content was only 29 mg kg^{-1} between 1990–1994 and 1995–1999.

However, care is needed when interpreting the results from such a soil testing database. The uncontrolled sampling strategy can involve heterogeneity in spatial distribution of samples and does not consider farmers' motivations for sampling a given field for soil testing. The statistical bias generated is difficult to assess even if partly offset by the great number of samples. To test the initial hypotheses regarding the sampling strategy (random spatial distribution, independent time series), precise georeferencing of the soil samples would be very helpful. Progressively, laboratories tend to provide such georeferenced data. In a near future, we may expect to have a sufficient number of precisely positioned data to be able to give up the administrative spatial support. Finer approaches like geostatistics could be then developed and digital maps of soil properties produced. Such a methodology is presented in Chapter 27.

Nevertheless, soil testing databases mostly provide a partial view of the soil cover because only chemical properties are considered and only topsoil horizons of cultivated land are sampled. Legacy data on agricultural topsoil could be integrated with existing soil surveys and human and natural factors which influence soil properties such as climate, geology, geomorphology (s.c.o.r.p.a.n terms of McBratney et al., 2003). The relevance of the digital soil maps in terms of spatial variability and evolution of soil dynamic characteristics should be improved.

23.5 Conclusion

We showed that legacy data initially obtained neither for mapping nor for monitoring purposes can be used to study the spatiotemporal variability of soil parameters. Gathering a large collection of validated data, the BDAT represents a key tool to assess the effect of human activity on spatial and temporal variability of soil. Indeed, marked regional trends but also intra-cantonal variability were underlined by aggregating spatially pH analytical results from the BDAT within discrete spatial entities. Different evolution dynamics of the extractable phosphorus status were identified for two regions, and the increase of soil phosphorus content observed in Brittany was validated by the comparison with agricultural practices. Using soil testing databases can be helpful for soil properties monitoring through wide areas (regions or countries) where other soil information is sparse. Finally, the combination of soil test results with existing soil surveys and ancillary data giving information on soil and soil properties variability could be a very interesting contribution of this kind of database to digital soil mapping.

Acknowledgments The BDAT programme benefits financial support from the French Ministry of Agriculture. We would like to thank the commercial laboratories for data transmission.

References

Baxter, S.J., Oliver, M.A., Archer, J.R., 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 and Management, 22, 383–392.

Bellamy, P.H., Loveland, P.J., Bradley, R.I., Lark, R.M., Kirk, G.J.D., 2005. Carbon losses from all soils across England and Wales 1978–2003. Nature, 437, 245–248.

Lemercier, B., Walter, C., Schvatz, C., Saby, N., Arrouays, D., Follain, S., 2006. Suivi des teneurs en carbone organique et en phosphore extractible dans les sols agricoles de trois régions françaises. Analyses à partir de la Base de Données des Analyses de Terre. Etude et Gestion des Sols, 13(3), 165–180 (in French).

Lettens, S., Orshoven, J., Wesemael, B., Muys, B., Perrin, D., 2005. Soil organic carbon changes in landscape units of Belgium between 1960 and 2000 with reference to 1990. Global Change Biology, 11(12), 2128–2140.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma, 117, 3–52.

Saby, N., Schvartz, C., Walter, C., Arrouays, D., Lemercier, B., Roland, N., Squividant, H., 2004. Base de Données des Analyses de Terre : Procédure de collecte et résultats de la campagne 1995–2000. Etude et gestion des sols, 11(4), 235–253 (in French).

Schvartz, C., Walter, C., Claudot, B., Bouédo, T., Aurousseau, P., 1997. Synthèse nationale des analyses de terre réalisées entre 1990 et 1994. 1. Constitution d'une banque de données cantonale. Etude et Gestion des Sols, 4(3), 191–204 (in French).

Walter, C., Schvartz, C., Claudot, B., Bouédo, T., Aurousseau, P., 1997. Synthèse nationale des analyses de terre réalisées entre 1990 et 1994. 2. Descriptions statistique et cartographique de la variabilité des horizons de surface des sols cultivés. Etude et Gestion des Sols, 4(3), 205–220 (in French).

Wheeler, D.M., Sparling, G.P., Roberts, A.H.C., 2004. Trends in some soil test data over a 14-years period in New Zealand. New Zealand Journal of Agriculture Research, 47, 155–166.

Chapter 24
Online Soil Information Systems – Recent Australian Experience

N.J. McKenzie, D.W. Jacquier and L.J. Gregory

Abstract Australian agencies are starting to provide online access to soil information through the Australian Soil Resource Information System (ASRIS – www.asris.csiro.au). ASRIS has been designed to integrate soil information collected using both conventional and digital methods. Here we review our experience in developing the system and focus on the importance of good standards for data collection and exchange. There is a clear need for an international standard (in the form of a GML schema) to enable efficient exchange of soil data. We also comment on the problem of market failure and its affect on investment in soil information.

24.1 Introduction

Australian agencies are starting to provide online access to the best available information on soil and land resources in a consistent format. This is being achieved through the Australian Soil Resource Information System (ASRIS – www.asris.csiro.au). ASRIS has been designed to integrate soil information collected using both conventional and digital methods. The system is described by McKenzie et al. (2005, 2007). Here we review our experience in preparing standards and securing institutional support because these are fundamental to a successful national system.

24.2 Online Soil Information Systems

A premise of most soil information systems is that users want to maximize the value of past and future investments in computing systems and data. OGC (2005) identify three needs that follow from this:

- the need to share and reuse data in order to decrease costs, get more or better information, and increase the value of data holdings

N.J. McKenzie
CSIRO Land and Water, G.P.O. Box 1666, Canberra, ACT 2602, Australia
e-mail: Neil.McKenzie@csiro.au

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

- the need to choose the best tool for the job, and the related need of reducing technology and procurement risk (i.e. avoiding reliance on a single software company)
- the need for more people with less training to benefit from using soil information for a wide range of purposes.

The current revolution in digital communication depends heavily on standards for software and data. Responsibilities for software reside with specialists in information technology, while for data they rest, in our case, with soil scientists involved in digital soil mapping.

In the past, the requirements of soil classification systems provided a motivation for collecting consistent soil data according to defined standards. For example, many countries adopted aspects of Soil Taxonomy and modified their systems for data collection and storage accordingly. The World Reference Base is having a similar, albeit less dramatic effect.

There is now much less emphasis on soil classification. Most soil information-systems for land evaluation provide estimates of the primary variables that characterize how a soil functions (e.g. nutrient supply, plant available water capacity, potential for erosion, transport of contaminants). While many agencies are changing how soil information is collected and analysed, there has been limited discussion and action on how we move towards consistent methods for sampling, measuring and reporting. If this is not addressed as a matter of urgency, we will not be able to fully exploit the potential of online systems for soil information. Most importantly, potential users of soil information will continue to make decisions on land planning and management without regard to soil factors.

24.3 Australian Soil Resource Information System

The Australian Soil Resource Information System (ASRIS) has been developed to provide primary data on soil and land to meet the demands of a broad range of users including natural resource managers, teachers, planners, researchers, and community groups. The online system provides access to the best available soil and land resource information in a consistent format across the country – the level of detail depends on the survey coverage in each region. More specifically, ASRIS provides the following.

- *A spatial hierarchy of land units*. This has seven main levels of generalization. The upper three levels provide descriptions of soils and landscapes across the complete continent while the lower levels provide more detailed information, particularly on soil properties, for areas where mapping has been completed. The lowest level relates to an individual site in the field.
- *A consistent set of land qualities*. These are described for each land unit. Descriptions from the lowest level are used to generate summaries for higher-level units. The land qualities relate to the intrinsic capability of land to support various

uses – the land qualities relate to soil thickness, water storage, permeability, fertility, salinity and erodibility.

- *A soil profile database*. These fully characterized sites are representative of significant areas and environments.
- *Estimates of uncertainty*. These are provided with most data held within ASRIS. A distinction is made between attribute uncertainty (due to the measurement or estimation procedure for a given soil material) and spatial uncertainty (due to the natural variation across a landscape). The estimates are provided to encourage formal analysis of the uncertainty of predictions generated using ASRIS data (e.g. crop yield, runoff, land suitability for a range of purposes).

ASRIS is being released in stages. The upper levels of the hierarchy have been recently completed for the whole country. There will be a restricted coverage at lower levels. Data will also be available for several thousand representative profiles. Most data in ASRIS are from conventional surveys so vector and point formats are common. Grid formats are used for environmental data (e.g. remotely sensed images, digital terrain variables) and some soil properties including those described by Johnston et al. (2003). The intention is to eventually provide fine-grain grids with estimates for the most important functional properties of soils. This will make the system consistent with the proposed global soil information system (www.globalsoilmap.net).

At present, ASRIS provides access to soil information with a minimum of interpretation. This is useful for those with technical training. However, the real benefits of the system will become apparent when a series of interpreted views of the data are prepared. Priorities for the coming two years include views relating to soil acidification (time to critical pH, lime requirement), soil erosion by wind and water, and soil carbon.

24.4 Lessons to Date on Measurement and Mapping

Australia has well-established protocols for some aspects of soil survey. McDonald et al. (1990) provide standard field methods for describing landform, vegetation and soil morphology. Laboratory methods for soil chemistry (Rayment and Higginson, 1992) are widely used but protocols for soil physical measurement are more recent (McKenzie et al., 2002). There has been less consistency in the method of soil survey (e.g. Beckett and Bie, 1978; Gunn et al., 1988). Most agencies use some form of integrated survey but there are many subtle differences in field sampling, mapping technique, approaches to classification and land evaluation. These differences are reflected in data structures and database designs of the main state and territory agencies responsible for soil survey. With this background, it was inevitable that a range of technical problems would be encountered during the design and implementation of ASRIS. The most significant to date are as follows.

- Existing protocols did not specify minimum data sets for soil profile description and laboratory characterization. Agencies have data sets with particular strengths and weaknesses (e.g. strong on plant nutrition but weak on soil physics) but none are comprehensive.

- The soil water regime is central to most aspects of land evaluation but few agencies have even rudimentary data on soil hydraulic properties and bulk density. This is partly due to the lack of protocols prior to 2002 when survey programs were most active.
- The existing guidelines for survey provide minimal direction on sampling. Few surveys record the purpose and method of sampling and this creates many problems when combining data sets. For example, some agencies sample relatively natural sites (e.g. roadside reserves) whereas others sample soils with land use typical of the region. Reliable estimation of, for example, soil organic carbon and pH at a regional scale has not been possible as a result. Statistically based sampling has been rare and there are multiple sources of bias in the state, territory and national databases.
- Each agency has developed its own classification system for soil profiles and mapping units. Only a few agencies have formal systems for correlation so the task of reconciling mapping is difficult and in some regions it is more efficient to start again.
- While most technical specialists involved in survey have a good appreciation of soil variation, this has not been recorded effectively. The uncertainties of outputs from surveys have not been communicated.

We have addressed these problems in several ways. Most of our effort has been directed to the Technical Specifications for ASRIS (McKenzie et al., 2005). As noted earlier, these define mapping units, a consistent set of land qualities, estimation procedures, and the beginnings of a system for stating uncertainty. In the longer term, survey practice must change with digital soil mapping becoming the norm. A significant step towards this is publication of new guidelines for surveying soil and land resources in Australia (McKenzie et al., 2008).

24.5 The Promise of Interoperable Systems

Most software for web browsing and managing spatial data can now process text encoded in the eXtensible Mark-up Language (XML). XML is a language for creating self-describing data files. These files have headers with instructions on how to interpret data that follow the header. Scores of industries and professional domains have developed XML schemas (schemas are essentially formats) to share information between organizations with diverse information systems. The Open Geospatial Consortium (OGC) has since developed the Geography Markup Language (GML) and this is becoming the standard XML encoding for geospatial information. GML separates content from presentation and the latter is entirely under program control for individual devices.

GML makes it possible to resolve many of the difficulties caused by incompatible data models that are typically a feature of local, regional and national soil information systems. The task for each custodian is to map their data to an agreed GML schema for soil information. One-to-one mapping between all data models is unworkable when thousands of models are involved. GML enables a many-to-one solution.

The task for those involved with digital soil mapping is to first agree on a GML schema for soil information. Each participating agency must then map their data to the agreed GML schema and expose this to the rest of the world through a web map service, web feature service, or web coverage service. This enables interactive exchange of soil information.

With ASRIS we have gone part of the way in developing a standard data model for soil attributes that is agreed between contributing state and territory agencies in Australia. We are yet to develop standards for vector and raster data. At present, data from state and territory agencies are imported to ASRIS via several means (e.g. spreadsheets, database tables, GIS coverages) and considerable effort is needed to manipulate it into the required format. Our long-term goal is to have an interoperable system where new data are imported automatically. Achieving this requires a GML schema for soil data.

In our view, it is imperative for the soil science community to agree on an international GML schema for soil data. This is a critical step in developing the proposed global soil information system (www.globalsoilmap.net). The GML schema will avoid duplication and provide many benefits. Most importantly, it will ensure soil information is readily accessible to other disciplines.

24.6 Barriers to Investment

Online soil information systems place new technical demands on soil survey agencies. The systems are also expensive to establish and require secure investment over the long term. Contrast this with conventional surveys where funds are required primarily at the time of survey. Obtaining the necessary funds for developing and maintaining online systems is a major challenge that is made more difficult by several institutional barriers to investment.

It is widely claimed by soil scientists that investment in soil and land resource information is too small. The resulting economic and governance costs include direct ones such as wasted expenditure on land development, forgone development, damage to infrastructure, environmental rehabilitation, litigation and compensation, and indirect costs of resource degradation including loss of ecosystem services. Various studies report very large benefit-cost ratios for land resource information (e.g. Olson and Marshall, 1968; Hallsworth, 1978; Klingebiel, 1996). Benefit-cost ratios often exceed 40:1 and in some instances exceed 100:1 (e.g. Australian studies by ACIL, 1996).

In Australia there has been a significant increase in demand for soil information amongst industry and government (NLWRA, 2002). Despite this, public investment in land resource survey has declined. There appears to be a substantial disconnection between demand and supply. We contend that *market failure*[1] is distorting both public and private investment in soil and land resource information regardless of whether conventional or digital methods are used. This is supported by the following observations.

- Consumers of soil information are not able to weigh up what they are getting for the price (lost production is not seen, degradation is often insidious).
- The beneficiaries of soil information are many and varied but they rarely pay for the information directly – online systems provide information at minimal cost.
- There is often a mismatch between the time needed to gather appropriate data (years) and the time scale of the decision-making process (days to months).
- Soil information has a long life and the stream of benefits is unpredictable – at the time of investment, all costs are known but benefits are not.
- Some major benefits only accrue when survey coverage is complete.

Several other observations relate specifically to the private sector.

- The efficiency of private survey at the enterprise scale (e.g. farm or plantation) is greatly enhanced by regional-scale information because it provides context and useful environmental data.
- The benefit from soil and land information is greatly enhanced when it contributes to a broader regional view – individual surveys and monitoring generate greater benefit when they fit together.

The absence of clear market signals to investors has resulted in governments at all levels failing to have coherent strategies for investing in natural resource information. Similarly, private sector work is fragmented and the substantial amount of information collected by private companies is poorly utilized. Public investment in Australia, as in most countries, is strongly influenced by neo-classical economics. We need to understand the underlying theory and principles if we are to attract the investment that most soil scientists contend is necessary. This requires an effective dialogue with decision makers. Edwards (2002) summarizes the challenge: '. . .to convince economists that government should take the running on any given issue, you must first convince them that there is market failure and, second, that government would actually make things better rather than worse'.

24.7 Conclusions

Many of the challenges encountered during the development of ASRIS are likely to be relevant to other national systems. Implementing digital soil mapping and taking full advantage of online GIS will not happen automatically. It requires strong demand from users of information, an even stronger commitment from individuals

[1] Market failure is a technical term used by economists to describe situations where the market (in this case, for soil information) when left to itself does not allocate resources efficiently.

and organizations with an interest in providing soil information, and substantial investment. Online soil information systems require much stronger discipline in the collection and management of soil data than soil scientists have shown to date. This is particularly challenging because methods of survey are in transition and the new methods for digital soil mapping are diverse. While the eventual result may be conceptually simple (e.g. functional soil properties predicted at points in three-dimensional space and through time), the current mix of old and new data is complex. A priority for the digital soil mapping community is to establish an agreed GML schema for soil data. Finally, ASRIS has gained wide support in Australia because it is solving problems common to all contributing agencies. Most notably, it provides an efficient way of supplying soil information to users across the country. This in turn provides a powerful incentive for coordination collaboration, and investment.

References

ACIL, 1996. The development of an economic framework and instruments for assessing the benefits and costs of land resource assessment in Australia. Final Report to the Australian Collaborative Land Evaluation Program, CSIRO Division of Soils, Canberra.

Beckett, P.H.T., Bie, S.W., 1978. Use of soil and land system maps to provide soil information in Australia. CSIRO Aust. Div Soils Tech. Paper No. 33.

Edwards, L., 2002. How to argue with an economist – reopening political debate in Australia. Cambridge University Press, Cambridge.

Gunn, R.H., Beattie, J.A., Reid, R.E., van de Graaff, R.H.M., 1988. Australian soil and land survey handbook: guidelines for conducting surveys. Inkata Press, Melbourne.

Hallsworth, E.G., 1978. Benefits and costs of land resource survey and evaluation. Commonwealth and State Government Collaborative Soil Conservation Study 1975–1977, Report 5, AGPS, Canberra.

Johnston, R.M., Barry, S.J., Bleys, E., Bui, E.N., Moran, C.J., Simon, D.A.P., Carlile, P., McKenzie, N.J., Henderson, B.L., Chapman, G., Imhof, M., Maschmedt, D., Howe, D., Grose, C., Schoknecht, N., Powell, B., Grundy, M., 2003. ASRIS – the database. Australian Journal of Soil Research 41, 1021–1036.

Klingebiel, A.A., 1996. Cost and returns of soil surveys. Soil Conservation 32, 3–6.

McDonald, R.C., Isbell, R.F., Speight, J.G., Walker, J., Hopkins, M.S., 1990. Australian soil and land survey field handbook, 2nd Edn. Inkata Press, Melbourne.

McKenzie, N.J., Coughlan, K., Cresswell, H.P., 2002. Soil physical measurement and interpretation for land evaluation. Australian Soil and Land Survey Handbook Series Vol. 5. CSIRO Publishing, Melbourne.

McKenzie, N.J., Jacquier, D.W., Maschmedt, D.J., Griffin, E.A., Brough, D.M., 2005. The Australian Soil Resource Information System: technical specifications. National Committee on Soil and Terrain Information/Australian Collaborative Land Evaluation Program, Canberra (www.asris.csiro.au)

McKenzie, N.J., Grundy, M.J., Webster, R., Ringrose-Voase, A., 2008. Guidelines for surveying soil and land resources (2nd edn). Australian Soil and Land Survey Handbook Vol. 2. CSIRO Publishing, Melbourne.

NLWRA, 2002. Australia's Natural Resources: 1997–2002 and beyond. National Land and Water Resources Audit, Canberra.

OGC, 2005. The importance of going 'open'. Open Geospatial Consortium White Paper, July 2005 (www.opengeospatial.org).

Olson, G.W., Marshall, R.L., 1968. Using high-intensity soil surveys for big development projects: a Cornell experience. Soil Science 105, 223–231.
Rayment, G.E., Higginson, F.R., 1992. Australian laboratory handbook of soil and water chemical methods. Inkata Press, Melbourne.

Chapter 25
Digital Soil Mapping Using Legacy Data in the Eden Valley, UK

T.R. Mayr, R.C. Palmer and H.J. Cooke

Abstract The National Soil Resources Institute has a considerable amount of legacy data in the form of auger bore observations and detailed soil maps. Both have limitations due to inconsistencies in mapping, extent and spatial distribution of the data. Expert knowledge and quality assessment of the inference model can be used to analyse the available training data as well as the resulting map to identify shortcomings. Expert knowledge will identify soils which are either under predicted or missing from the training dataset, whereas the quality assessment will identify soils and landscape units that are missing from the training data. In addition, the methodology provides the means to assess accurately the number and locations of any additional samples required. Using this framework, legacy data can be a valuable source of information in Digital Soil Mapping.

25.1 Introduction

The National Soil Resources Institute (NSRI) has a considerable legacy of traditional soil maps and point observations following 75 years of soil mapping in England and Wales. The National Soil Map (NATMAP), representing a five-year mapping programme (Soil Survey of England and Wales, 1983), contributed 150 000 geo-referenced auger bore descriptions held electronically in the NSRI computerised Land Information System (LandIS). These soil observations are either transects across the landscape or clustered within the sample farms that were investigated during the mapping phase for NATMAP.

Although only 25% of the area of England and Wales has been mapped in detail, there are examples of detailed soil mapping for most soil landscapes in the two countries. The 117 mapped areas in the 1:25 000 scale mapping programme (1968–1980) were chosen for their pedological, geomorphological and agricultural interest and to provide an understanding of soil patterns in all landscapes for future

T.R. Mayr
Cranfield University, National Soil Resources Institute, Building 37, Cranfield MK43 0AL, UK
e-mail: t.mayr@cranfield.ac.uk

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

county, regional and national soil maps (Findlay, 1970). These areas correspond to the Ordnance Survey 1:25 000 scale map series. They provide an excellent basis for developing digital soil mapping methods (McBratney et al., 2003) as a means of extending the current coverage of detailed soil maps in England and Wales (see also Chapter 6).

NATMAP auger bore data and, to a lesser extent, the detailed maps incorporate inconsistencies derived from the methods of mapping employed by the individual soil surveyors and the purpose for which the data were collected. This paper examines potential solutions to allow for these inconsistencies and suggests how diverse data can be most effectively used in digital soil mapping.

25.2 Materials and Methods

Legacy data (soil maps and point data) have been used in the past to derive soil-landscape relationships. In the first approach (Mayr et al., 2001), map legends and map unit definitions were taken to represent the mental models used by soil surveyors in the field. A series of rules were devised linking the soil map units to environmental spatial data which would have been followed by the surveyor. Once formulated, these rules can be used to guide subsequent re-interpretations of spatial information for the same area or can be used in similar landscapes elsewhere (Bui et al., 1999). This approach was essentially akin to those advocated for soil resource inventories by Thompson and Beckman (1959) and Favrot (1989) and has also been previously reported by Mayr and Palmer (2006). In the second approach (Mayr et al., 2006) a soil-landscape model was developed, which incorporated similar decision making processes to those used in the field by a soil surveyor. This approach depended on existing point observation data accurately reflecting the spatial distribution of soils in the landscape. Some of the data were biased by the sampling strategy used by the soil surveyor and some methods of coping with this bias are discussed in this paper.

25.2.1 Study Site

The Eden, rising on the western flanks of the Cumbrian Pennines, south of Kirkby Stephen, flows north westwards towards Carlisle before discharging to the Solway Firth. The area south of Lazonby and the Eamont confluence has been chosen for the study, which extends to approximately 689 km^2 (Fig. 25.1). The north-eastern part of the catchment is drained by short, relatively steep streams from the Pennines; the south-western part includes tributaries of the Lyvennet system which arise on Ravensworth Fell and headstreams of the Eden originating on Mailerstang Common and passing through Kirkby Stephen. The watershed along the Pennines scarp rises to over 1000 m OD, while the river falls to about 70 m OD near Lazonby. Rainfall exceeds 2000 mm per annum on the eastern watersheds, with a field capacity period of over 11 months and an accumulated temperature of less than 750 day degrees

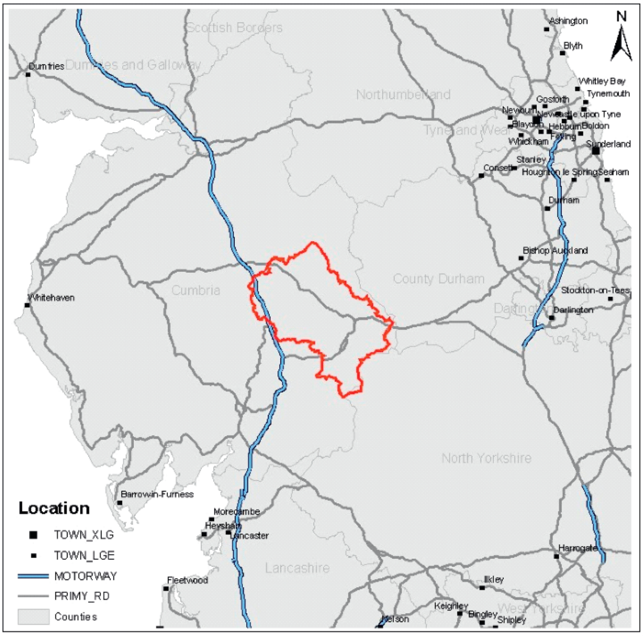


Fig. 25.1 Location map (See also Plate 30 in the Colour Plate Section)

above 0 °C. The lower valley has about 850 mm annual rainfall, a field capacity period of 215 days and an accumulated temperature of about 1300 day degrees above 0 °C.

25.2.2 Covariates

The small-scale soil map (1:250 000 scale NATMAP), geology (BGS), lithology (BGS) – for further discussion see Chapter 14, 5 m DTM (Nextmap), as well as a bioclimatic map (1:625 000 scale) were available for the project. Extensive DTM analysis (see also Chapter 10) was carried out using TAPES, TOPAZ and LandMapR software programs as well as custom-written applications for relief intensity and texture. In total, 76 evidence layers were available for modelling. Entropy reduction was used to reduce the dimensionality of often very highly correlated datasets.

25.2.3 Training Data

All auger bore observations were compiled in an EXCEL spreadsheet using information from LandIS and from auger bore observations archived as paper records. As the prime objective of this methodology was to map soil classes, all observations that had not been ascribed to a soil series were discarded. Also, as some of the auger bore information pre-dated NATMAP, some of the data did not incorporate the soil series rationalisation that was carried out as part of the NATMAP programme in order to eliminate significant inter-regional correlation problems (Hollis and Avery, 1997).

All soil series identified within the auger bore records have therefore been correlated and re-named, where necessary, to fit the current soil classification protocols (Avery, 1973; Avery, 1980; Clayden & Hollis, 1984). The final number of observations available for modelling was 2400. The training dataset was completed by attaching the covariates using Hawth's Analysis Tool for ArcGIS (www.spatialecology.com).

25.2.4 Expert Knowledge

As NATMAP was produced during a very short period of time it forms a coherent dataset due to consistent series definitions and mapping guidelines. In addition, it is the only soil map which covers all of England and Wales and consequently provides a very convenient framework in which digital soil mapping can be undertaken. Therefore, NATMAP map units play an integral part in any digital soil mapping approach.

In order to minimise bias in the training datasets, it was important that the local composition of all NATMAP units were defined as accurately as possible in terms of the areal extent of each constituent soil series. In order to describe the thematic space, the following information was assembled for each NATMAP soil association:

- lead and constituent soil series from national NATMAP legend;
- national average proportions of constituent soil series from LandIS;
- listing of other soil series mentioned in Regional Bulletins when describing NATMAP unit;
- identification of all auger bores recorded within the association within the catchment and its buffer;
- identification of all National Soil Inventory observations within that association nationally;
- identification of all soil series mapped within the association on detailed 1:50,000 scale soil maps that have been mapped and published since the production of NATMAP.

25.2.5 Inference

Bayesian Belief Networks emerged as the primary tool as it offered a flexible approach to reducing the uncertainty associated with soil survey. It used the best available evidence of soil variation for the area, and adopted the best current knowledge of its meaning. The procedure was analogous to conventional soil survey in that it

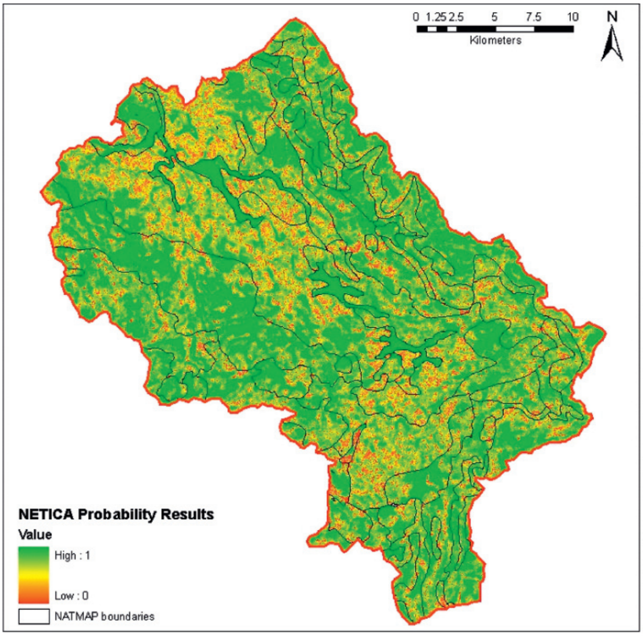


Fig. 25.2 Model quality assessment (See also Plate 31 in the Colour Plate Section)

used a wide range of disparate evidence according to prior knowledge of its meaning. It showed a similar flexibility in that it adapted to conditions of varying availability of evidence. In contrast to conventional soil survey, the rule-based method provided a numerical estimate of the probability of occurrence for a given attribute as it varied across the study area, by means of explicit associations with the evidence (Fig. 25.2).

In addition, the software provided mechanisms for incorporating expert knowledge by manipulating both the *a priori* NATMAP composition as well as the *a priori* distribution of soil series, both based on the training data. The first was achieved by addressing the probabilistic relationships between individual soil series and NATMAP units; the second by addressing the spatial extent of individual soil series.

25.2.6 Quality Assessment

In order to assess independently the results from any inference system beyond the internal success rate based on entropy reduction, three tests were undertaken:

a) Spatial assessment:
   The spatial support was assessed on the density and spatial distribution of the auger bores that were available for incorporation into the model.

b) Thematic assessment:
   The thematic support was assessed on the number of auger bores that were available for each predicted soil series.
c) Feature space assessment:
   The feature space analysis was based on mapping the composite of all discrete node states of the covariates covered by the training data within each NATMAP unit.

The three assessments detailed above were combined by calculating the vector length in the 3D space provided by the three assessments.

25.3 Results and Discussion

25.3.1 Expert Knowledge

The results of the NATMAP analysis, based on local expert knowledge, are given in Table 25.1 for one NATMAP unit. Soil series in upper case notation are taken from the NATMAP legend, whilst the other soil series have been identified by the NATMAP analysis of all available data. This approach will have provided a much more robust approximation of the composition of NATMAP units than would have been possible from reviewing all available auger bore observations as in many cases the distribution of bores were biased by the sampling approach taken by individual surveyors. In some cases once soil surveyors have decided on the best soil association to represent a block of land, they would use their auger bores to help identify the minor constituents of that soil association. In this way constituent soil series that may occupy a relatively small proportion of a soil association can be represented by the most bores within LandIS. Using this approach, the majority of auger bores would represent a relatively small proportion of the soil association because larger sample numbers are used to identify the dominant constituents compared with the dominant series. For example (Table 25.2), the distribution of auger bores in relation to the predicted spatial extent of individual soil series based on the above approach

Table 25.1 Results of *NATMAP analysis* for Denbigh 1 association (541j)

Soil series	Map symbol	Soil Subgroup	Estimate of %age area within soil association
POWYS	Ph	3.13	36.0
SANNAN	Sn	5.42	20.0
Brickfield	Br	7.13	12.0
BARTON	BT	5.41	10.0
Cegin	Ca	7.13	8.0
MANOD	Mj	6.11	5.0
Hafren	HN	6.54	5.0
Wilcocks	Wo	7.21	2.0
Hiraethog	Hi	6.51	2.0

Table 25.2 Estimates versus actual auger bores

Series	Expected Distribution	Actual Distribution
Clifton	12.2	12.5
Brickfield	10.4	13.8
Winter Hill	8.5	0.0
Salwick	6.9	9.1
Waltham	5.5	2.3
Wilcocks	5.4	6.4
Nercwys	4.3	3.3
Newbiggin	4.3	0.4
Enborne	4.0	1.9
Wick	3.8	6.4
Quorndon	3.2	0.8
Crwbin	2.4	1.1
Malham	1.9	0.0
Arrow	1.9	1.2
Wharfe	1.6	0.9
Revidge	1.4	1.8
Manod	1.2	4.0
East Keswick	1.2	0.0
Belmont	1.1	0.7
Newport	1.1	2.7
Cegin	1.1	3.7
Claverley	1.0	0.0
Flint	1.0	0.2
Salop	1.0	0.4

has shown that Brickfield, Salwick and Wilcocks series are over sampled whereas Winter Hill, Waltham and Nercwys are under sampled.

The NATMAP analysis provided the tool for analysing the training database. It highlighted soil series that were missing from the auger bore collection (because of surveyor bias in the original sampling) as well as identifying those soil series that were missing from the national overview of soil associations but might normally have been expected to occur locally in a particular landscape (inclusions). In this way, this procedure identified soil series that needed to be sampled and provided a mechanism for prioritising the sampling as well as identifying those series that did not require further sampling. The expert NATMAP analysis also allowed any soil series with few observations to be merged with more widespread soils based on a 'similar soils' concept.

The NATMAP analysis provided an identification of:

- auger bores that were removed because of lack of certain key information;
- those soil series that would be present within a soil association but for which there were no auger bores within the training database;
- those records identifying rare soils that were re-classified according to the similar/dissimilar soils concept to more extensive soils within the catchment;
- areas within the 'thematic space' with no or very sparse data;
- areas where additional soil sampling was required.

In addition, the NATMAP analysis provided quantitative information that was required to refine both the soils series-NATMAP relationships as well as the *a priori* probabilities of the distribution of soil series in the Belief Network. The digital soil maps reflect the expected distribution of soils rather better than the distribution of soils based on the training data alone. In this way, the resultant digital soil maps should reflect the range of soils as well as their extent more accurately than would be the case if only the training database were used.

25.3.2 Quality Assessment

Spatial analysis
There were clear differences in the density and spatial distribution of auger bore observations across the study site (Fig. 25.3) with nested bores from sample farms clearly visible. The auger bore density associated with the fells is lower because the uplands had been already mapped prior to NATMAP using air photo interpretation.

Thematic analysis
The number of auger bores that supported the predicted soil series along the escarpments and on the fells were much lower (often $<$ 10 bores/soil series) than in the

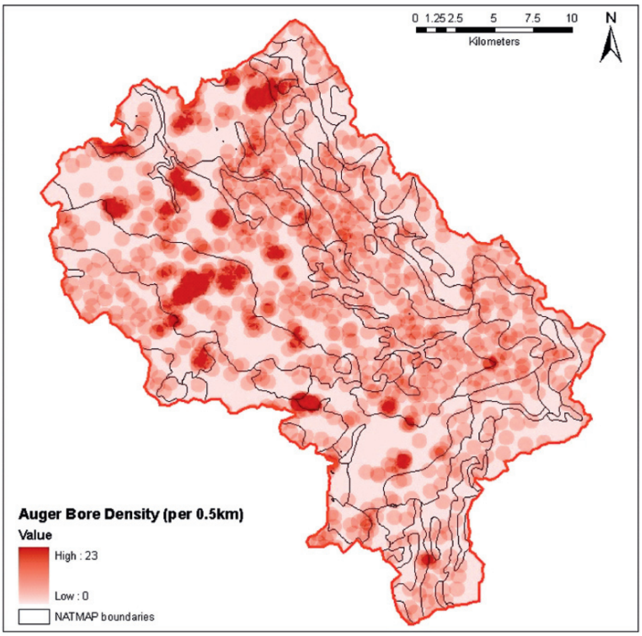


Fig. 25.3 Spatial assessment (See also Plate 32 in the Colour Plate Section)

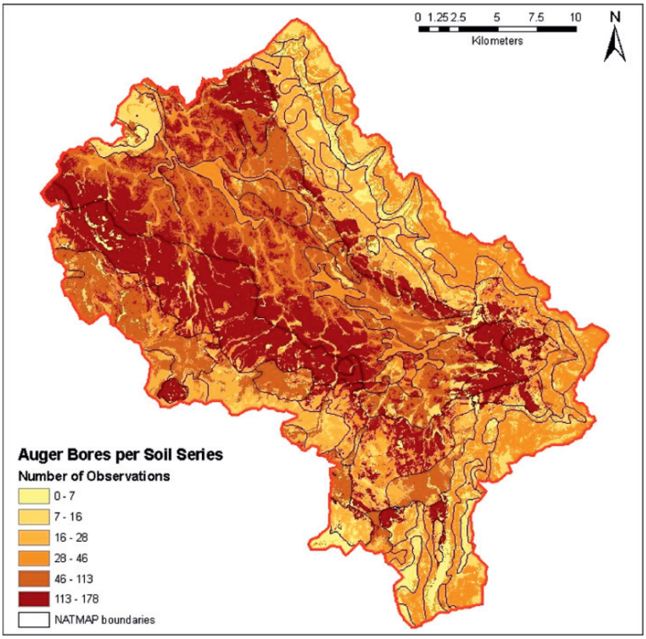


Fig. 25.4 Thematic assessment (See also Plate 33 in the Colour Plate Section)

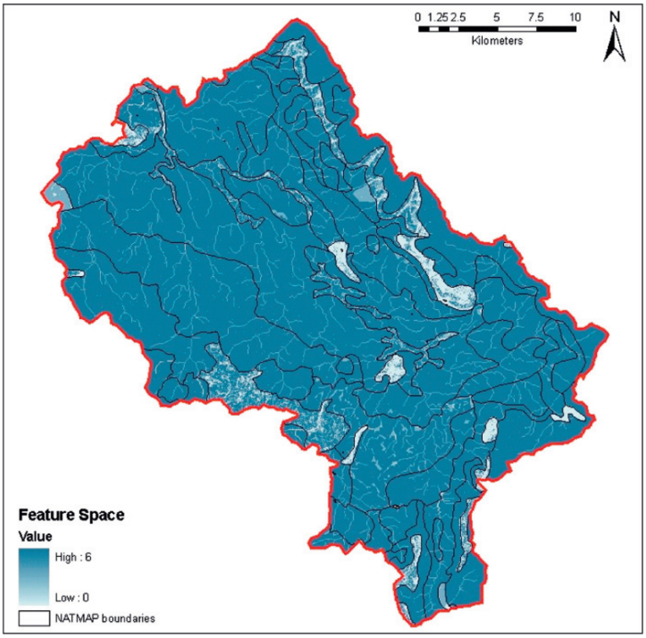


Fig. 25.5 Feature space assessment (See also Plate 34 in the Colour Plate Section)

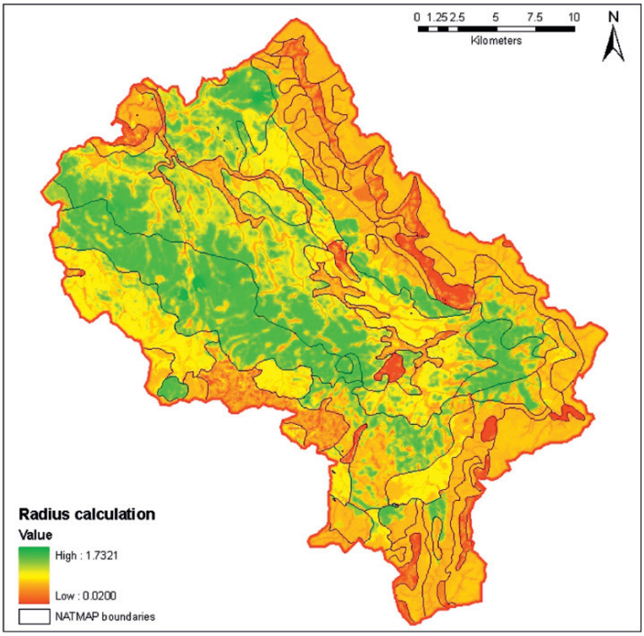


Fig. 25.6 Combined quality assessment (See also Plate 35 in the Colour Plate Section)

valley bottom (always over 50 bores/soil series and locally > 100) (Fig. 25.4). This reflects, in part, the priority given at the time to agricultural areas.

Feature space analysis
The results illustrated differences in the feature space covered by the models for each of the NATMAP units (Fig. 25.5). The analysis clearly identified those landscape units that were sparsely covered by the feature space. In this way the NATMAP analysis also provided a means to help prioritise additional soil sampling.

The combined quality assessment clearly identified those landscape units that were poorly covered by the model (Fig. 25.6). Consequently, the analysis identified areas which require additional sampling as well as the feature space in which the sampling has to be taken place.

25.4 Conclusions

This research has shown that digital soil mapping based on legacy soil survey data is fraught with problems. However, as long as there is a clear understanding of the mapping rules and conventions under which the soil data were collected, diverse soil datasets can be harmonised for the purposes of digital soil mapping. In this case many of the methods of harmonisation were undertaken within the NATMAP analysis and so enhancing the role of legacy data in producing reliable digital soil maps. Providing legacy data are analysed and used in conjunction with expert knowledge, they provide a valuable data set for digital soil mapping and hence are a potential rather than a curse.

Acknowledgments Funding is acknowledged for aspects of this work by the Department for Environment, Food and Rural Affairs (Project SR0120) and the Department for Environment, Food and Rural Affairs in conjunction with the Scottish Executive Environmental and Rural Affairs Department (Project SP0531). We would also like to thank Dr. Bob MacMillan for access to his software.

References

Avery, B.W., 1973. Soil classification in the Soil Survey of England and Wales. Journal of Soil Science 24: 324–338.

Avery, B.W., 1980. Soil Classification for England and Wales. Soil Survey Monograph No. 14, Harpenten, lawes Agricultural Trust.

Bui, E.N., Loughhead, A. and Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. Aust. J. Soil Res. 37: 495–508.

Clayden, B., Hollis, J.M., 1984. Criteria for differentiating soil series. Soil Survey Technical Monograph No.17, Harpenden.

Favrot, J.C., 1989. A stragety for large scale soil mapping: the reference areas method. Science du sol 27: 351–368.

Findlay, D.C., 1970. Making 1:25,000 Soil Maps. Agricultural Research Council, Soil Survey of England and Wales.

Hollis, J.M., Avery B.W., 1997. History of Soil Survey and Development of the Soil Series concept in the UK. Advances in GeoEcology 29: 109–144.

Mayr, T.R., Palmer, R.C., Lawley, R.S., Fletcher, P., 2001. New methods of soil mapping. Final report for Defra contract SR0120.

Mayr, T.R., Black, H., Towers, W., Palmer, R.C., Cooke, H.J., Freeman, M., Wood, C., Wright, S., Lilly, A., Jones, R., DeGroote, J., Hornung, M., 2006. Novel methods for spatial prediction of soil functions within landscapes. Final report for Defra and SEERAD contract No. SP0531.

Mayr, T.R., Palmer, R.C., 2006. Digital soil mapping: an England & Wales perspective. In: Lagacherie, P., McBratney, A.B., Voltz, M. (eds.) Digital Soil Mapping, An Introductory perspective. Developments in soil science, vol 31. Elsevier, Amsterdam 600pp.

McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma, 117(1–2): 3–52.

Soil Survey of England and Wales, 1983. Soils of England and Wales. Scale 1:250 000. Rothamsted Experiment Station, Harpenden.

Thompson, G.G., Beckman, C.H., 1959. Soils and Land Use in the Toowoomba Area, Darling Downs, Queensland. Soils and Land Use Series No. 28, CSIRO, Australia.

Chapter 26
Delineating Acidified Soils in the Jizera Mountains Region Using Fuzzy Classification

L. Boruvka, L. Pavlu, R. Vasat, V. Penizek and O. Drabek

Abstract Soil acidification represents a serious problem in mountainous areas of the Czech Republic. It is mainly caused by acid parent materials, high precipitation, the type of vegetation, and acid deposition. These factors act in different combinations and result in different soil conditions. The aim of this chapter is to distinguish areas in the Jizera Mountains with different levels of soil acidification and sensitivity using fuzzy classification.

A set of 98 sampling sites was analysed and sampling density was approximately one site per 2 km^2. Samples were collected from surface organic horizons (O), depth ranged from 4 to 22 cm depending on site conditions. Soil analysis included active and exchangeable soil pH, total content of C, N, and S, pseudototal content of Ca and Mg (after aqua regia digestion), and the ratio of absorbances of soil sodium pyrophosphate extract at the wavelengths of 400 and 600 nm as indicator of humus quality (A_{400}/A_{600}). Moreover, concentrations of exchangeable Al in KCl extract and organically bound Al in $Na_4P_2O_7$ extract were determined.

Soil classes were calculated using fuzzy *k*-means method with extragrades. Five classes were selected. The first class with high exchangeable Al content, high S and N, and low Ca, represents the area that was most affected by the acid deposition. The second class with the lowest pH represents strongly acid soils that have very high sensitivity to acidification, but with smaller acid deposition. The third class with high Ca content includes the areas that were limed in the past. The fourth class includes principally the sites with the highest S and N deposition that are populated by grass. The fifth class includes the areas with high Mg content; its distribution corresponds to beech forests that have more favourable effects on soils than spruce forests. Fuzzy classification distinguished soils with strongest sensitivity to acidification. Positive effect of beech forest, grass cover, and liming on surface organic soil horizons is shown.

L. Boruvka
Department of Soil Science and Geology, Czech University of Life Sciences in Prague, CZ-165 21, Prague 6 – Suchdol, Czech Republic
e-mail: Boruvka@af.czu.cz

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

26.1 Introduction

Soil acidification represents a serious problem in the mountainous areas of the Czech Republic. In some areas, acidification led to a complete forest decline. Soil acidification is a natural process induced by acid parent material, high precipitation, and forest vegetation cover. Moreover, it is accelerated by acid deposition (for example Kram et al., 1997; Uhlirova et al., 2002; Klimo et al., 2006). Although S deposition is currently decreasing due to desulfurization of thermal power stations, a different trend can be observed in the deposition of N as a result of more road traffic (Hrkal et al., 2006). Soils affected by acid deposition have diminished ability to neutralize continuing inputs of strong acids, provide poorer growing conditions for vegetation, and extend the time needed for ecosystems to recover from acid deposition (Driscoll et al., 2003). Acid deposition alters soils by depletion of Ca, Mg and other nutrient cations, mobilization of potentially toxic Al forms, and increasing the accumulation of S and N in soil. Soil recovery is a long term process (Galloway, 2001). Tao et al. (2002) based the assessment of soil sensitivity and vulnerability to acidification on cation exchange capacity and base saturation. The sensitivity can be evaluated also by critical loads, that is the highest depositional load that will not cause chemical changes in soil leading to long-term harmful effects on ecosystem structure and function. However, it is a complex issue with a number of uncertainties as reviewed by An et al. (2001) and Skeffington (2006). Ameliorating materials like limestone or dolomite were applied in the most endangered areas of the Czech Republic (Sramek et al., 2006) but the effect of liming is often disputable (Formanek and Vranova, 2002) because of consequent stronger organic matter decomposition, nitrogen leaching, tree root flattening etc.

The above mentioned factors can form a wide range of combinations, differing in soil sensitivity to acidification, buffering ability, and damage to the ecosystem. Spatial distribution of soil acidification can exhibit a strong variation in space. A number of methods have been invented to study and map soil spatial distribution (see Section 1.4, Chapter 2). In our previous papers, we studied spatial distribution of acidification indices in the Jizera Mountains region as an area strongly influenced by acidification and tried to assess the effect of stand factors on this spatial distribution (Boruvka et al., 2005a,b,c). The aim of this chapter is to distinguish parts that differ in the level of acidification status and sensitivity using unsupervised fuzzy classification with extragrades.

26.2 Material and Methods

A set of 98 sampling sites was analysed (Fig. 26.1). Sampling density was approximately one site per 2 km^2. Two objectives were followed in the sampling scheme design: 1) to cover the whole area more or less evenly, and 2) to include different categories of vegetation, altitude, slope aspect etc. Most sites were covered by spruce forest (*Picea abies*), a smaller part was covered by beech (*Fagus sylvatica*)

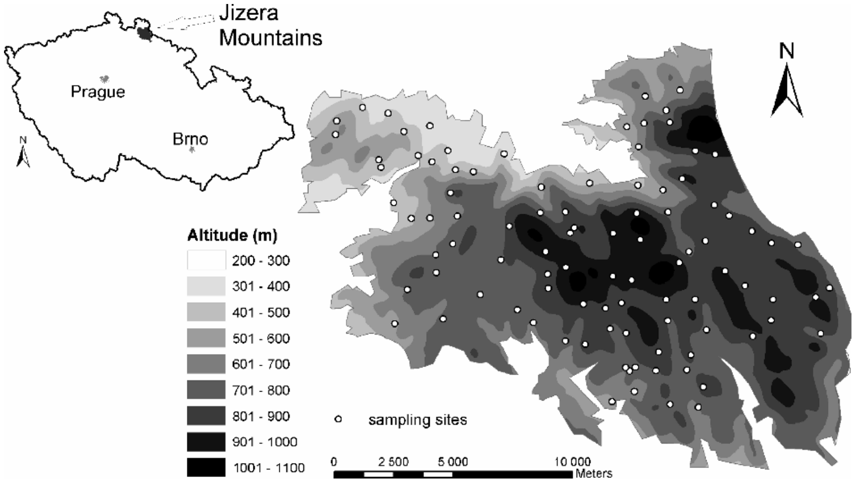


Fig. 26.1 Location of study site and sampling sites in the Jizera Mountans of the Czech Republic

forest. Clear-cut area in the most elevated parts of the region was populated by grass (*Calamagrostis villosa*). All soils (mainly Cambisols and Podzols) were formed on granitic bedrock, so that the effect of parent rocks can be omitted in this particular area. Detailed description of soil data can be found in Mladkova et al. (2004).

Soil samples were collected from surface organic horizons (O), the depth ranged from 4 to 22 cm depending on site conditions. These horizons are affected by anthropogenic influence the most (for example Boruvka et al., 2005d). One sample was collected from an area 50 $\times$ 50 cm at each sampling site. The analyses included active and exchangeable soil pH (pH-H_2O and pH-KCl, respectively), total content of C, N, and S, pseudototal content of Ca and Mg (after aqua regia digestion), and the ratio of absorbances of soil sodium pyrophosphate extract at the wavelengths of 400 and 600 nm as indicator of humus quality (A_{400}/A_{600}). Moreover, concentrations of two Al forms were determined, namely exchangeable (in 0.5 M KCl extract; Al_{KCl}) and organically bound (in 0.05 M $Na_4P_2O_7$ extract; $Al_{Na4P2O7}$). For details see Mladkova et al. (2004). In addition, C/N ratio was calculated, though it cannot serve as a good indicator of humus quality as it is influenced by the level of N deposition. S/Ca ratio indicating the ratio between S as the principal acidificant and Ca as a base element was also calculated.

Unsupervised classification was chosen; for advantages and disadvantages of this approach see Section 10.5.1. Fuzzy soil classes were calculated using fuzzy k-means method with extragrades (McBratney and de Gruijter, 1992; see also Section 20.2), using the program FuzME, version 2.1 (Minasny and McBratney, 2000). All measured and calculated soil characteristics listed above were used as the input data. Maps of the spatial distribution of membership values were consequently created using ordinary kriging. Maps of class membership values were created using ArcMap 8.1 (ESRI, Inc.) software.

26.3 Results and Discussion

Generally, the soils under study were extremely acid, with high S and N accumulation and high concentration of toxic Al forms (Table 26.1).

For the fuzzy classification, five classes were selected. It provided better results compared to 3 and 4 classes with respect to fuzziness performance index, modified partition entropy, and separation validity function (McBratney and de Gruijter, 1992). Class centres are shown in Table 26.2. Classes and their differences can be described as follows:

Class *a* shows high Al_{KCl} content, high S and N concentration, low Ca content, and high S/Ca ratio. According to this characteristic and the spatial distribution of membership values (Fig. 26.2), this class represents the area most affected by acid deposition.

Class *b* with the lowest pH values represents strongly acid soils with very high sensitivity to acidification, but with smaller acid deposition compared to the previous

Table 26.1 Basic statistical parameters of the studied dataset

Characteristic (units)	Mean	St.dev.	Min.	Max.
pH-H_2O	3.9	0.26	3.5	4.6
pH-KCl	3.2	0.24	2.8	3.8
Mg (mg kg^{-1})	846.6	424.3	214.0	2160.0
Ca (mg kg^{-1})	578.9	824.6	129.0	7770.0
C (%)	28.9	6.8	8.2	45.9
S (%)	0.33	0.11	0.08	0.68
N (%)	1.48	0.36	0.33	2.16
A_{400}/A_{600}	7.42	1.07	4.89	10.39
Al_{KCl} (mg kg^{-1})	1236.1	360.6	66.8	2890.9
$Al_{Na4P2O7}$ (mg kg^{-1})	5042.8	2002.1	1969.4	11403.6
C/N	19.78	1.74	15.27	24.53
S/Ca	12.13	8.69	0.83	51.68

Table 26.2 Fuzzy class centres

Characteristic (units)	Class *a*	Class *b*	Class *c*	Class *d*	Class *e*
pH-H_2O	3.91	3.70	3.95	4.01	3.98
pH-KCl	3.17	2.93	3.20	3.31	3.24
Mg (mg kg^{-1})	699.3	740.8	662.7	643.6	1447.0
Ca (mg kg^{-1})	312.4	480.3	570.8	587.3	402.9
C (%)	31.18	30.81	23.54	33.01	21.84
S (%)	0.37	0.33	0.27	0.40	0.26
N (%)	1.57	1.44	1.26	1.78	1.16
A_{400}/A_{600}	7.79	7.67	6.76	7.22	6.99
Al_{KCl} (mg kg^{-1})	2102.3	1322.2	971.3	1116.7	1062.7
$Al_{Na4P2O7}$ (mg kg^{-1})	3130.3	2610.7	3066.7	5518.0	2811.8
C/N	19.97	21.46	18.67	18.66	18.77
S/Ca	16.98	10.22	8.24	12.16	10.18

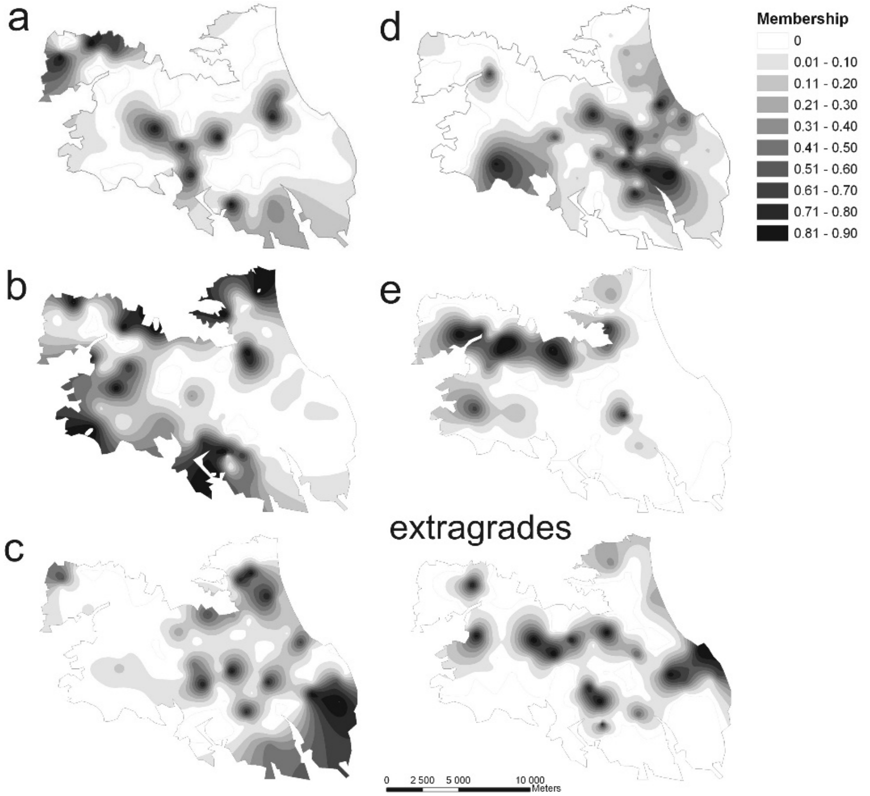


Fig. 26.2 Kriged maps of membership values for the five fuzzy classes (**a–e**) and extragrades

class, and it has lower S and N concentrations and lower C/N and S/Ca ratios. The concentration of toxic Al forms is also smaller than in class *a*. This class exhibits highest membership values in lower altitudes under spruce, which might indicate the positive effect of lower rainfall thanks to less intensive leaching of base cations from soils.

Class *c* with high Ca content includes the areas that were limed in the 1980s and 1990s. We could not take into account limed and unlimed areas in the sampling design, as the information is not available. Soils in this class have the lowest Al_{KCl} content and S/Ca ratio. The effect of liming is, however, obvious in the organic horizons, but the effect was not found in mineral soil horizons (Boruvka et al., 2005a). Higher membership values of class *c* are distributed mainly in the eastern part of the region, where liming was more intensive.

Class *d* includes the sites with the highest S and N deposition that are mainly vegetated by grass that has replaced the declined and clear-cut forest. Soils under grassland have better soil chemical conditions and organic matter able to bind Al; $Al_{Na4P2O7}$ concentration is therefore increased. As the cleared forest areas were limed, the concentration of Ca is high and the S/Ca ratio is only medium. pH values are the highest in this class. The highest membership values can be seen in the most elevated areas of the region.

Class *e* includes soils with high Mg concentration, low S and N accumulation and a low S/Ca ratio. Although the N content is lower, the C/N ratio is also relatively low compared to other classes, which may indicate a better quality of humus. This is supported also by slightly lower value of the A_{400}/A_{600} ratio. Concentrations of both Al forms are relatively low compared to other classes. Distribution of the highest membership values corresponds to beech forests that have less acidifying effects on soils than spruce forests due to better litter quality and lower acidificant interception from dry deposition.

The extragrades represent mainly scattered locations with various properties.

26.4 Conclusions

Fuzzy classification was used for forest soil acidification assessment. Five fuzzy classes were formed with different sensitivity and human impact. The less sensitive are classes *d* and *e*, where soils exhibit better buffering capacity. The main reason is the difference in vegetation cover, namely: grass (class *d*) and beech forest (class *e*). Positive effect of liming was shown in class *c* and *d*. Classes *a* and *b* can be assessed as the most sensitive, class *a* being more impacted by acid deposition than class *b*. Delineation of areas with different deposition load and sensitivity to acidification in the maps can be used for further decision making in forest management (Fig. 12.1).

Acknowledgments The data were obtained in frame of grant projects No. 1G57073 and QC 1250 of the Czech Ministry of Agriculture. Data processing was supported by the grant No. 526/06/1182 of the Czech Science Foundation.

References

An, J., Zhou, L., Huang, M., Li, H., Otoshi, T., Matsuda, K., 2001. A literature review of uncertainties in studies of critical lodas for acidic deposition. Water Air Soil Pollut. 130, 1205–1210.

Boruvka, L., Mladkova, L., Drabek, O., 2005a. Factors controlling spatial distribution of soil acidification and Al forms in forest soils. J. Inorg. Biochem. 99, 1796–1806.

Boruvka, L., Mladkova, L., Drabek, O., Vasat, R., 2005b. Factors of spatial distribution of forest floor properties in the Jizera Mountains. Plant Soil Environ. 51, 447–455.

Boruvka, L., Mladkova, L., Penizek, V., Drabek, O., 2005c. Forest soil acidification assessment using principal component analysis and geostatistics. Pedometrics 2005, Naples, Florida, September 12–14, 2005, p. 5–6 (abstract).

Boruvka, L., Podrazsky, V., Mladkova, L., Kunes, I., Drabek, O., 2005d. Some approaches to the research of forest soils affected by acidification in the Czech Republic. Soil Sci. Plant Nutr. 51, 745–749.

Driscoll, C.T., Driscoll, K.M., Mitchell, M.J., Raynal, D.J., 2003. Effects of acidic deposition on forest and aquatic ecosystems in New York State. Environ. Pollut. 123, 327–336.

Formanek, P., Vranova, V., 2002. A contribution to the effect of liming on forest soils: review of literature. J. For. Sci. 48, 182–190.
Galloway, J.N., 2001. Acidification of the World: natural and anthropogenic. Water Air Soil Pollut. 130, 17–24.
Hrkal, Z., Prchalova, H., Fottova, D., 2006. Trends in impact of acidification on groundwater bodies in the Czech Republic; an estimation of atmospheric deposition at the horizon 2015. J. Atmos. Chem. 53, 1–12.
Klimo, E., Materna, J., Lochman, V., Kulhavy, J., 2006. Forest soil acidification in the Czech Republic. J. For. Sci. 52(Special Issue), 14–22.
Kram, P., Hruska, J., Wenner, B.S., Driscoll, C.T., Johnson, C.E., 1997. The biogeochemistry of basic cations in two forest catchments with contrasting lithology in the Czech Republic. Biogeochemistry 37, 173–202.
Mladkova, L., Boruvka, L., Drabek, O., 2004. Distribution of labile aluminium forms in soils of Jizera Mountains region. Plant Soil Environ. 50, 346–351.
McBratney, A.B., de Gruijter, J.J.A., 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. J. Soil Sci. 43, 159-175.
Minasny, B., McBratney, A.B., 2000. FuzME version 2.1. Australian Centre for Precision Agriculture, The University of Sydney, NSW 2006 [online] (http://www.usyd.edu.au//su/agric/acpa).
Skeffington, R.A., 2006. Quantifying uncertainty in critical loads: (A) Literature review. Water Air Soil Pollut. 169, 3–24.
Sramek, V., Materna, J., Novotny, R., Fadrhonsova, V., 2006. Effect of forest liming in the Western Krušné hory Mts. J. For. Sci. 52 (Special Issue), 45–51.
Tao, F., Hayashi, Y., Lin, E., 2002. Soil vulnerability and sensitivity to acid deposition in China. Water Air Soil Pollut. 140, 247–260.
Uhlirova, H., Fadrhonsova, V., Biba, M., Lochman, V., 2002. Deposition and movement in forest ecosystems of selected substances with connection to the food chain. Chem. Listy 96, 598–606.

Chapter 27
Incorporating Legacy Soil pH Databases into Digital Soil Maps

S.J. Baxter and D.M. Crawford

Abstract Soil data and reliable soil maps are imperative for environmental management, conservation and policy. Data from historical point surveys, e.g. experiment site data and farmers fields can serve this purpose. However, legacy soil information is not necessarily collected for spatial analysis and mapping such that the data may not have immediately useful geo-references. Methods are required to utilise these historical soil databases so that we can produce quantitative maps of soil properties to assess spatial and temporal trends but also to assess where future sampling is required. This paper discusses two such databases: the Representative Soil Sampling Scheme which has monitored the agricultural soil in England and Wales from 1969 to 2003 (between 400 and 900 bulked soil samples were taken annually from different agricultural fields); and the former State Chemistry Laboratory, Victoria, Australia where between 1973 and 1994 approximately 80,000 soil samples were submitted for analysis by farmers. Previous statistical analyses have been performed using administrative regions (with sharp boundaries) for both databases, which are largely unrelated to natural features. For a more detailed spatial analysis that can be linked to climate and terrain attributes, gradual variation of these soil properties should be described. Geostatistical techniques such as ordinary kriging are suited to this. This paper describes the format of the databases and initial approaches as to how they can be used for digital soil mapping. For this paper we have selected soil pH to illustrate the analyses for both databases.

27.1 Introduction

Legacy soil data can provide a rich source of information about the state of soil over space and time (see Section 1.3 about legacy data as well as Chapters 23 and 25). For example, the fertility status of agricultural topsoil has been monitored for several

S.J. Baxter
Department of Soil Science, The University of Reading, Whiteknights, Reading, England, RG6 6DW, UK
e-mail: sam.baxter2@googlemail.com

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

decades in different countries and with the advance of computation, databases at State and national scale are being created (McBratney et al., 2003).

Two databases that have been digitally archived are the Representative Soil Sampling Scheme (RSSS) in England and Wales (Skinner and Todd, 1998), and the State Chemistry Laboratory database (SCL) in Victoria, Australia (MacLaren et al., 1996). The RSSS has annually recorded soil information from agricultural land, from 1969 to 2003, except for 1984. Tested fertility properties included soil pH, amongst others. The pH data have been analysed by classical statistics (Skinner and Todd, 1998; Skinner et al., 1992; Church and Skinner, 1986). In Victoria, Australia, the former SCL has analysed approximately 80,000 soil samples submitted by farmers between 1973 and 1994, for soil pH and other soil fertility tests. Initial statistical analyses have been reported by MacLaren et al. (1996). This paper focuses on data collected in 1991 for England and Wales and for data collected from 1973 to 1994 for Victoria. Reasons for this are explained in Section 27.2.

For both databases, analyses have been performed using administrative regions as boundaries (e.g. Skinner and Todd, 1998). However, such databases could provide a wealth of information at finer spatial scales using digital soil mapping and geostatistical techniques. The databases could be used to: explore spatial and temporal changes of the soil properties; create digital soil maps using optimal interpolation methods such as ordinary kriging; the uncertainties in the datasets could be quantified; and the results stored in a GIS. However, such databases are not always in a form that is suitable for spatial analysis and digital soil mapping. The purpose of this study is to unfold some of the issues in the use of legacy soil databases for digital soil mapping. The two databases mentioned above will be discussed and some preliminary results given, in particular in relation to their spatial coordinates and their suitability for geostatistical analysis.

27.2 Material and methods

The method of sampling for England and Wales is described in Church and Skinner (1986). Each year a representative selection of farms was visited and soil samples taken from four randomly selected fields. New farms were regularly introduced for sampling so that farmers' practices were not influenced as the data from the survey became known to them. Between 400 and 900 bulked soil samples were taken annually from different agricultural fields.

In Victoria, the analytical results, enterprise and location nearest the sample site (usually a farmer's field excluding atypical areas) were recorded. While the samples were taken randomly within the sampled area of the field, the fields were not selected at random. Some bias in the data may result from a preponderance of samples taken to assess fields suspected of having poor fertility although in latter years, a greater emphasis was placed on whole farm monitoring and nutrient management.

For England and Wales, prior to and including 1981, the farm only was identified by National Grid coordinates and not the four individual fields randomly selected at each farm for sampling. After 1981 each field was identified individually by National Grid coordinates. For spatial analysis this meant that each field formed a separate site for the analysis, whereas for the earlier data an average of the four values had to be used for each farm site. This has implications for comparing the spatial variation in the earlier years of the survey with that from the later years as the sample support differs. We selected data from 1991 (as these had field specific coordinates) and assessed the effect of averaging the data to approximately each farm site by averaging the data within a radius of 5 km. These were compared with using the full data (raw data) for each field. Temporal analysis of pH between 1971 and 2001 has been done elsewhere (Baxter et al., 2006); the concern here is to assess the effect of averaging data to a farm level.

The Victoria data has spatial coordinates derived from the nearest locality, that is, the nearest town, village, hamlet or sign post marking a location. There are approximately 10,000 of these in Victoria, of which 2,237 had soil pH samples associated with them. For a preliminary analysis, we used the mean value of the samples at each locality which represents all of the soil samples taken between 1973 and 1994 as no temporal trend was detected over the State. For this database the sample support differs between locality. However, because of the volume of data available, this dataset was used to ascertain whether a broad State wide assessment of the soil could be made.

The geostatistical analyses included variography and ordinary kriging (Webster and Oliver, 2001). Experimental variograms were computed and modelled using GenStat (Payne, 2000). Variograms of pH were computed for 1991 for England and Wales for the raw and averaged data, and for the data averaged at each locality in Victoria for 1973–1994. Interpolations were made every 2.5 km for England and Wales and every 0.025 decimal degree in Victoria on a square grid by ordinary punctual kriging. Thus three variables are examined.

27.3 Results and Discussion

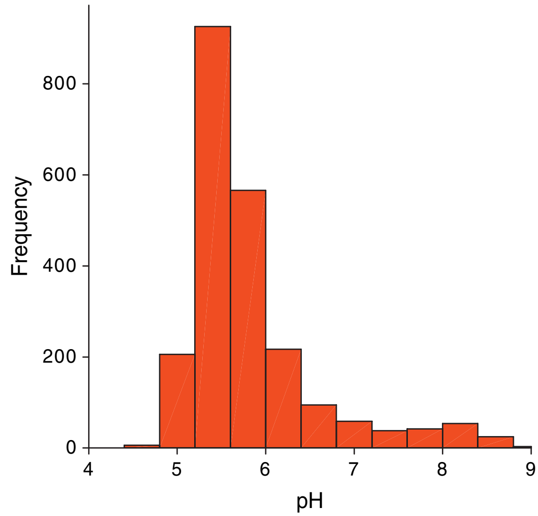
27.3.1 Summary statistics

Table 27.1. shows the summary statistics of pH computed from the raw and averaged data for England and Wales in 1991. It also shows the summary statistics for pH in Victoria. In England and Wales there were 716 fields and 181 farms originally

Table 27.1 Summary statistics for pH

pH	No. of values	Mean	Minimum	Maximum	Standard deviation	Skewness
Eng + Wales 1991	716	6.5	3.65	8.35	0.94	0.08
Eng + Wales 1991 averaged to 5 km	177	6.5	3.8	8.2	0.87	0.09
Victoria 1973–1994	2, 237	5.83	4.45	9.2	0.75	1.96

Fig. 27.1 Histogram of pH in Victoria, Australia



sampled in 1991. After averaging the data within a radius of 5 km there was 177 values. The number of averaged data is close to the number of farms. There is little difference in the mean values of the raw and averaged data (Table 27.1). The standard deviations are smaller for the averaged data, as would be expected, because the sample support has increased and the local sampling effects have been removed.

The average pH for Victoria was 5.83, the maximum value for Victoria was large, pH 9.2 (Table 27.1.). The median was pH 5.6 in Victoria. Figure 27.1 shows a histogram of the pH from Victoria. The distribution was positively skewed with most samples being acid but some samples being extremely alkaline. The latter come from areas in Victoria which naturally have moderately alkaline surface soils and an alkaline trend down the profile. In some fields, the latter have been exposed by erosion. The pH from England and Wales was more normally distributed with skewness values close to 0.

27.3.2 Variography

All three variables were spatially autocorrelated. The experimental variograms of pH for England and Wales showed evidence of spatial trend, concave shapes were observed at longer lags; this violates the assumption in geostatistics that the variable is random. The trend was removed by fitting a linear surface on the coordinates and the variograms were computed again on the residuals. Figure 27.2 shows the experimental variograms and their fitted models computed from the raw and averaged data of the linear residuals for pH (pH_{res}) for England and Wales sampled in 1991 and for pH from Victoria. Table 27.2 shows the parameters of the models fitted to the experimental values using GenStat (Payne, 2000) and Fig. 27.2. shows the experimental values and fitted models. For the England and Wales pH_{res} the sill variances,

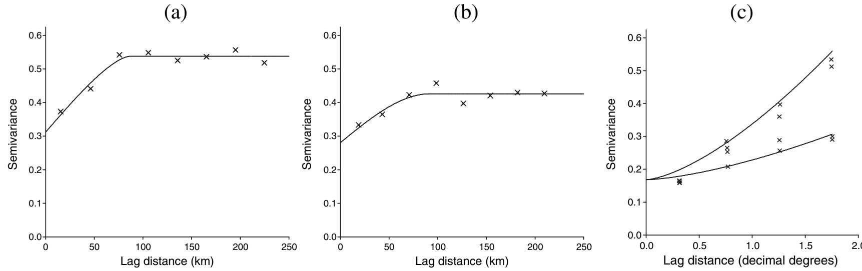


Fig. 27.2 Variograms of (**a**) the linear residuals of pH (pH_{res}) for 1991 for England and Wales; (**b**) the linear residuals of pH (pH_{res}) for 1991 for England and Wales with the data averaged within a 5 km radius; and (**c**) pH from Victoria for 1973–1994. The experimental values are indicated by the symbols, the model by the line

Table 27.2 Variogram model parameters for pH

pH	Model	c_0	c	a (km)	% nug
Eng + Wales 1991 linear residuals	Circular	0.3117	0.2261	86.10	58.0
Eng + Wales 1991 linear residuals averaged to 5 km	Spherical	0.2806	0.1454	92.0	65.9

pH	Model	c_0	phi	B	C	α
Victoria 1973–1994	Affine power	0.1683	0.3271	7.98	0.0363	1.486

Where c_0 is the nugget variance, c the sill of the autocorrelated variance, a the range of spatial dependence, α the exponent, phi the direction of maximum gradient and B and C are the gradients in the directions of maximum and minimum change, respectively.

c_0 and c, of the raw data are larger than the averaged data and the nugget:sill ratios ($c_0 : c_0 + \mathrm{c}$) are larger for the averaged data. This indicates that there has been some loss of detail in describing the variation of this property with the averaged data. The range of spatial dependence for the averaged data is longer (92 km) than that for the raw data (86 km). Averaging the data has smoothed the variation; nevertheless the variograms are similar therefore it is possible to discern the spatial structure with the smaller set of data.

Variograms computed from the raw and averaged England and Wales data for 1991 with 177 sites were compared with those computed from the data of the National Soil Inventory (NSI), which had about 5,670 sites taken on a 5 km grid (Oliver et al., 2002). The NSI variograms showed spatial dependence at two distinct scales, represented in the variogram as two spatial components: a short range and a long range component. The results suggest that the small number of sites in the RSSS, which are sparsely distributed over England and Wales, has provided a reasonable summary of the longer scale of variation present. The variogram model for pH_{res} from the NSI data had a long-range component of 85 km which is similar to the RSSS (Table 27.2).

The spatial variation of pH in Victoria showed directional differences. The variation was greater in the north-south direction compared to the east-west direction. Therefore the spatial variation was quantified using an affine power function variogram model (Table 27.2, Fig. 27.2).

27.3.3 Kriging

Figure 27.3 shows the maps of the kriged predictions for England and Wales. Sample locations are shown by crosses for the average data in Fig. 27.3(a), these are not shown on the map derived from the raw data, Fig. 27.3(c), as the crosses would be stacked up at this scale within farms. To obtain predictions on the original scale of measurement the linear trend was added back to the predictions of the residuals. The redder areas of the maps show where the values of the properties are large and the bluer areas where they are small. Figure 27.3 also shows the map of kriging variances for 1991 averaged values for England and Wales. The larger kriging variances show where the kriged predictions are less reliable; they decrease around the sampling sites. They are large around the edge of England and Wales and where the data was sparse. The map of kriged predictions using the raw data for England and Wales in 1991 shows more detail than the one using the averaged data as would be expected. However, the averaged one shows a reasonable level of detail at the national scale. The results of this assessment along with a comparison of the pH reported from the NSI indicate that it is

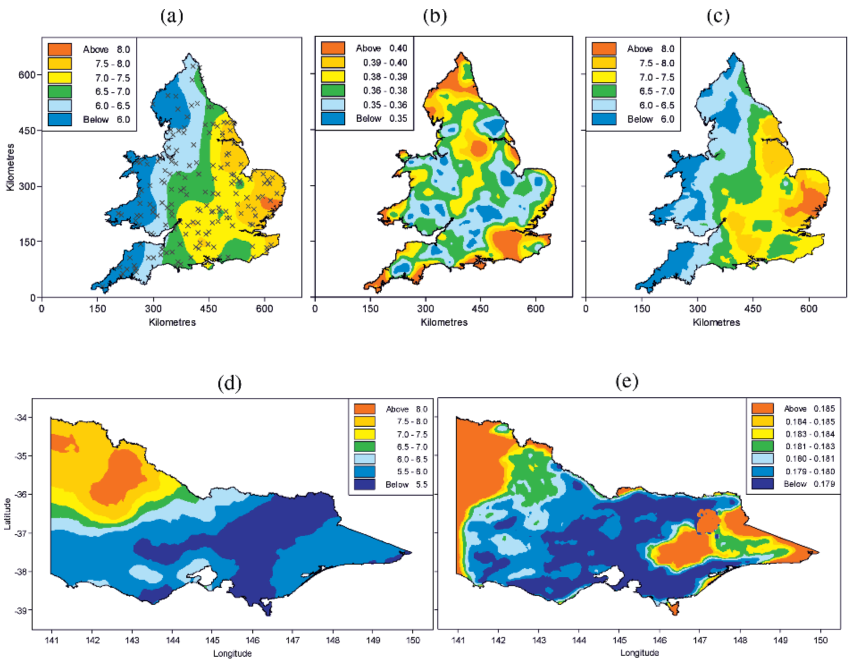


Fig. 27.3 England and Wales maps (**a**) kriged predictions of pH 1991 using the data averaged within a 5 km radius, sample locations are plotted as crosses; (**b**) kriged estimation variances of pH 1991 using the data averaged within a 5 km radius; and (**c**) kriged predictions of raw pH 1991. Maps of kriged predictions in Victoria, Australia of (**d**) pH; and (**e**) kriged estimation variances of pH (See also Plate 36 in the Colour Plate Section)

feasible to base further geostatistical analyses on the averaged data. This means that the data for earlier years can be analysed in the same way as those for latter years.

The values of pH are larger in the east and smaller in the west in England and Wales (Fig. 27.3). This can be explained by the land use, climate and geology. The agricultural land in the west is mainly grassland with higher annual rainfall than in the east. Under arable cropping in the east the pH is higher. Also the main chalk and limestone areas lie in a sequence running approximately east-west across southern England and turning northwards to run roughly north-south through eastern England (Skinner et al., 1992). For details of the temporal changes of pH over time at this resolution see Baxter et al. (2006).

The map of kriged predictions for Victoria has lower values in the east and south, and higher values in the north west (Fig. 27.3). These patterns are also linked to climate, with the lowest pH values (< 5.5) corresponding with the highest rainfall in the Great Dividing Range, Gippsland and the Otway ranges. Alkalinity (pH > 7.5) is observable in the drier north west. Figure 27.3 shows the estimation variances for pH in Victoria. These are large in the north west and east. This region has large tracts of public land and larger farms with bigger fields and therefore fewer agricultural soil samples will have been taken.

27.4 Conclusions

Legacy soil databases can provide a rich and valuable source of information for digital soil mapping. This paper has shown the potential for two soil databases to be used for digital soil mapping. Though not designed for spatial analysis, spatial autocorrelation was present and the variables can be kriged. The England and Wales farm averages and Victoria mean values appear to be adequate for describing large-scale trends.

Further work is needed to quantify the accuracy of soil pH in Victoria by comparing the predictions with exact spatially referenced measurements. There are numerous opportunities to explore the spatial and temporal variation of the soil properties in these databases. The soil properties could be mapped in relation to their soil texture or type of farming. The soil properties could be coregionalized with climate and terrain attributes. Future research could be done to explore how topsoil samples from agricultural soils could be integrated with soil profile information. Such an approach could enable a comprehensive assessment of the state of the soil in natural and managed landscapes at the State and national level.

Acknowledgments The RSSS work was funded by Defra and we thank Alan Todd of Rothamsted Research for providing the data. The RSSS soil samples were collected by ADAS staff. This paper was supported by the Soil Health Project at the Department of Primary Industries, Victoria – led by Richard MacEwan.

References

Baxter, S. J., Oliver, M. A., Archer, J. R., 2006. The Representative Soil Sampling Scheme of England and Wales: The spatial variation of topsoil pH and nutrient status between 1971 and 2001. Soil Use and Management 22, 383–392.

Church, B.M., Skinner, R.J., 1986. The pH and nutrient status of agricultural soils in England and Wales, 1969–83. Journal of Agricultural Science Cambridge 107, 21–28.

MacLaren, G., Crawford, D.M., Maheswaran, J., 1996. Temporal and spatial changes in soil acidity, fertility and salinity. National Landcare program final report. Agriculture Victoria, Melbourne, Australia.

McBratney, A.B., Mendonça-Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Oliver, M.A., Loveland, P.J., Frogbrook, Z.L., Webster, R., McGrath, S.P., 2002. Statistical and geostatistical analysis of the national soil inventory of England and Wales. The Technical Report, Defra. http://www2.defra.gov.uk/research/Project_Data/More.asp?I=SP0124&M=CFO& V=SSLRC [2006, May 08]

Payne, R.W., (ed.), 2000. The Guide to GenStat: Part 2, Statistics. VSN International, Oxford.

Skinner, R.J., Church, B.M., Kershaw, C.D., 1992. Recent trends in soil pH and nutrient status in England and Wales. Soil Use and Management 8, 16–20.

Skinner, R.J., Todd, A.D., 1998. Twenty-five years of monitoring pH and nutrient status of soils in England and Wales. Soil Use and Management 14, 162–169.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists. John Wiley & Sons, Chichester, UK.

Chapter 28
The Digital Terrain Model as a Tool for Improved Delineation of Alluvial Soils

V. Penizek and L. Boruvka

Abstract Typical examples of azonal soils are Fluvisols and Gleysols that occur around watercourses; they are bound to the alluvial part of landscape and have a characteristic spatial manifestation. They are good examples of a strong relationship between landform and soil. Here we wish to verify the efficacy of different relief characteristics derived from a digital terrain model (DTM) for the delineation of hydromorphic soils around small watercourses. The study is focused on choosing the most appropriate terrain attributes and their combinations. The study area consists of a small 83 km^2 catchment. A DTM with 10 m by 10 m pixels was derived from contours with a 2 m vertical interval. Three methods were compared: (1) combination of drainage area and slope curvature, (2) compound topographic index (CTI) and (3) combination of drainage area and height above the watercourse. The success of methods was verified by comparison of the width of estimated alluvial soils and alluvial soils extent delineated in detail soil map. Detailed comparison of the maps created showed discontinuities in predicted alluvial plain. The delineation based on compound topographic index provided was the worst. The alluvial plain was strongly underestimated (on average by 43%). Discontinuities of the alluvial plain were very frequent. Steep valley bottoms around smaller watercourses, that cause relatively low CTI values even near the watercourses, are the reason of this failure. The third method that was supported by the assumption that alluvial soil can be present only at some level above the watercourse with consideration of the size of the watercourse was the most successful. The extent of alluvial soils was underestimated by less than 22% and there were no discontinuities in the alluvial plain delineations. This study shows that terrain attributes can be a useful aid for delineation of soils strongly related to terrain.

V. Penizek
Department of Soil Science and Geology, Czech University of Life Sciences, CZ-16521 Prague 6 – Suchdol, Czech Republic
e-mail: Penizek@af.czu.cz

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

28.1 Introduction

Mapping soil is both expensive in time and money. Digital soil mapping can overcome this problem by exploiting easily-accessible auxiliary data, representing soil-forming factors. The choice of the auxiliary data is therefore driven by the prevailing soil-forming factors. The problematic of exploitation of auxiliary data is further discussed in Chapter 2. This study focused on delineation of alluvial soils represented by Fluvisols and some Gleysols. Alluvial soils occur almost worldwide. They are good examples of a strong relationship between soil and particular landscape units (Demek, 1988). Alluvial soils developed under one prevailing specific forming factor – relief, differ significantly from surrounding zonal soils. Different conditions of their development lead to their specific spatial distribution in the landscape. These soils are special not only because of soil-forming processes, but also by their environmental and agricultural properties.

Alluvial soils are usually chemically and physically fertile and are important for agricultural production. They are especially important in countries and regions with limited soil resources such as Africa (FAO, 1998). FAO (1993) describes these soils as a soil "with a reasonably moderate to high inherent fertility status that occupy flat, easily worked land. Therefore Fluvisols are very productive for a wide range of dryland crops and paddy rice on the various parts of the floodplain." For these reasons Fluvisols have been extensively reclaimed in some areas including Thailand, Surinam and Indonesia.

Delineating these soils on a map is an important source of information on their extent, and can be used for the assessment of their exploitation, protection of natural resources or agricultural use. Unfortunately, in many places in the world this information is not available, or is very limited. Traditional soil maps are available only at small scales in many countries in the world (see Section 33.1.1 for details). Small-scale soil maps usually overestimate the extent of alluvial soil due to map generalization. (For example the delineated extent of Fluvisols in the Czech Republic increases from 17% on the 1:50 000 map to 26% on the 1:250 000 map (Penizek et al., 2006). New techniques can be used for proper delineation of these soils.

Earlier studies showed that the description of these soils by modern methods is rather complicated and can significantly influence spatial prediction of other soils (Penizek and Boruvka, 2004). The spatial distribution of alluvial soils in the landscape can strongly influence the spatial prediction of other soils that are developed under a set of soil-forming factors (zonal soils) and description of their spatial variability. This is a possible reason to study these soils separately. Spatial delineation of these soils in landscape is the first step of their characterization. Methods integrating relief attributes are the most appropriate.

Relief units (elements and forms) can be classified by combination of different terrain characteristics. Alluvial plains are one of the basic relief units. They are characterized as relatively flat area around watercourses with a mainly concave transition to the surrounding landscape. There are typical relief properties of alluvial plains: high values of contributing area, low slopes, no or concave curvature, or high values of the compound topographic index (Park et al., 2001). Delineation of alluvial plains by digital elevation model (DEM) or digital terrain model (DTM) can be done using these characteristics. These characteristics can be differently combined and different criteria can be chosen (MacMillan et al., 2000; Park et al., 2001; Kozak et al., 2004). DTM/DEM data are relatively well accessible worldwide. Elevation data from remote sensing obtained from stereo aerial photographs or satellite images are the most important source of DEM especially in regions with limited geographic information (sources of DEM are discussed in Sections 10.4 and 2.2.1).

In this Chapter we aim to exploit different relief characteristics derived from DTMs for the delineation of the extent of hydromorphic soils around small watercourses. The study is focused on choosing the most appropriate relief characteristics and their combinations.

The Czech Republic has relatively good information about soils and their distribution. However, studying the relation between soil distribution and relief and finding techniques to derive soil distribution from terrain parameters can be used in the areas where soil information is rather scarce and limited (similar problematic is discussed in Chapter 4).

28.2 Material and Methods

28.2.1 Input Data and Area of Interest

The study area (83 km^2) located in Southern Bohemia ($49^\circ\ 26.62'$ N, $14^\circ\ 43.757'$ E) represents a typical catchment with a small watercourse (Fig. 28.1). A DTM with 10 m by 10 m pixels was derived from contour lines with a vertical interval of 2 m (in flat areas 1 m) and summit heights. Additionally, streams were used to support the interpolation of DTM. The size of pixel was estimated by contour density (Hengl, 2006). The DTM was processed by ILWIS Academic 3.2 (ITC, 2001).

28.2.2 Methods

Three methods were compared for alluvial-plain delineation: (1) combination of drainage area and slope curvature (Park et al., 2001), (2) compound topographic index (CTI) and 3) combination of drainage area and height above the watercourse.

Drainage area and slope curvature (AC)

This method combines characteristics of drainage area (As) and slope curvature (Cs) and variability of the slope curvature (Cs). Six landforms (including alluvial plain) can be described/classified by this method (Park et al., 2001). Alluvial plain is characterized as an area with low curvature, but with high drainage area. A negative slope curvature is a third characteristic.

Compound topographic index (CTI)

CTI indicates areas with higher soil moisture in low parts of landscape close to watercourses; on the other side, areas close to water divide are characterized by low values of CTI. CTI is calculated as:

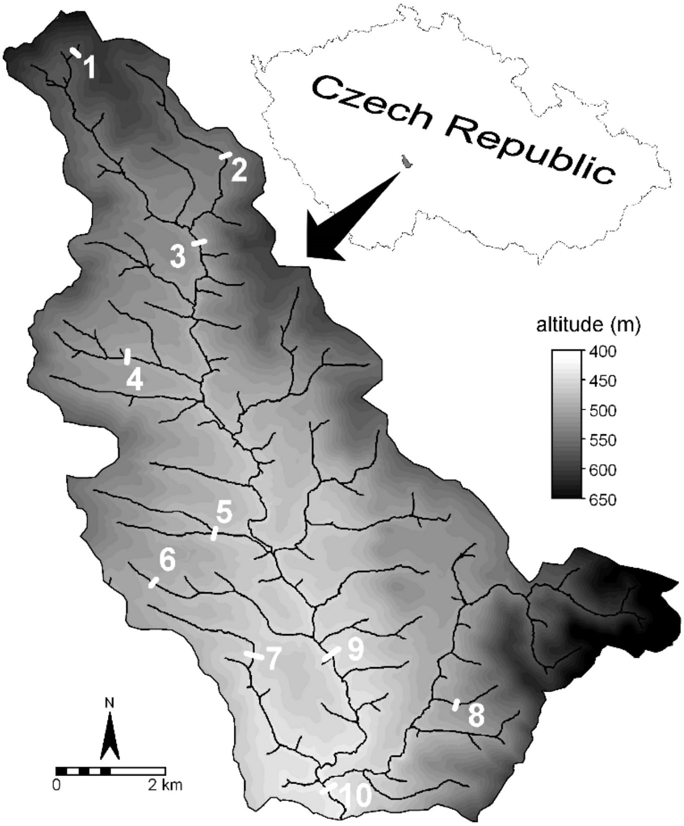


Fig. 28.1 Area of interest; numbers indicate control cross-sections

$$CTI = \ln\left(\frac{As}{\tan\beta}\right),$$

where As is drainage area and β is slope (McKenzie and Ryan, 1999). High values of CTI can indicate hydromorphic soils on the alluvial plain.

Drainage area and height above watercourse (AI)

This method is based on consideration that the presence of alluvial soils is limited by height above a watercourse in relation to the watercourse size. The watercourses form the alluvial soil by periodic flooding (a soil-forming factor for Fluvisols) and stagnation of groundwater close to the surface (a soil-forming factor for Gleysols). The presence of alluvial soil is then given by height of any point above the watercourse ($h_r - h_t$) where h_r is altitude at given point and h_t is altitude of the watercourse, and drainage area (As), which characterizes the size of the watercourse. The so-called "alluvial index" is described as:

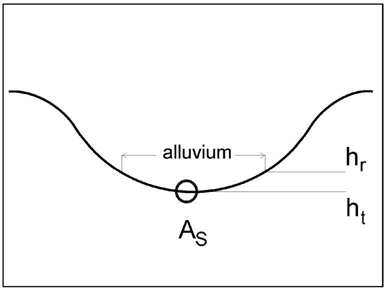


Fig. 28.2 Delineation of an alluvial plain by drainage area As and height above a watercourse ($h_r - h_t$)

$$AI = \frac{\log(As)}{(h_r - h_t)},$$

where As is drainage area, h_r is altitude at given point and h_t is altitude of the watercourse (Fig. 28.2).

28.2.3 Verification and Validation of Results

The success of the three methods was tested by comparison of the width of estimated alluvial soils with 10 cross sections of an existing 1:5 000 soil map. Cross-sections were chosen in regard to different shape of the valley (given by the steepness of the valley) and the size of the watercourse (given by the drainage area). Another comparison was based on visual assessment of the maps created.

28.3 Results and Discussion

All three methods described above were used for alluvial-plain delineation. The method described by Park et al. (2001) that combines drainage area and slope curvature was modified. Additionally a buffer of 150 m as a maximal width of valley at the study area was set to eliminate prediction of alluvial soil at places far from the alluvial plain that fulfill the limits given by Park et al. (2001). The control cross-sections showed that the alluvial plain width was underestimated by 24.5% on average (Table 28.1). Moreover, detailed comparison of the created map showed places where the alluvial plain appeared discontinuous. Another problem was constriction of the alluvial plain at the watercourse junctions (Fig. 28.3a).

Delineation of alluvial soils by the compound topographic index provided the worst result. Soils on the alluvial plain are hydromorphically influenced. Alluvial soils were delineated as areas with a CTI larger than 7, based on knowledge from previous studies (Penizek, 2004; 2005). The CTI map was processed as the average of a 3-by-3 pixel moving window. The alluvium was strongly underestimated (on average by 43%) by this method (Table 28.1). Discontinuities of the predicted

Table 28.1 Prediction of alluvial plain width at control cross-sections by methods. AC – combination of drainage area and slope curvature, CTI – compound topographic index and AI – combination of drainage area and height above watercourse

Width of alluvial plain at control cross-section (m)					Deviation between map and prediction					
					(%)			(m)		
cross-section	map	AC	CTI	AI	AC%	CTI%	IA%	AC-m	TI-m	AI-m
1	0	40	35	55	–	–	–	–40	–35	–55
2	70	70	35	55	0.0	50.0	21.4	0	35	15
3	70	80	50	55	14.3	28.6	21.4	–10	20	15
4	120	70	60	80	41.7	50.0	33.3	50	60	40
5	125	85	55	75	32.0	56.0	40.0	40	70	50
6	65	65	50	55	0.0	23.1	15.4	0	15	10
7	140	90	55	100	35.7	60.7	28.6	50	85	40
8	85	55	35	60	35.3	58.8	29.4	30	50	25
9	205	120	100	190	41.5	51.2	7.3	85	105	15
10	75	90	65	75	20.0	13.3	0.0	–15	10	0
		Average deviation			24.5	43.5	21.9	32	48.5	26.5

Comment: positive values indicate underestimation of predicted width of alluvium, negative values indicate overestimation of predicted width of alluvial plain in comparison to mapped state.

alluvial plain were very frequent (Fig. 28.3b). Steep valley bottoms around smaller watercourses, that cause relatively low CTI values even near the watercourses, are the reason of this failure.

The third method was based on the assumption that alluvial soil can be present only at some level above the watercourse in relation to the size of the watercourse. This method that combines height above watercourse and drainage area was the most successful. A value of $A = 1.5$ was set for delineation of the alluvial plain. The extent of alluvial soils was underestimated by less than 22% (Table 28.1) and there were no discontinuities in alluvial plain delineation (Fig. 28.3c). In one case the alluvial soil was predicted where these soils were not indicated on the original map. Even so this method provides the best result. Overall overview of success of the methods is given in Table 28.1 and Fig. 28.5; an example of prediction along two cross-sections is presented in Fig. 28.4.

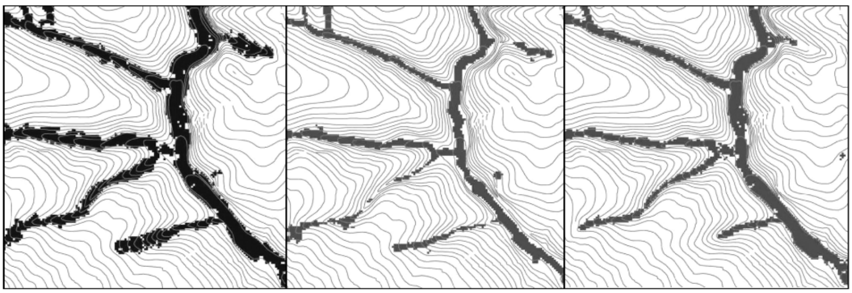


Fig. 28.3 Detail of delineation of alluvial soil by combination of drainage area and slope curvature (**a**), compound topographic index (**b**), combination of drainage area and height above watercourse (**c**)

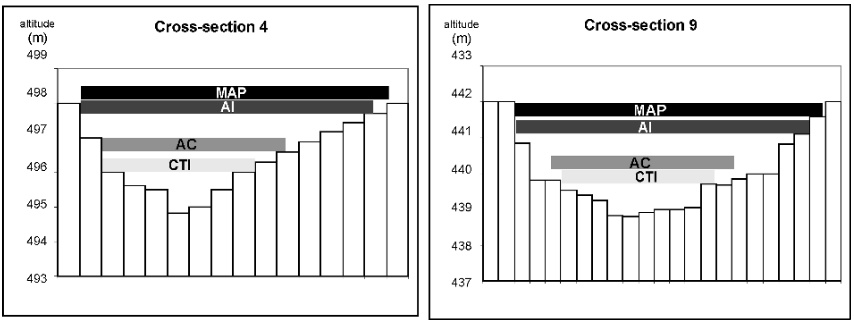


Fig. 28.4 Example of delineation of alluvial soils at two control cross-sections. MAP- extent of alluvial soils at reference map, AC – combination of drainage area and slope curvature, CTI – compound topographic index and AI – combination of drainage area and height above watercourse

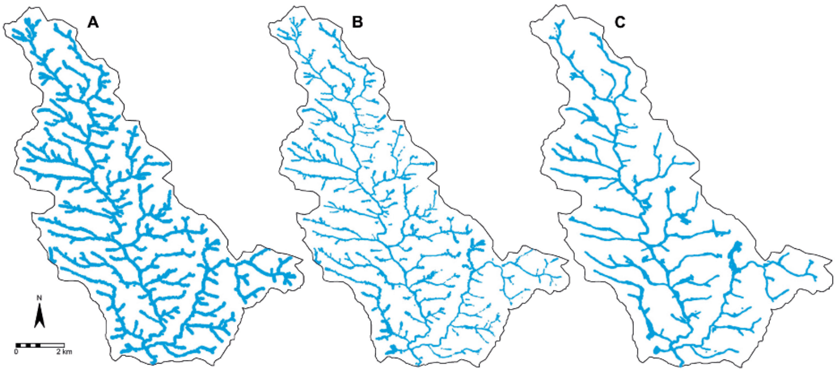


Fig. 28.5 Delineation of alluvial soils at study area by combination of drainage area and slope curvature (**A**), compound topographic index (**B**) and combination of drainage area and height above watercourse (**C**)

28.4 Conclusions

This study shows that terrain attributes can be a successful tool for delineation of soil classes strongly related to a specific type of terrain. The results of the three methods also showed that the success of prediction can be influenced by the choice of proper input parameters. Different characteristics of auxiliary data can differently serve to specific aims. One method/parameter can by successful for some study, but in other study it can fail. Exploitation of general-purpose methods or limits seems to have limited applicability. Training of the method in a small area and its application to a larger extent could provide the best prediction.

Nevertheless, digital soil mapping that exploits auxiliary data can be very useful and can overcome some problems of traditional soil survey and mapping (see also Chapter 11). This method can for example be used for the successful updating of existing medium- and small- scale maps (scales coarser than 1:50 000), where

Fluvisols and Gleysols around smaller watercourses are usually largely overestimated due to map generalisation. Exploitation of auxiliary data has large potential especially in the parts of world with limited information about natural resources.

Acknowledgments This study was supported by the Czech Science Foundation (Grant No. 526/06/1182) and the Czech Ministry of Education, Youth and Sports (Grant No. MSM 6046070901).

References

Demek, J., 1988. Obecna geomorfologie. Academia, Praha. 480p. (in Czech).

FAO, 1993. Soil Tillage in Africa: Needs and Challenges. Soils Bulletin 69. FAO, Rome.

FAO, 1998. Wetland characterization and classification for sustainable agricultural development. ISBN 0-7974-1928-4.

Hengl, T., 2006. Finding the right pixel size. Computers and Geosciences 32, 1283–1298.

ITC, 2001. Integrated Land and Water Information System 3.2 Academic. International Institute for Geo-information Science & Earth Observation (ITC), Enschede, the Netherlands.

Kozak J., Pridal M., Nemecek J., 2004. Upresneni rozlohy Fluvisolu na digitalnich mapach – mozna reseni. Pedologicke dny 2004, Roztoky u Krivoklatu. CZU, Praha. pp. 115–125. (in Czech).

MacMillan, R.A., Pettapiece, W.W., Nolan, S.C., Goddard, T.W., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Sets and Systems 113, 81–109.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.

Park, S.J., McSweeney, K., Lowery, B., 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103, 249–272.

Penizek, V., 2004. Vyuziti digitalniho modelu terenu pro mapovani nivnich pud pedometrickymi metodami. Final report of Grant GA AF No. 21130/1312/3192. (in Czech)

Penizek, V., Boruvka, L., 2004. Processing of conventional soil survey data using geostatistical methods. Plant, Soil and Environment 50, 352–357.

Penizek, V., 2005. Testovani ruznych pedometrickych metod pro zpracovani vysledku pudniho pruzkumu. Ph.D. thesis. Czech University of Agriculture, Prague. (in Czech)

Penizek, V., Boruvka, L., Kozak, J., 2006. DTM as a tool for correction of alluvial soil delineation on middle- and small-scale maps. 18th World Congress of Soil Science, July 9–15, 2006. Philadelphia, USA.

Chapter 29
Building a Digital Soil Data Base of the Solimões River Region in the Brazilian Central Amazon

W.G. Teixeira, W. Arruda, H.N. Lima, S.A. Iwata and G.C. Martins

Abstract The region near the Solimões river in the Brazilian Central Amazon receives much attention because of oil and gas transport from the Urucu river Province to the refinery in Manaus. Information about soil characteristics and its spatial distribution is important to allow secure intervention in the case of an accident (oil spill). The objectives of this chapter is to present the methodology used to built a soil digital data base of this region combining soil surveys that are mainly available as printed maps at different scales. First, the soil maps were scanned and vectorized and the soil units were identified. All information was put in a digital soil database, with scales from 1:250,000 to 1:10,000. The predominant soils near the borders of Solimões River are Eutric Fluvisols and Eutric Gleysols, whereas in the terra firme predominate yellow Ferralsols, Acrisols and Plinthosols occur. Some Podzols are found scattered in the area, normally at the base of short valleys. Anthrosols with rich antropic horizons also occur, which are called *Terra Preta de Índio*. A soil digital database using this approach to collect all information can be used to plan, monitor and reduce the impacts caused by the petroleum exploitation. It is also useful for land use planners in this region.

29.1 Introduction

The region near the Solimões river in the central Amazon receives much attention because of the exploration of oil and gas in the Urucu river Province. Petroleum is transported from Urucu until the town of Coari through a pipeline and then by ship to a refinery in Manaus. A net of researchers are studying different aspects in a cooperative project called – *Potenciais impactos e riscos ambientais do transporte de gás natural e petroleo na Amazônia* (PIATAM). Piatam is a major socio-environmental research program created to monitor the Urucu oil and natural gas production and

W.G. Teixeira
Embrapa Amazônia Ocidental – Rod AM 010 – Km 29 – Manaus – AM, 69011 – 970
e-mail: wenceslau@cpaa.embrapa.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

transportation activities. Urucu is the largest Brazilian in-land oil-rich province, located in the middle of the Brazilian Amazon. For the region, information about soil characteristics and its spatial location is a key factor to allow a secure intervention in the case of an accident like an oil spill.

Information about soil characteristics and soil maps for this region is rare and are scattered in technical reports (EMBRAPA, 1990; CETEC, 1986; BRASIL, 1978; IPEAM, 1970).

The objective of this study was to build a digital database with information about soil characteristics and their localization in the Solimões river area of the Central Amazon. This soil georeferenced digital database may be used to plan, monitor and reduce the impacts caused by the petroleum exploitation on the environment and the people living in the region. It also may help researchers in identifying areas where soil should be surveyed and more data are to be collected. In this chapter, we show the methodology used to convert printed soil maps in digital maps and how to combine the available soil surveys for a region in an unique digital database. Some pedological aspects of the soil in the region is also discussed.

29.2 Material and Methods

The study area is located in the border of the Solimões river comprising an area around 12 million hectares in the Central Amazon.

The first step was a search in libraries for published soil survey available for the central Amazon. It was identified four printed soil survey reports comprising this region and a digital soil database:

i. an exploratory soil survey – Folha Manaus SA-20 – Project Radambrasil – Soil map published at scale of 1:1000,000 – (BRASIL 1978);
ii. a reconnaissance soil survey of part of the city of Manacapuru – PDRI Project – Soil map published at scale of 1:100,000 – (CETEC, 1986);
iii. a reconnaissance soil survey of the region near the Road AM 070 (Cacau Pirêra to Manacapuru (Road AM 070) – Soil map published at 1:120,000 (IPEAM, 1970);
iv. semi-detailed soil survey for the Experimental Research station of Caldeirão – Embrapa Amazônia Ocidental – with a soil map published at scale of 1:10,000 (EMBRAPA, 1990).

A digital soil database is also available for this region: The Digital Soil Database of Legal Amazon – (DSDLA) – (SIPAM, 2004) has details compatible with the scale of 1:250.000.

The soil maps were scanned and vectorized and the soil units were identified. It was conducted in ArcGis 9.2 (ESRI, USA). Figure 29.1 illustrated the sequence used to transform printed soil maps in soil digital maps.

The SIPAM database was used as base to georeference the other soil surveys.

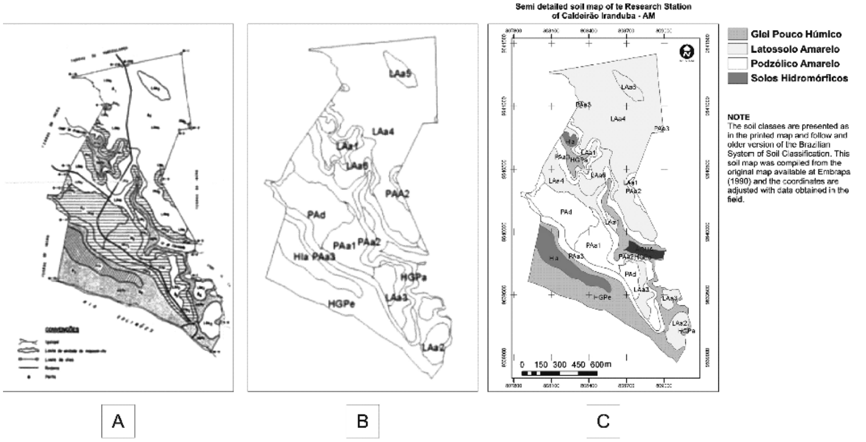


Fig. 29.1 Illustration of the sequence used to transform a printed soil maps in a soil digital map. (**A**) The printed map is scanned; (**B**) Transform the *boundary lines* between the soil units in lines vector to be identified in SIG soft wares (**C**) Organize the legend and the distinction between soil units using labels and representative color for different soil classes. This example is illustrated with the semi detailed soil map of the Experimental Research Station of Caldeirão – Embrapa Amazônia Ocidental (EMBRAPA, 1990)

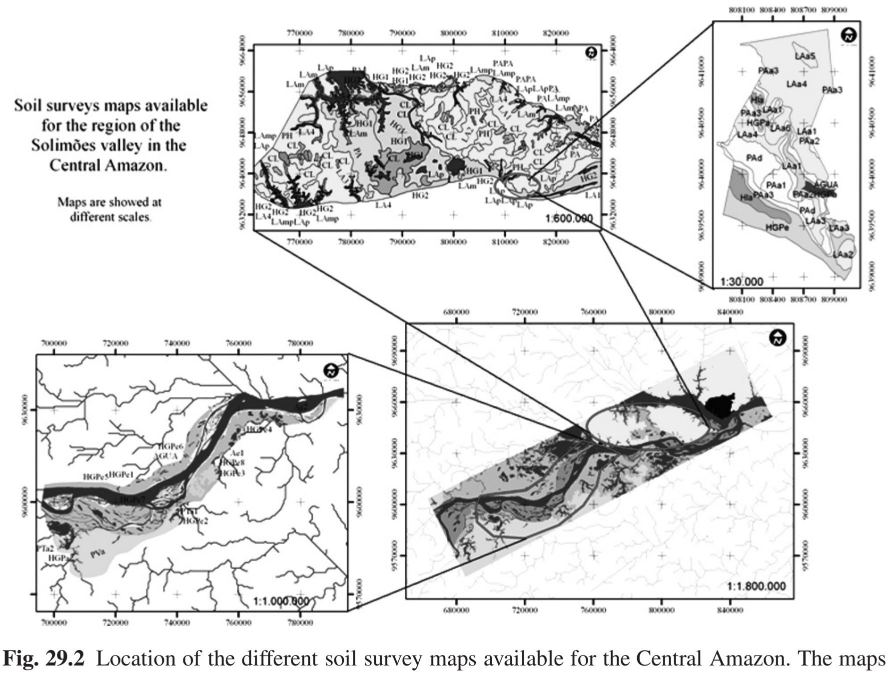


Fig. 29.2 Location of the different soil survey maps available for the Central Amazon. The maps have different site sampling intensity ranging from compatible with a semi-detailed legend (Research Station of Caldeirão – to 1:10,000) to an exploratory soil survey (SIPAM Digital Soil Data Base – 1:250,000) (See also Plate 37 in the Colour Plate Section)

The Fig. 29.2 shows the localization of all the soil surveys used in this study. For some soil maps it was necessary to rectify the boundaries of soil classes units. It was done using topographic features identified from digital elevation model generated from radar images obtained from the Shuttle Radar Topographic Mission (SRTM). We also used optical images from Landsat 5 TM which make it possible to identify vegetation types typically related with some classes of soil (e.g. *campinara* vegetation and Podzols, and typical *Igapó* vegetation for flooded areas).

The classifications of soil were kept as it was published. Finally, all the information was gathered in an unique digital soil database, with scales ranging from 1:250,000 to 1:10,000.

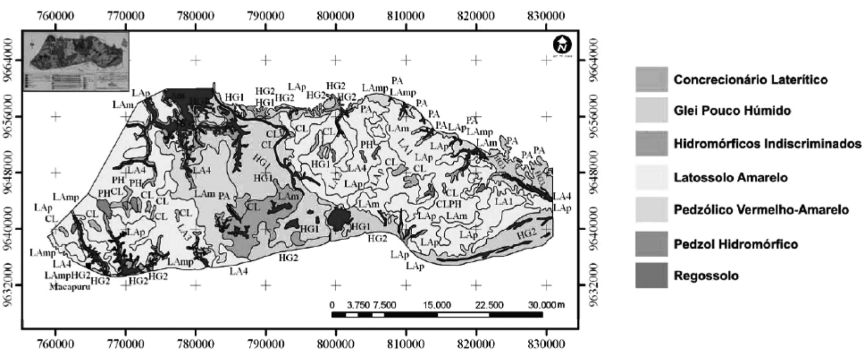
29.3 Results and Discussion

The Experimental Research Station of Caldeirão near the city of Manacapuru has a semi-detailed soil survey (Fig. 29.1). In this survey, soil units were identified in the scale of 1:10,000 and are composed typically by an unique soil class and each soil unit. It was observed a predominance (near the border of the Solimões) of Eutric Gleysols (*Gleissolos eutróficos – labeled as HGPe*). In this semi detailed soil survey was possible to separate some area with Dystric Gleysols (*Gleissolos distróficos labeled as HGPa*), those soils occur mostly in depressed areas and has a clayey soil texture. This soil map shows also a soil unit with an Acrisol with an antropic A horizon (*Podzólico Amarelo A antrópico* – labeled as Pad). This soil is locally called as *Terra Preta de Índio* and the A horizon is characterized by the dark color and often by presence of postherds, lithic artifacts and charcoal pieces. It has a high fertility and large amount of soil organic carbon. Moreover, it also shows different physical soil properties (Teixeira and Martins, 2003). The more widely accepted theory about the origin of those epipedons is that they were improved by Amerindian populations in pre-Colombian Indian settlements. Those Anthrosols have been found mainly in the "*terra firme*" (*i.e.*, areas never flooded) (Sombroek, 1966; Lima et al., 2002). Large areas of those Antrhosols where covered by sediments in the floodplains or where destroyed by the lateral movement of the river – a phenomenon locally named as *terras caídas*. An example of an area with an Anthrosol covered by sediments in the floodplains in this study area was described by Teixeira et al. (2006).

Figure 29.2 shows details of the maps that are gathered in an unique database. This digital database eases the access to information about the soil in this region. The use of zoom functions permits to visualize details where such information is available. Figure 29.2 also shows an unpublished exploratory soil maps compiled from the digital soil database from SIPAM. This map shows a predominance of Eutric Fluvisols (*Solos aluviais eutróficos*) and Eutric Gleysols (*Gleissolos eutróficos*). In the scale of this exploratory soil survey (1:250,000) it is very rare to be possible to map an unique soil class in the map units. It has a consequence that this kind of map is only for a rapid general appraisal of an area. For some purpose as intervention in an oil spill, exploratory soil surveys will be not useful and more detailed and up-to-date soil maps are necessary especially near the Purus river (affluent of the left border of Solimões river). In this area of the central Amazon an unique exploratory soil map is available. However, some areas in the central Amazon is already well described and mapped as the interval between the cities of Manacapuru and Iranduba, where three soil surveys are available.

Figure 29.3 show the original (small box) reconnaissance map of soils near the Roadway AM-070 – from Cacau-Pirêra to Manacapuru and also shows the renewed reconnaissance soil survey of the region between Cacau Pirêra and Manacaparu. Siltic Eutric Gleysols (*Gleissolos*) are dominant in a intricate pattern that makes the surveyor note as in some part as indiscriminate hydromophic soils. In the left border of Solimões river Dystric yellow Ferralsols (*Latossolo Amarelo*) dominate in association with Plinthosols (*Concrecionário Lateritico*, nowadays classified in the Soil Brazilian Classification System as *Plintosolo*) and Acrisols (*Podzólicos, nowadays* classified in the Soil Brazilian Classification System as *Argissolos*).

Figure 29.4 shows the renewed reconnaissance soil map of the borders of the Solimões river near the city of Manacapuru. At the border of the Solimões River Eutric Gleysols (*Gleissolos eutróficos*) and Eutric Fluvisols (*Solos Aluviais eutróficos*) are the dominant soil types. In the bluff areas at both sides of the Solimões river Acrisols (*Podzólicos*) and Plintosols (*Plintossolos*) are the major soil types. An increase in the proportionate extent of Acrisols in the terra firme occurs in direction to the nascent of the Solimões (Tables 29.1–29.3). The original publication of the reconnaissance survey (CETEC, 1986) also distinguishes map units based on relief and flood period. The same soil class occur in the same relief were separated when it has a different flood permanency. It was used a system of classification based on



NOTE
The soil classes are classified following and older version of the Brazilian System of Soil Classification.
This map was compiled from the digital soil data base publisehd by Ipeam (1:250.000).
The compilation and organization of the legend was done by Wenceslau Teixeira and Warley Arruda.

Fig. 29.3 Renewed reconnaissance soil map from Cacau-Pirêra to Manacapuru (Roadway AM 070) – Published by IPEAM (1970). The original map of the reconnaissance soil map from Cacau-Pirêra to Manacapuru (Roadway AM 070) is showed in the *small box left* side (See also Plate 38 in the Colour Plate Section)

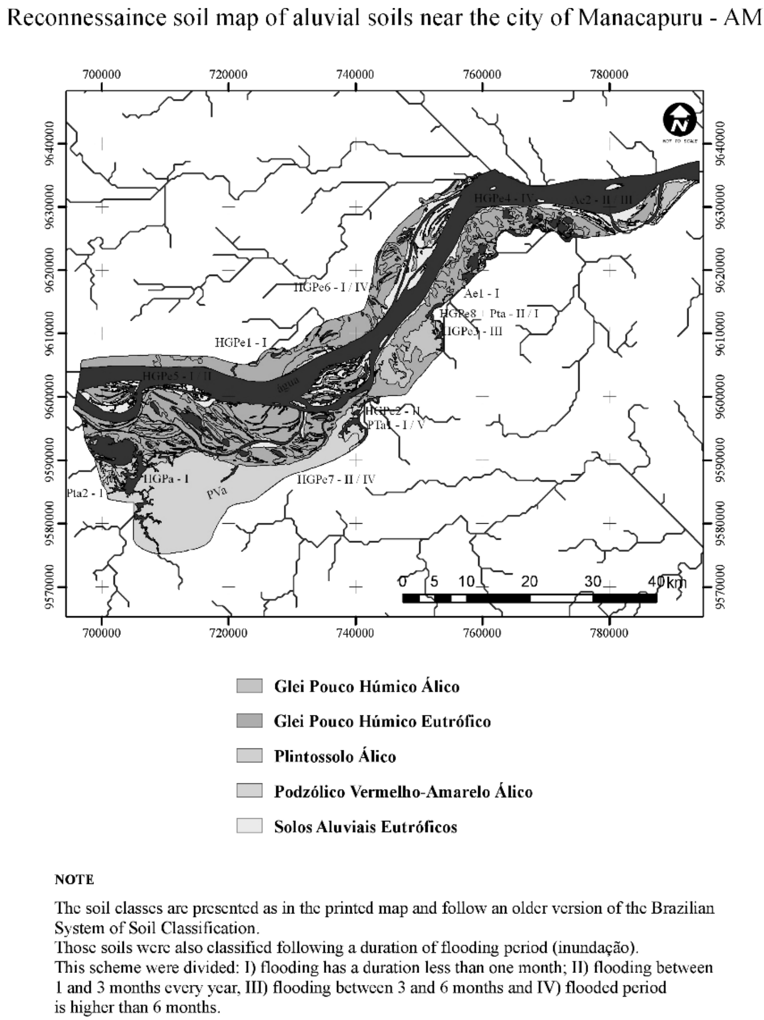


Fig. 29.4 Renewed reconnaissance soil map of the soil near the *border* of Solimões in the city of Manacapuru (CETEC, 1986) (See also Plate 39 in the Colour Plate Section)

flood period: less than one month (class I); among one and three months (class II); among three and six months (class III) and more than six months (class IV). In this area, the flood criterion is very important as the flooded area may occur in a large portion of the terrain.

Tables 29.1, 29.2 and 29.3 show the symbols used in the maps, the number of units mapped and the measured area relative to the total area. This type of table and statistics are laborious in traditional mapping but easy in a digital map using geographic information system (GIS). Soil map units in digital format allow comparison with other thematic databases like, for example, cross studies of land use systems and soil class (Soares et al., 2007).

Table 29.1 Map units symbols, dominant soil type, number of units mapped, measured area and percent of each soil class mapped from the renewed semi detailed soil survey of the Research Station of Caldeirão – Embrapa Amazônia Ocidental (EMBRAPA, 1990)

Map unit symbols	Dominant soil type[§]	Number of units mapped	Area (ha)	%
LAa4	Ferralsol	2	62.8	36
HGPe	Eutric Gleysol	1	24.3	14
Pad	Acrisol	2	15.9	9
PAa3	Acrisol	4	13.3	8
HGPa	Dystric Acrisol	3	11.5	7
PAa2	Acrisols	2	11.0	6
Hla	Gleysols and Fluvisols	2	9.0	5
PAa1	Acrisolo	1	8.2	5
LAa6	Ferralsol	1	5.3	3
LAa3	Ferralsol	2	5.2	3
LAa2	Ferralsol	1	3.7	2
LAa1	Ferralsol	4	3.2	2
LAa5	Ferralsol	1	2.6	1

[§] FAO – ISRIC (2003)

This soil database is freely available in a digital format (–shape file) which makes it possible to visualize in ArcGis (ESRI, USA) or free software as Springer (INPE, Brazil).

Table 29.2 Map units symbols, dominant soil type, number of units mapped, measured area and percent of each soil class mapped from the renewed reconnaissance soil survey of area between Cacau-Pirêra and Manacapuru – Brazilian Central Amazon (IPEAM, 1970)

Map unit symbols	Dominant soil type[§]	Number of units mapped	Area (ha)	%
LA1	Ferralsol	1	18699	15
LA3	Ferralsol	1	14698	12
HG1	Gleysol	8	13192	11
Lap	Ferralsol	16	12057	10
HI	Gleysol and Fluvisol	64	12022	10
HG2	Gleysol	13	11547	9
PA	Acrisol	12	11096	9
Lam	Ferralsol	10	7981	7
LA2	Ferralsol	1	7087	6
Lamp	Ferralsol	7	6306	5
HG3	Gleysol	4	2592	2
LA4	Ferralsol	6	2554	2
CL	Plinthosol	16	2397	2
PH	Podzol	8	959	1
R	Regosol	2	87	< 0.5

[§] FAO – ISRIC (2003)

Table 29.3 Map units symbols, dominant soil type, number of units mapped, measured area and percent of each soil class mapped from the renewed reconnaissance soil survey of county of Manacapuru – Brazilian Central Amazon (CETEC, 1986)

Map unit symbols	Dominant soil type[§]	Number of units mapped	Area (ha)	%
PVa	Acrisol	6	27674	24
Ae2	Fluvisol	101	15105	13
HGPe4	Gleysol	95	14357	12
HGPe1	Gleysol	9	12895	11
HGPe7	Gleysol	23	12555	11
HGPe8	Gleysol	6	7522	6
HGPe6	Gleysol	21	7512	6
HGPe5	Gleysol	12	6649	6
HGPe3	Gleysol	5	5829	5
Ae1	Fluvisol	6	2314	2
HGPe2	Gleysol	11	2005	2
HGPa	Gleysol	15	1259	1
PTa1	Plinthosol	9	1096	1
PTa2	Plinthosol	2	556	< 0.5

[§] FAO – ISRIC (2003)

An ongoing project is the revision of the legend to follow the Brazilian System of Soil Classification from 1999. This project also included a databank called SOLOAMA, containing information about the soil physical, chemical, morphological and mineralogical properties, partly available in soil survey reports, scientific articles and doctoral and master theses and soil data collected during PIATAM excursions in the Central Amazon.

A secondary objective of this project is to reduce the risk of loss of information from former soil surveys and the conducting of new soil survey for the region. We incorporated legacy soil data into new georeferenced databases and this approach to rescue and renewal soil data is discussed in this book by Rossiter (see Chapter 6).

29.4 Conclusion

The approach to collect all available soil information in a database may be used to plan, monitor and reduce the impacts caused by the petroleum exploitation on the environment and the people living in the region Central Amazon. It may also be useful for land use planners.

A limitation of a soil data base combining soil surveys of different scales is the printing as the scale is different for different areas. This database is typically used in a computer with a software that allows zooming the map. It may cause confusion for non-soil specialist showing that some areas are more variable than other, which is a consequence of a larger scale used in semi-detailed soil surveys which identified more soil classes than in exploratory soil surveys.

References

BRASIL. 1978. Ministério das Minas e Energia. Levantamento Exploratório de Solos. Folha Manaus SA-20 – Projeto Radambrasil. Rio de Janeiro, 628 p. + maps.

CETEC. 1986. Fundação Centro Tecnológico de Minas Gerais. Levantamento de reconhecimento de solos e aptidão agrícola em áreas abrangidas pelo PDRI – AM. Município de Manacapuru CETEC. Belo Horizonte, 185 p. + maps.

EMBRAPA. 1999. Sistema brasileiro de classificação de solos. Rio de Janeiro, Centro Nacional de Pesquisa de Solo, 412 p.

EMBRAPA. 1990. Levantamento semi-detalhado dos Solos da Estação Experimental do Caldeirão (Iranduba). Report Embrapa CPATU – Belém, 56 p. + maps.

FAO – EC – ISRIC. 2003. World soil resource map. FAO, Rom. Map at 1:25.000.000 scale.

IPEAM. 1970. Os solos da área Cacau-Pirêra – Manacapuru. IPEAM. Belém. 198 p. +1 map. (Série: Solos da Amazônia, v. 2–3).

Lima, H. N., Schaefer, C. E. G. R., Mello, J. W. V., Gilkes, R. J., Ker, J. C. 2002. Pedogenesis and pre-Colombian land use of "Terra Preta Anthrosols" ("Indian Black Earth") of Western Amazonia. Geoderma, 110(1–2): 1–17.

SIPAM. 2004. Base Pedológica da Amazônia Legal-Convênio SIVAM – IBGE.SIPAM, Brasília. Digital data base at 1:25.000.000 scale.

Soares, C.B.S., Teixeira, W. G., Pinto, W. H. A., Costa, L.A. 2007. Geotecnologia aplicada ao estudo da dinâmica do uso e cobertura da terra na Área de Proteção Ambiental (APA) Encontro das Águas – Iranduba (AM). In: Anais do XXII Simpósio Brasileiro de Sensoriamento Remoto. Florianópolis. INPE, pp. 3163–3170.

Sombroek, W. 1966. Amazon soil: a reconnaissance of the soils of the Brazilian Amazon region. Wageningen: Centre for Agricultural Publications and Documentation.

Teixeira, W. G., Martins, G. C. 2003. Soil physical characterization. In: Lehmann, J., Kern, D.C., Glaser, B., Woods, W. I. (eds.), Amazonian Dark Earths: Origin, Properties and Management. Kluwer, Dordrecht, pp. 271–286.

Teixeira, W. G., Martins, G. C., Lima, H. N. 2006. An Amazonian dark earth profile description from a site located in the floodplain (Várzea) in the Brazilian Amazon. In: Morcote, G., Mora S., Franky C. (eds.), Pueblos y paisajes antiguos de la Selva Amazónica. Universidad Nacional de Colombia-Taraxacum, Bogotá, pp. 293–300.

Chapter 30
Enhancing the Use of Remotely-Sensed Data and Information for Digital Soilscape Mapping

L. Le Du-Blayo, P. Gouéry, T. Corpetti, K. Michel, B. Lemercier and C. Walter

Abstract The lack of soil maps in Brittany in the north west of France, leads to an approach based on the inference of soilscape units which can be delimited and characterised with relatively fewer field observations than conventional survey. Whereas geology and landform are generally used data to map soilscape units, natural and agricultural landscapes indicate relevant information on soils within them. Remote sensing is obviously the main source of data to map landscape units at regional scale, but one must look carefully how to analyse landscape units, including soil properties, without simply focusing on land-use class. The proposed method for landscape classification is based on a specific classification system developed at regional and local scales, including the role of landscape patterns using object-oriented classification. Post-classification processing is then developed to generalise the results and define mixed landscapes. Finally fusion techniques are tested to examine the probability of common soilscape boundaries arising from different environmental factors (geology, elevation, landscape).

30.1 Introduction

A central purpose of the French national program *Inventory Management and Soil Conservation* is to provide digital soil maps at regional extents. For Brittany, soil mapping is not extensive. The existing polygon maps, which are not all digitised yet, cover approximately 20% of the area at 1:100 000, and less than 15% at 1:25 000. So this highly developed agricultural region has quite poor soil reference maps and Bottom Up synthesis as describe in Section 13.3.1 are not possible for the region. Faced with this lack of spatial data on soils, and the cost-to-time ratio necessary to collect sufficient new soil observations, one solution is to use external factors

L. Le Du-Blayo
Equipe COSTEL, UMR CNRS 6554 (Littoral Environnement Télédétection et Géomatique), Université de Rennes 2, place du recteur H. le Moal, 35 043 Rennes cedex France
e-mail: laurence.ledu@uhb.fr

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

(environmental covariates as noted in Section 1.2) to deduce soil spatial patterns, which can then be verified locally. The program maps soilscape units (*Unités de Pédo-Paysages*: UPP) at 1:250 000, and populates each UPP with a definition of local soil classes (*Unités Typologiques de Sols*: UTS). This concept of a soilscape unit (Lagacherie et al., 2001) results from the general principle, established in various disciplines, that functional links and pressures between environmental and human factors produce a specific landscape at the regional scale (Le Du-Blayo 2007).

In Brittany, where about 80% of the land area is agricultural, the main agricultural landscape patterns (size and shape of fields, cropping and pasture land, density of hedgerow network, waste land etc.) are closely related to soil attributes. The landscape and landuse mosaic often reveals soils features that other factors like geology or elevation do not (Gaddas, 2001). For example we can find three small areas of vegetable crops precisely located on originally basic soils on aeolian surface deposits, with no link to other environmental factors. The landscape/landuse pattern is also an important factor for DSM (Digital Soil Mapping) because human activities drive soil evolution (urbanisation, planted conifer woodland, intensive agriculture etc.). Thus, according to the *scorpan* model (McBratney et al., 2003, see Section 2.2.1), in a region like Brittany, *o* (organisms, vegetation or fauna or human activity) is a key determining factor for deriving landscape/landuse units for successful soilscape mapping (as described in Section 3.3.1).

Remote sensing is a readily available source of information on landscape/landuse, accessible nearly all around the world and often *gratis*. It is well adapted to work where there is a dearth of soil maps and data. The question is then to set up an appropriate method for landscape/landuse unit mapping (Le Du, 2000) that can be subsequently used for inferring soil units.

30.2 Material and Methods

The methodology comprises three main steps (Fig. 30.1):

(1) landscape/landuse classification using remotely- sensed data;
(2) post-classification processing in order to obtain compact landscape/landuse units; and,
(3) fusion with other environmental data to extract soilscape units at the regional extent.

30.2.1 Classification of Landscapes

In the context of this chapter ‘landscape’ refers to a combination of landform and landuse, more formally, a spatial combination of specific elements (house, hedge, river, wood. . .) with a specific spatial structure (regular or irregular, large or small parcels. . .). Thus, a *landscape class* is defined at local scale by a composition

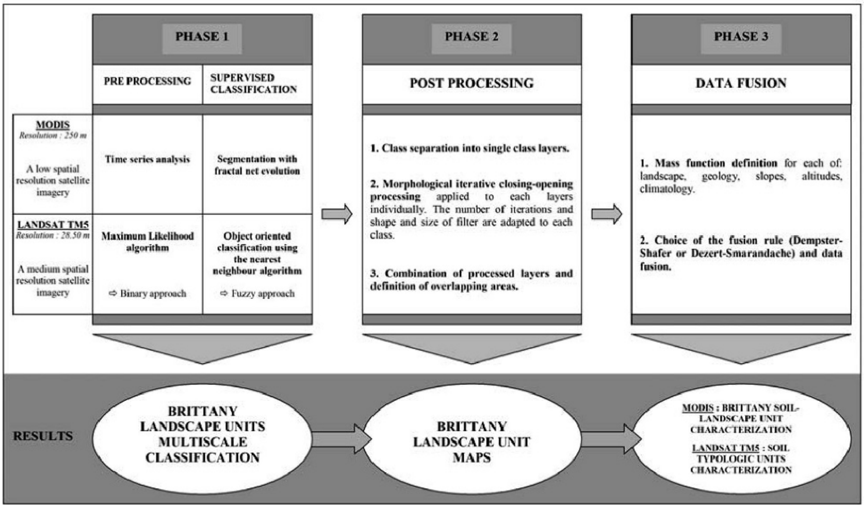


Fig. 30.1 A diagram of the approach

(land use classes) and a spatial topology. At the regional scale, landscape units are defined by a composition of *landscape class* (open field, bocage, suburban area...) with a specific spatial organisation (homogeneous/heterogeneous, linear/mosaic etc.).

Remote sensing is often a unique source of data concerning land use. Here we have focussed on two specific scales.

(1) A whole-of -Brittany scale for delimiting the UPP (1:250 000). To work at this scale, we used MODIS (MODerate resolution Imaging Spectroradiometer) imagery with a low spatial resolution (250 m by 250 m for bands 1, 2, 3 and 500 m by 500 m for bands 4, 5, 6, 7). This imagery is well adapted for regional-scale work because this it shows the main landscape contrast and covers Brittany in a single scene. This is very important for homogeneous data and simplifies the image preparation (Lecerf et al., 2006). Spring, summer, autumn and winter images are combined using Time Series Analysis (TSA) to integrate the seasonal specificities of landscapes.

(2) A local scale for delimiting the UTS using medium spatial resolution satellite imagery. This yields precise landscape classes at a local scale for the UTS. To work at this scale, we used Landsat 5 and 7 imagery with a 30 m by 30 m spatial resolution which shows the main details of landscapes patterns and allows and object-oriented approach to image segmentation. Due to the size of the region studied, the characteristics of Landsat imagery and the availability of acquired scenes with acceptable cloud cover, we worked with a mosaic of six different images, taken on different dates, to cover the whole of Brittany.

The landscape unit enable a synthesis of landscape composition and pattern which can reveal a soil system (Girard and Girard 2003). For example, the European land-use classification, Corine Land Cover, does not reveal the set of parameters required to perform soilscape mapping. Some classes like 'vegetable crops' might enable inference of derived factors like soil hydromorphy. But overall landcover class alone does not enable construction of a pedologic map. Additional and homogeneous landscape information are required all over the area of interest.

This is the reason why a specific landscape classification has been derived combining land use, patterns and structures, and surface moisture. These classes can be hierarchically developed according to the scale of interest.

For MODIS imagery, the classification accuracy is based on statistics of field reference sample plot data and knowledge of the Brittanic landscapes. This step needs many iterations before defining final the representative samples for each class, with a good spread, size and form. The area and number of reference areas varies according to landscape diversity and their inner heterogeneity; for Brittany we have defined 109 reference areas ranging from 750 m square to 4000 m square.

On landsat images, we preferred object–oriented analysis because pixel-based classification is not efficient on images obtained from medium spatial resolution satellite sensors (Whiteside and Ahmad, 2005). Object-oriented approach for the segmentation and classification of Landsat TM images with eCognition software recognises the role of spatial patterns in landscape discrimination, e.g., linear woodland in a valley, large parcels of cereals on plains, small parcels of grass land on hills, etc.

30.2.2 Map of Landscape Units

As landscapes in Brittany are quite complex, with an important diversity in classes and a remarkable heterogeneity in their spatial organisation, classifications based on MODIS and also on Landsat images, can not be use as the final result for mapping landscape units. Post-classification processing is necessary to produce continuous surfaces and to smooth boiundaries for image vectoring (Bou Keir et al., 2004). Common post-classification techniques, such as filters to eliminate small clusters, are inappropriate because of the specific shapes of some landscapes units, e.g., compact shapes (urban zones and woodlands) versus linear shapes (valleys), and well delimited classes versus mixed zones ('bocage'). Unlike land cover maps where parcels have clear boundaries and a single component (crop, grassland, etc.), landscape units are often a specific combination of different classes, and landscapes unit boundaries often have a fuzzy transition (Robbez-Masson et al., 1999).

The method proposed here separates all classes into individual binary layers, which are then individually processed. The purpose of the process is to generalise the spatial representation of each class. This is done using successive iterations of morphological openings and closings (Serra, 1982, 2004) which in turn eliminates small clusters and merges clusters together. Different structuring elements are tested until the size and shape corresponds to the specific shape of the studied class and cartographic scale (e.g., circular structuring elements for open field class, smaller circular elements for forests, linear structuring elements for valleys, etc.). The iterative steps are retained for possible use in a further post-processing step discussed below.

The last step of post-processing is to combine all the processed layers. The raw combination of post-processed layers presents some areas where two or more classes overlap, and some other regions with no class allocated to them.

The combination of overlapping classes is done using a set of ‘expert’ deductive rules for particular cases (e.g., class A overlaps class B, class B overlaps class A, class A mixed with class B, class A not compatible with class B etc.) with particular consequences. If A is mixed with B it may mean that we have to consider a combination between these two classes as a landscape unit itself or as a transitional area between an A landscape unit and a B landscape unit.

Regions with no class allocated are filled using previous step result of intermediate images conserved for each class, recursively, until raw we achieve a full image with no gaps.

30.2.3 Soilscape Boundaries

The landscape map is not sufficient in itself and we have to deal with other environmental data in order to obtain soilscape maps. The most common information available in every country with poor environmental survey are geology and landform, which we choose to fuse with landscape classes to extract soilscape units at a regional scale. To that end, we suggest the use of the theory of evidential fusion. The main principle of data fusion is to combine information from different sensors related to the same observed phenomenon by having a special care to the situations where data are dubious or are in conflict. As we do not have the final soilscape classification, which is not known *a priori*, we focus only on the extraction of the most probable delimitation of soilscape units.

We chose to work on evidential fusion because its theoretical foundations are well adapted to our application. The most popular approach was proposed in Shafer (1976) and is known as the Dempster-Shafer (DS) evidential theory. Unlike methods generally used (e.g. Bayesian techniques), the DS theory was the first one to introduce uncertainty in modelling and which allows expressions of ignorance. The main principle is to deal with two kinds of information imperfection: *probabilistic uncertainty* and *imprecision*. Let us examine these two concepts.

Uncertainty: for a given set of hypothesis, a frame of discernment that integrates the uncertainty is defined. This latter is the set of all possible combinations. For example, if one notes {A,B,C} a set of three hypotheses (in our specific case, this set has only two assumptions and is {boundary, not-boundary}), the frame of discernment is: $\{A, B, C, A \cup B, A \cup C, B \cup C, A \cup B \cup C\}$ and corresponds to all situations that we are likely to deal with. The symbol $\cup$ represents the union of hypotheses and in then related to the uncertainty. For example, $A \cup C$ corresponds to uncertain situations where A and C both hold.

Imprecision: the imprecision is defined through *mass functions*. For each element of the frame of discernment, we have to define a mass function related to the importance of the corresponding hypothesis. This mass is in the interval [0,1] (1: totally sure, 0: totally unsure). The sum of all mass functions has to be 1. For each source of information, we have to define the mass functions for every hypotheses of the frame of discernment. The Dempster rule of fusion combines all mass functions to outputs that are the fused mass functions related to each hypothesis of the frame of discernment (Shafer, 1976).

Although the DS theory has proved to be powerful for combining many sources of information, it suffers from an important limitation: paradoxical information is not taken into account. Indeed, some famous examples have showed that when we have a conflict between information, the DS is likely to exhibit unexpected results. Smarandache and Dezert (2004) have then proposed the Dezert-Smarandache (DSm) theory that extend the DS by introducing the paradox in the frame of discernment. This paradox is represented by the intersection $\cap$ of two hypotheses. A new fusion rule (named the Dezert rule) that integrates this paradoxical information is then defined (Corgne et al., 2003; Smarandache and Dezert, 2004).

30.3 Results and Discussion

30.3.1 Landscape Classification

Landscape classes are defined by expert knowledge of Brittanic landscapes, according to existing landscape classes in other French regions (particularly in the Landscapes Atlases of the Ministry of Environment). A set of 150 reference areas haven been recognised and 109 finally selected after statistical analysis. Depending on the diversity and heterogeneity of landscapes, sample size is adapted. The selection is progressively improved by analyzing the nature of unclassified pixels after each iteration. As landscape class is not a defined object in itself (like land use for example) but a global composition with fuzzy statistical boundaries, a binary validation on a set of testing samples is not sufficient. Elevation is included in the classification processing as a thematic layer to force the emergence of landscape units so that the results (Fig. 30.2a) clearly show the main landscape units at regional scale, but also the complexity of landscape combinations and heterogeneity.

The proposed classification from low-resolution satellite imagery is improved with the use of Landsat imagery and the object-oriented image processing. Segmentation of specific patterns (linear forests in the valleys), topographic information (slope), and particular spectral responses (forest, meadow. . .) improve the accuracy of classification (e.g., specific landscapes units in the valley). Again the technical choices are focused on landscape characteristics and not land cover so that it helps in defining the UTS locally.

(a)

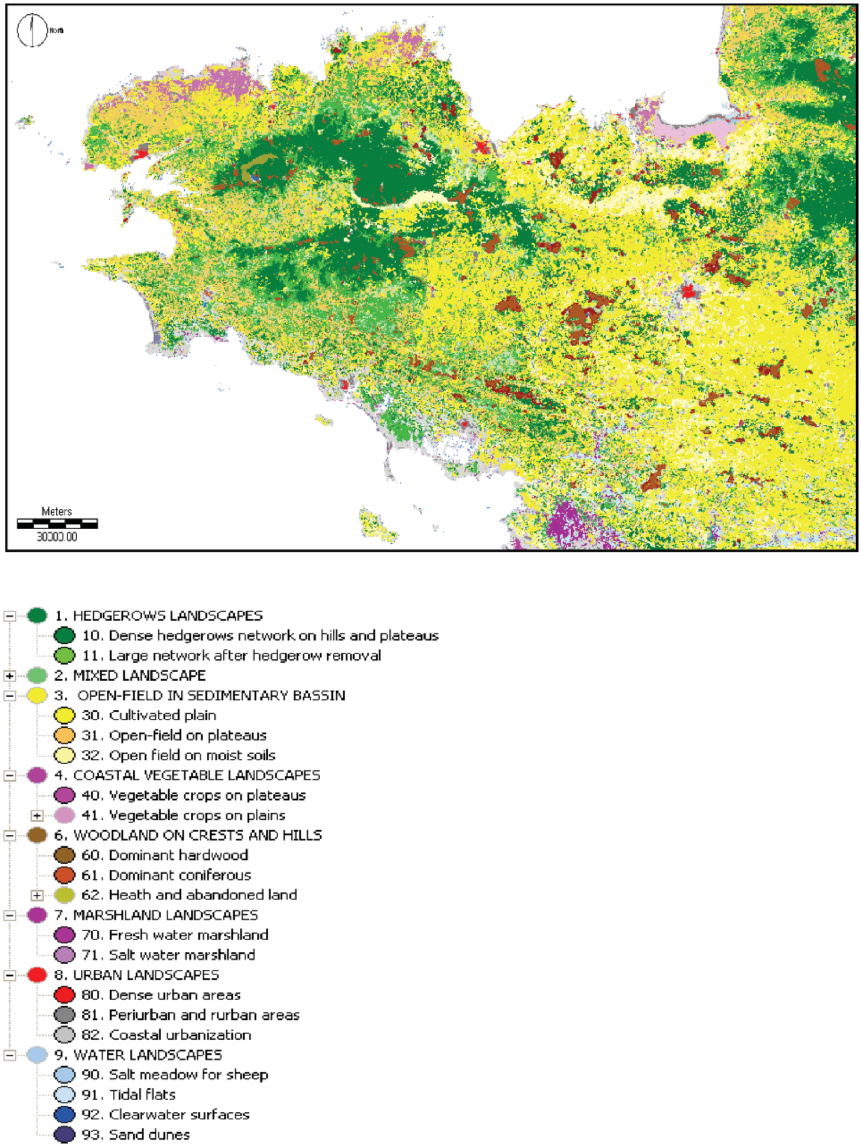


Fig. 30.2a Landscape classification realized by means of MODIS satellite images at regional scale (Phase 1) (See also Plate 40 in the Colour Plate Section)

30.3.2 Map of Landscape Units

The post-classification image (Fig. 30.2b) presents a simplified view of landscape units, ready to be combined with other factors to obtain soilscape units (UPP). Then final combination of overlapping areas depends on the topology of real landscapes. For example, when classes A and B are combined in significant proportion, they are both conserved in a transition zone or transformed into a new class; or when A is included in B, A is deleted except when it is a special class (e.g., urban, woodland etc). Using these expert rules, the final result is very close to the reality and complexity of local and regional landscape units (Canevet et al., 1990).

(b)

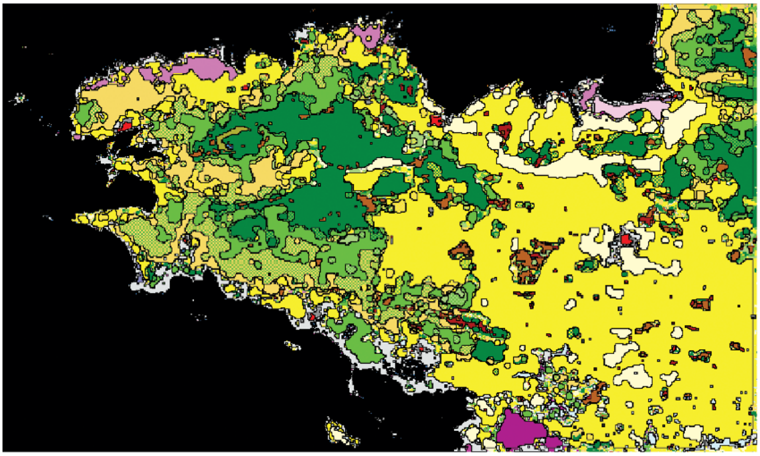


Fig. 30.2b Post-processing techniques applied to landscape classification (Phase 2) (See also Plate 41 in the Colour Plate Section)

30.3.3 Soilscape Boundaries

In our specific application, the inputs are the three choropleth or polygon maps (geology, relief as altitude classes, and landscape). The hypothesis are B, N (B = Boundary, N = Not boundary). The frame of discernment is $\{B, N, B \cup N, B \cap N\}$. The mass function of the hypothesis B is defined by $m(B(\boldsymbol{x})) = \tau^2/(\tau^2 + d^2(x))$ where $d(x)$ is the distance between the point x and the nearest boundary and τ^2 is a stretch parameter. Such a mass function is close to 1 when x is near a boundary and decreases to 0 when x is far from a boundary. The mass function for the hypothesis N is defined as: $m(N(x)) = 1 - m(B(\boldsymbol{x}))$. When we are dealing with a zone which is larger than a typical soilscape unit, the hypothesis N has an important influence in the inside of the region ($m(N) \sim 0$). Hence, this prevents the creation of new boundaries. Actually, from a soilscape unit scale point of view, the information inside such a region is not information of 'no boundaries' but is rather uncertain information. As a consequence, to cope with this problem, we applied the defined mass functions only when we are near a boundary, *i.e.*, $d(x) < \sigma$, σ being a distance to define. We define mass functions as $m(B(x)) = m(N(x)) = 0$ and $m(B \cup N(x)) = 1$.

Results are presented in Fig. 30.3. White represents 0 and black represents 1. Figure 30.3(d) represents the fused map. It indicates the probability of the hypothesis B (presence of a boundary) based on the three input maps.

Two extensions to the methodology are required. The first is the transformation of the probability map into delineated soilscape units. This can be done using some morphological processes such as watershed segmentation (Serra, 1982, 2004). The second consists in fusing on the *nature* of the soilscape units instead of fusing on the *borders*. This would improve the quality of the results but this requires isolating the number and the nature of each soilscape unit in a prior step.

Using this fusion approach, we could then add other environmental data in order to improve the convergence of common boundaries. In Table 30.1 we list the accessible data which may help experts to decide on soilscape boundary location in Brittany.

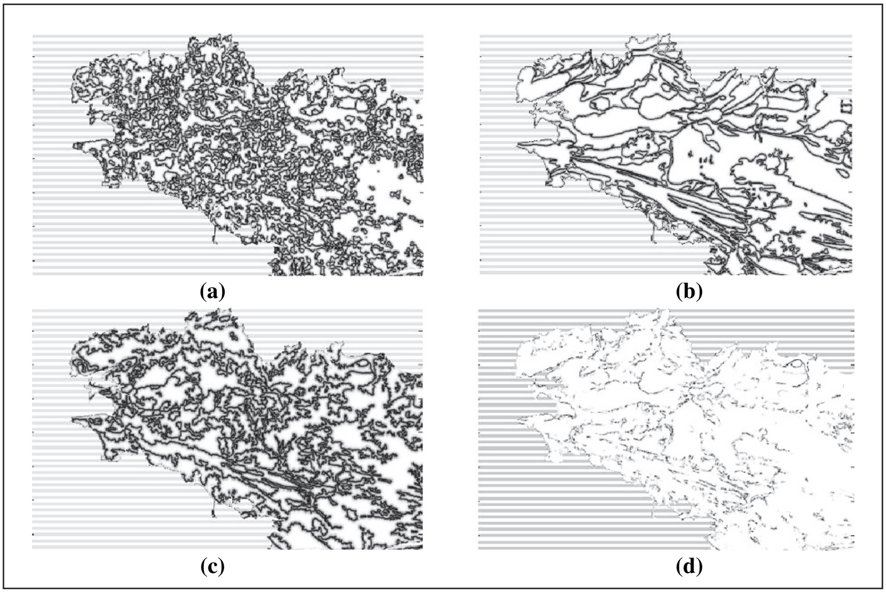


Fig. 30.3 (Phase 3) Mass functions related to the presence of boundaries for (**a**) landscape classification (**b**), geology and (**c**) altitude. Fused map (**d**) indicates the chance of a boundary, *white* represents 0 and *black* 1

Table 30.1 Resolving the lack of soil data using other environmental data

Scale	Map	s	c	o	r	p	a	n
1:250 000	Soilscape units	Data from regional survey in progress	Total rainfall. Max temp. Days $< 0°$	Landscape pixel classification and post classification on MODIS Incl. Surf. moisture	Elevation slopes	Geology Spectral Radiometry K, Th, U	Multidate	Distance to coast
1:50 000	Soil type units	Local validation in progress	Aspect	Landscape object oriented classification on Landsat TM Incl. Surf. moisture	Aspect Curvatures	Surface deposits	Multidate	Distance to talweg

30.4 Conclusions

Remote sensing proved its performance in landscape mapping, both at regional and local extents, as we overcame land-use complexity with specific classes, reference areas and classification techniques. Beyond the accuracy of landscape classification, the object-oriented image analysis has a great potential for DSM for at least two reasons: first because landscape patterns are more stable than the land use they contain, second because these patterns are linked with soil types and can influence soil evolution (erosion, acidification. . .). Post-classification processing is a crucial step for the recognition of the final spatial units and gives a interesting analysis in progressively transitional and mixed landscapes. These results on landscape unit boundaries have to be combined with other factors of the *scorpan* model. Fusion techniques can provide information on shared boundaries from different environmental factors and the probability of soilscape boundaries.

Acknowledgments This work is part of a research program *Sols de Bretagne* run by the SAS (Sol Agronomie Spatialisation) unit at the INRA Agrocampus in Rennes. It is known as *GIS-sol* an dis supported by supported by: IFEN, INRA, MEDD, MAAPAR, ADEME, Conseil Régional de Bretagne, and Conseils Généraux des départements Bretons.

References

Bou Keir R., Girard M.C., Khawlie M., 2004. Utilisation d'une classification structurale OASIS pour la cartographie d'unités de paysage dans une région représentative du Liban. J. Can. Télédétection 30, 617–630.

Canevet C., Laurent L., Le Queau J.-R., Le Rhun P.-Y., Mounier J., Pihan J., Prioul C., Remy F., Le Moignic G., 1990. Atlas de Bretagne, Rennes, coédition Institut culturel de la Bretagne – Skol Vreizh – INSEE, 64 p.

Corgne S., Hubery-Moy L., Dezert J., Mercier G., 2003. Land cover change prediction with a new theory of plausible and paradoxical reasoning. Fusion 2003 Conference, Cairns, Australia, 8–11 July 2003, 1141–1148.

Girard M.C., Girard C.M., 2003, Processing of remote sensing data. Balkema, The Netherlands, 487 p.

Gaddas F., 2001. Proposition d'une méthode de cartographie des pédopaysages – application à la moyenne vallée du Rhône. Thèse de l'Institut National Agronomique de Paris-Grignon, 212pp. http://www.inapg.inra.fr/ens_rech/ager/recherche/theses/these_gaddas.pdf.

Lagacherie P., Robbez-Masson J.M., Nugyen-The N., Barthès J.P., 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma 101, 105–107.

Lecerf R., Hubert-Moy, L., Corpetti, T., Dubreuil, V., 2006. Détermination et suivi de la couverture hivernale des sols à l'échelle régionale par télédétection: évaluation des données EOS/MODIS en paysage fragmenté, Colloque Interaction Nature–Société, analyses et modèles, UMR 6554 LETG Nantes.

Le Du L., 2000. Unités de paysage et télédétection. In Action paysagère et acteurs territoriaux, GESTE n°1, Université de Poitiers, pp. 109–119.

Le Du-Blayo L., 2007. Le paysage en Bretagne, enjeux et défis, 351 p.

McBratney A., Mendonça Santos M.L., Minasny B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Robbez-Masson J.-M., Foltete J.-C., Cabello L., Flitti M., 1999. Prise en compte du contexte spatial dans l'instrumentation de la notion de paysage – application à une segmentation géographique assistée. Revue Internationale de Géomatique 9, 173–195.

Serra J., 1982. Image analysis and mathematical morphology Vol. I. Academic Press, London, 600 p.
Serra J., 2004. http://cmm.ensmp.fr/∼serra/cours.htm (Lectures on Mathematical Morphology).
Shafer G.A., 1976. Mathematical theory of evidence. Princeton University Press, Princeton, New Jersey.
Smarandache F., Dezert J., 2004. Advances and applications of DSmT for information fusion: from evidence to plausible and paradoxical reasoning. American Research Press, Rehoboth.
Whiteside T., Ahmad D W., 2005. A comparison of object-oriented and pixel-based methods for mapping land cover in northern Australia. Proceedings of SSC2005 Spatial Intelligence Innovation and Praxis: The National Biennial Conference of the Spatial Science Institute, September 2005, Melbourne, pp. 1225–1331.

Chapter 31
The Use of GIS and Digital Elevation Model in Digital Soil Mapping – A Case Study from São Paulo, Brazil

G.S. Valladares and M.C. Hott

Abstract This paper applied pedological mapping in an experimental center of "APTA-Frutas" in Jundiaí, São Paulo, Brazil, using morphometric parameters and GIS tools. The aim of this work was to obtain a preliminary legend of a soil map and to compare the preliminary map with maps made by the traditional soil survey methods. The area has 59 hectares and is located at a mountainous relief in the Atlantic Plateau. The original soil map of this area was made at 1:10 000. A digital elevation model (DEM) was generated with 4 m spatial resolution based on a topographical map at 1:10 000 scale, where the level curves are equidistant at 5 m. Based on the DEM we generated altitude, curvature and slope maps. In order to map the hydromorphic soils it was generated a buffer around the hydrography. We also calculated frequency distribution graphics of altitude, curvature and slope maps. After the interpretation of the frequency distribution, we defined classes to predict the soils types. The curvature map was divided into two class intervals ($<$ or $=$ 0 and $>$ 0), the altitude map was divided into four class intervals (690–703, 704–714, 715–730, and 731–757 m), and the slope map was divided into four class intervals (0–9, 10–19, 20–44, and 45–72%). The maps were reclassified and converted to shape files. The shape files were intersected with the others to generate the final preliminary soil map. The methodology was adequate for the preliminary mapping of some types of soils.

31.1 Introduction

This paper applied a pedological mapping methodology (digital soil mapping), in an experimental center of APTA-Frutas (São Paulo State Agribusiness Technology Agency-Fruits) in Jundiaí Municipality, SP, Brazil, using morphometric parameters and GIS (Geographic Information System) tools.

G.S. Valladares
EMBRAPA National Satellite Monitoring Research Center, Campinas, SP, Brazil 13070-115
e-mail: gustavo@cnpm.embrapa.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

The digital elevation models (DEM) (see also Chapter 15) provide information on topography, and derivative products, such as slope that through histograms or reference areas allows to compare with traditional soil map (Lagacherie et al., 1995), as well as to make rules that will be applied to a DEM (McBratney et al., 2003). Both reference areas and histograms need a wide knowledge of the study area to delineate samples (Lagacherie et al., 2001) or to classify soils.

The local landform or relief, represented through DEM, has a major impact on soils by controlling water and sediment movements (McKenzie and Ryan, 1999), together with other factors, such as parent rock.

The aim of our study was to propose a methodology to obtain a preliminary legend of a soil map, which may guide the pedologists in their fieldwork and augment their understanding of the soil-landscape relationship. Previous studies have investigated this topic (Arcoverde et al., 2005; Mühlethaler et al., 2005) and the focus of our work was to compare the preliminary map with the traditional soil maps and to provide an alternative to support decision-making in soil survey planning management.

31.2 Material and Methods

The study area has 59 hectares and is located at Jundiaí, approximately 75 km northwest of São Paulo, Brazil, in a mountainous relief in the Atlantic Plateau (Fig. 31.1). The study area receives 1,409 mm of rain per year with the majority falling between October and March. The land use and land cover are predominantly apple, vineyard, peach, citrus and natural vegetation (Atlantic Forest).

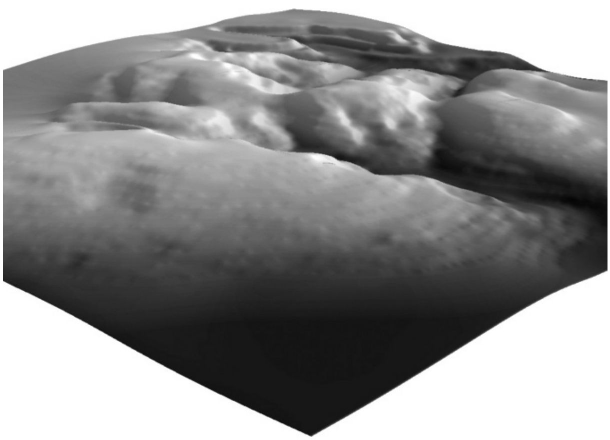


Fig. 31.1 Illustration of the DEM of CAPTA–Frutas, Jundiaí, SP, Brazil in 3D projection

The original soil map of the area was made at 1:10 000 scale (Valadares et al., 1971). It was digitalized and inserted in a GIS. The map's legend was converted to World Reference Base for Soil Resources -WRB (ISSS, 1998).

Using the TOPOGRID function with ArcInfo Workstation GIS available in ArcGIS 9.0 package (ESRI, 2004), a digital elevation model (DEM) with 4 m of spatial resolution (Fig. 31.2a) was generated, based on the 1:10 000 topographical map (Melo and Lombardi Neto, 1999), where the level curves are equidistant at 5 m. Based on the DEM, we generated derivated maps with ArcGIS software, like altitude, curvature and slope maps (Fig. 31.2a, b and c).

In order to map hydromorphic soils, we made a buffer with 7 meters around the hydrography (Fig. 31.5a). We have also made frequency distribution graphics representing altitude, curvature and slope maps. We defined classes to predict the

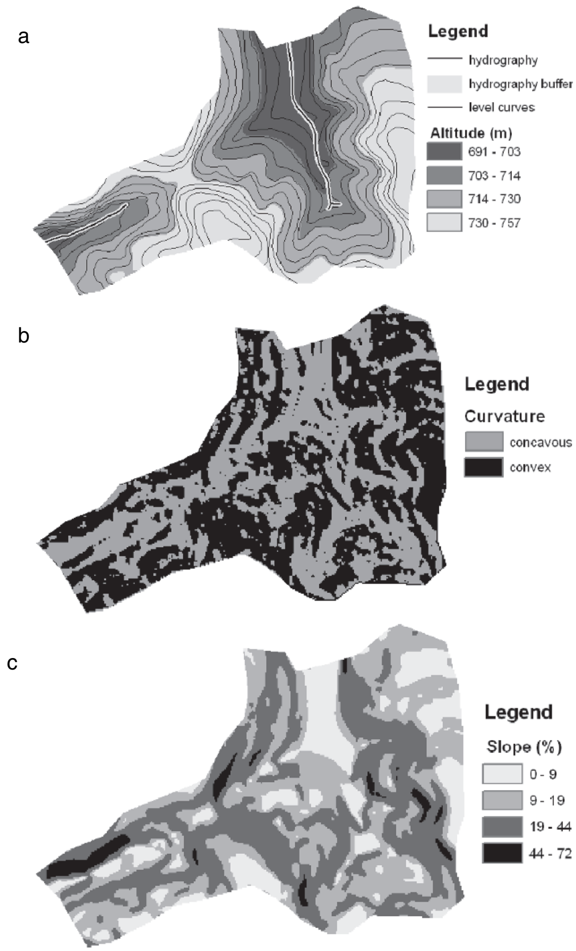


Fig. 31.2 Maps derived from DEM of CAPTA-Frutas, Jundiaí, SP, Brazil. (**a**) altitude with level curves, hydrography and buffer; (**b**) curvature; (**c**) slope

soils types. It was made after visual interpretation of natural breaks in the frequency distribution.

The joint interpretation of all the maps and the INTERSECT function were used to generate the preliminary soil map. The INTERSECT function was applied between the altitude and curvature maps to generated a first version of the preliminary soil map (psoil_1). Then psoil_1 was intersected with the slope map producing a second version of the preliminary soil map (psoil_2). In the last step, the p_soil_2 was intersected with the hydrographic buffer to generate the final preliminary soil map (see also Sections 19.2 and 34.2, using parameters derived from digital models).

31.3 Results and Discussion

The curvature map was divided into two classes in the study area, concave and convex ($<$ or $=$ 0 and $>$ 0), as the mountainous relief plain ground (near 0) is minimally representative. In the concave areas, soils like Dystric Gleysols or Orthic Acrisols are common, while Dystric Cambisols and Xanthic Ferralsols are predominant in the convex areas.

The altitude map varies from 690 to 757 m and was divided into four class intervals (690–703, 704–714, 715–730, and 731–757 m). Fig. 31.3a shows the frequency distribution for altitude. Within the class "690–703 m" all the Dystric Gleysols and a part of the Orthic Acrisols occur, while in the class "higher than 730 m" occur the Dystric Cambisols and the Xanthic Ferralsols. In both intermediate classes (704–714 and 715–730 m) the Orthic Acrisols, Dystric Cambisols and the Xanthic Ferralsols are common. It is not possible to differentiate exactly the soil types using the altitude map.

In the study area, slopes vary from 0 to 72%. The slopes were divided into four class intervals (0–9, 10–19, 20–44, and 45–72%). Fig. 31.3b shows the frequency distribution for the slope classes. Table 31.1 represents a matrix of soil types and the altitude and slope class intervals, without considering curvature.

The maps were reclassified and converted to shape files. In the shape file format, the INTERSECT function was applied to the maps (Fig. 31.4). Firstly, a map

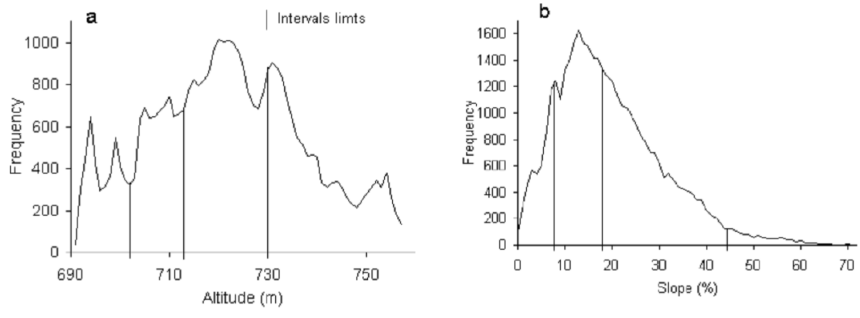


Fig. 31.3 Frequency distribution for altitude (**a**) and slope (**b**)

Table 31.1 Soil types based on altitude and slope from CAPTA-Frutas, Jundiaí, SP

Slope classes (%)	Altitude classes (m)			
	690–703	704–714	715–730	731–757
0–9	Dystric Gleysols	Orthic Acrisols	Xanthic Ferralsols	Xanthic Ferralsols
10–19	Orthic Acrisols	Orthic Acrisols	Orthic Acrisols	Orthic Acrisols
20–44	Orthic Acrisols	Orthic Acrisols	Orthic Acrisols	Orthic Acrisols
45–72	Orthic Acrisols	Orthic Acrisols	Orthic Acrisols	Orthic Acrisols

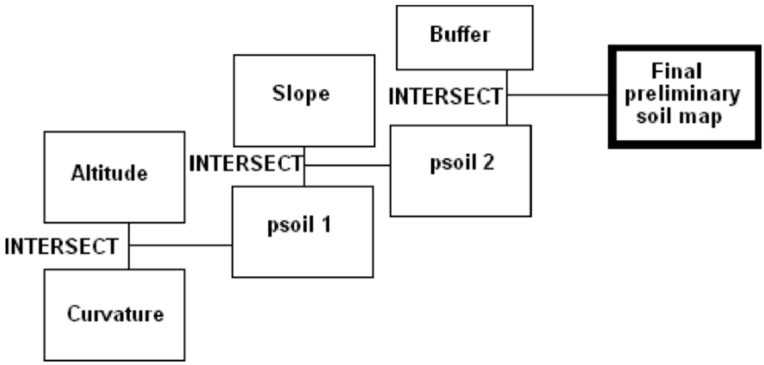


Fig. 31.4 Simplified flowchart for elaboration of the final preliminary soil map

(psoil_1) was generated with the intersection between the altitude and curvature shape files. This new shape file (psoil_1) was intersected with the slope shape file, generating a second version (psoil_2). In the last step, the psoil_2 shape file was intersected with the hydrographic buffer shape file to generate the final preliminary soil map (Fig. 31.5a). Table 31.2 shows the interpretation of the soil types after finishing all the maps' intersections.

The original soil map (Fig. 31.5b) was combined with the digital soil map using the intersect function. For the Dystric Cambisol, the equivalence area was 76%, and for the Dystric Gleysol the equivalence area was 74%. For the Orthic Acrisol, the equivalence was 55% and in the Xanthic Ferralsol the equivalence area was only 15%. The Xanthic Ferralsol was confused with the Dystric Cambisol, because both occur at the same altitude and have similar slope and curvature characteristics, which proved to be a limitation in the proposed approach. Table 31.3 shows that 59% of the area with the Xanthic Ferralsols were classified in the digital soil map as Dystric Cambisols soil and 26% as Orthic Acrisols soil.

In the lower altitude terraces with smaller declivities and concave forms near the streams, the wetlands with hydromorphic soils predominate, and it was classified as Dystric Gleysols (Fig. 31.5a). Comparing Fig. 31.5a and 31.5b, the Dystric Gleysols were overestimated in the northern part, where it was confused with the Orthic Acrisols soil area. Table 31.3 shows that 20% of the Dystric Gleysols area was classified as Orthic Acrisols and 6% as Dystric Cambisols in the preliminary digital soil map (see also example in Fig. 19.1).

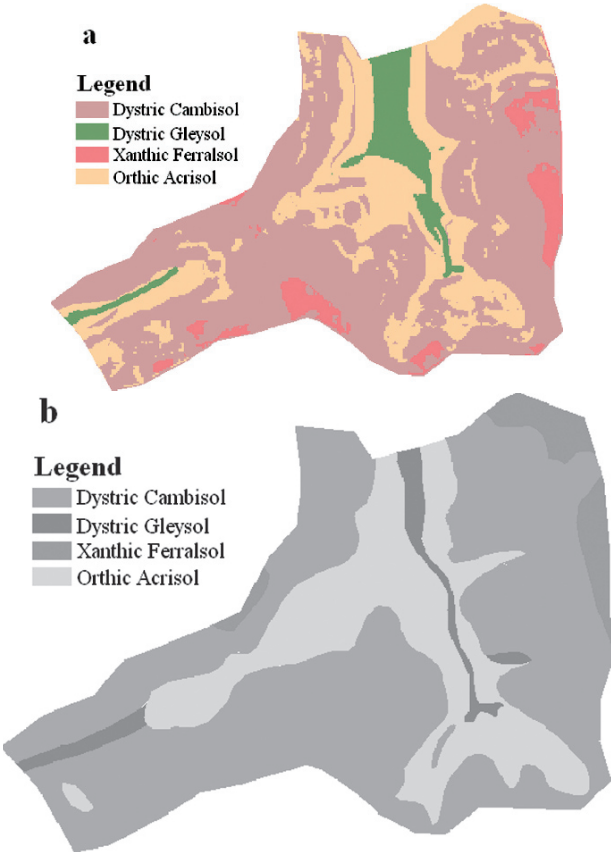


Fig. 31.5 Preliminary digital soil map derived from DEM (**a**), and final soil map elaborated by traditional soil mapping (**b**) of CAPTA-Frutas, Jundiaí, SP, Brazil (See also Plate 42 in the Colour Plate Section)

The Orthic Acrisols are located in the lower part of the slope. These area had previously been underestimated (Fig. 31.5a and 31.5b), where they had been confused with Dystric Cambisols and Dystric Gleysols. Table 31.3 shows that 31% of the Orthic Acrisols area was classified as Dystric Cambisols and 14% as Dystric Gleysols.

The Dystric Cambisols predominated in the study area and were located in the upperslopes and in the higher parts of the landscape. Fig. 31.5 represents the results for this soil type. Table 31.3 shows that 17% of the Dystric Cambisols area was classified as Orthic Acrisols and 7% as Xanthic Ferralsols.

Table 31.2 Soil types defined to produce the preliminary digital soil map after combinations between maps buffer around the hydrograph, curvature, altitude, and slope

Soil Types in Preliminary Digital Soil Map	Combinations hydrography buffer+curvature+altitude+slope
Dystric Gleysol	all sites with hydrography buffer, no buffer + concave + (609–703 m) + (0–9%), no buffer + convex + (609–703 m) + (0–9%) no buffer + concave + (704–714 m) + (0–9%), no buffer + convex + (704–714 m) + (0–9%), no buffer + concave + (609–703 m) + (10–19%), no buffer + convex + (609–703 m) + (10–19%),
Orthic Acrisol	no buffer + concave + (609–703 m) + (> 19%), no buffer + convex + (609–703 m) + (> 19%), no buffer + concave + (704–714 m) + (10–19%), no buffer + concave + (715–730 m) + (10–19%), no buffer + concave + (715–730 m) + (0–9%) no buffer + concave + (704–714 m) + (> 19%), no buffer + convex + (704–714 m) + (> 19%), no buffer + concave + (715–730 m) + (> 19%),
Dystric Cambisol	no buffer + convex + (715–730 m) + (> 19%), no buffer + concave + (> 730 m) + (10–19%), no buffer + convex + (> 730 m) + (10–19%), no buffer + concave + (> 730 m) + (> 19%), no buffer + convex + (> 730 m) + (> 19%) no buffer + convex + (704–714 m) + (10–19%), no buffer + convex + (715–730 m) + (> 19%),
Xanthic Ferralsol	no buffer + convex + (715–730 m) + (0–9%), no buffer + concave + (> 730 m) + (0–9%), no buffer + convex + (> 730 m) + (0–9%)

Table 31.3 Soil types correspondence area (%) for traditional and preliminary soils maps from CAPTA-Frutas, Jundiaí, SP

	Traditional Soil Map			
Preliminary Digital Soil Map	Dystric Cambisol	Dystric Gleisol	Xanthic Ferralsol	Orthic Acrisol
Dystric Cambisol	76	6	59	31
Dystric Gleysol	0	74	0	14
Xanthic Ferralsol	7	0	15	0
Orthic Acrisol	17	20	26	55
Total	100	100	100	100

31.4 Conclusions

The proposed methodology was adequate to identify some types of soils using GIS, and showed the importance of relief in the Atlantic Plateau soils' formation. In order to produce a detailed soil map using this methodology, additional fieldwork is necessary. For the Dystric Cambisol, the Dystric Gleysol, the Orthic Acrisol, and the Xanthic Ferralsol the equivalence area was respectively, 76%, 74%, 55% and 15%.

The Dystric Cambisols and the Xanthic Ferralsols predominated in the upper-slopes and in the higher parts of the landscape. In the lower altitude terraces, which have smaller declivities and concave forms near the streams, predominate the wetlands with hydromorphic soils, classified as Dystric Gleysols. The Orthic Acrisols are located in the lower part of the slope.

Soils are function of five formation factors: parent rock, relief, vegetation, climate and time. In this study we considered only the relief factor. For large areas with lesser scales, other soil formation factors may be included in the analysis, with the purpose of obtaining satisfactory results.

References

Arcoverde, G.F.B., Borges, M.E.S., Martins, E.S., Ramos, V.M., Guimarães, R.F., Carvalho Junior, O.A., Gomes, R.A.T., 2005. Mapeamento pedológico em relevos cársticos a partir da análise morfométrica. XII Simpósio Brasileiro de Sensoriamento Remoto. INPE, Goiânia, Brazil, pp. 1725–1732.

Environmental Systems Research Institute (ESRI). 2004. ArcInfo 9.0. Redlands, CA.

ISSS International Society of Soil Science Working Group, 1998. World Reference Base for Soil Resources. ISRIC, FAO, Rome.

Melo, A.R., Lombardi Neto, F., 1999. Planejamento Agroambiental do Centro Avançado de Pesquisa do Agronegócio de Frutas. IAC/APTA, Campinas.

Mühlethaler, B., Ramos, V.M., Carvalho Junior, O.A., Guimarães, R.F, Bettiol, G.M., Gomes, R.A.T., Martins, E.S., Reatto, A., 2005. Avaliação do uso da morfometria como suporte para a elaboração de mapa pedológico na bacia do Ribeirão da Pedreira – DF. XII Simpósio Brasileiro de Sensoriamento Remoto. INPE, Goiânia, Brazil, pp. 3167–3174.

Lagacherie, P., Legros, J.P., Burrough, P.A., 1995. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. Geoderma 65, 283–301.

Lagacherie, P. Robbez-Masson, J.M., Nguyen-The, N., Barthès, J.P., 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma 101, 105–118.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.

Valadares, J., Lepsch, I.F., Küpper, A., 1971. Levantamento pedológico detalhado da Estação Experimental de Jundiaí, SP. Bragantia 30, 337–386.

Chapter 32
Geomorphometric Attributes Applied to Soil-Landscapes Supervised Classification of Mountainous Tropical Areas in Brazil: A Case Study

W. Carvalho Junior, E.I. Fernandes Filho, C.A.O. Vieira, C.E.G.R. Schaefer and C.S. Chagas

Abstract The present study aimed to improve the recognition of patterns of soils organization in mountainous tropical landscapes, hence helping soil surveys. The study area is located in the northwest Rio de Janeiro State, with a total area of approximately 16.470 ha. In this concern, geomorphometric features that define the geomorphic signature of the soil-landscape, were used. Geomorphometric features includes: elevation, relative elevation, aspect, curvature, curvature plane, curvature profile, slope, flow direction, flow accumulation and drainage's Euclidian distance, being all these features obtained by geoprocessing techniques. Almost all attributes were obtained from a digital elevation model and, therefore, the primary elevation data were obtained from the topographic maps. Through these geomorphometric attributes, a geomorphometric signature of the landscape was elaborated, and the particularities of each soil-landscape unit improved the supervised classification. The results showed the feasibility of using geomorphometric attributes to perform a supervised classification, using either neural networks or a maximum likelihood algorithm for soil-landscapes classification of mountainous tropical areas. In addition, we showed that geoprocessing techniques used to extract geomorphometrics attributes can subsidize soil surveys, making soil mapping faster and less biased by subjectivity.

32.1 Introduction

The study of relationships between soil and landscape, through geoprocessing techniques, requires quantified geomorphometric parameters, which can be understood as a continuous numeric description of a surface (Wood, 2000). In geomorphologic terms, it can be understood as a group of values that describes the landform in a way that allows to distinguish topographically different landscape. This way, the terrain parameters need to be sensitive to the geomorphological processes.

W. Carvalho Junior
Embrapa Solos. Rua Jardim Botânico, 1024. Rio de Janeiro. RJ. Brazil. CEP 22.460-000
e-mail: waldircj@cnps.embrapa.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

The quantification of the morphologic features of the earth surface is essential for the knowledge of physical, chemical and biological processes that take place in the landscape (Blaszczynski, 1997). Therefore, landforms influence, water flow, sediment transport, and the nature and distribution of habitats of plants and animals. In addition, it also expresses weathering processes that act on the formation of the landscape.

In the traditional cartography, based on discreet models of space variability, attributes are considered to change abruptly, contrary to what occurs in the nature and is assumed in this study. Thus, continuous models are used in a different approach to represent the space variability of terrain attributes, assuming that these attributes vary gradually through the space.

The terrain digital analysis available through geographical information systems (GIS) is a fast and economic alternative that can be applied to the establishment of geomorphometric attributes in any portion of the earth surface, as a relatively simple and precise procedure, using a raster data structure (Sections 7.4 and 28.4)

The raster representation assumes that the geographical space can be represented as a regular grid of pixel or cell. Each pixel is then associated with a portion of the landscape. The resolution - or the scale of the raster data, is represented by the relationship between the cell size and its representation on the ground (Burrough, 1986). Chou (1997) highlights the advantages of using raster format such as: efficiency in the data processing; the varying sources of available data, such as satellites images, aerial photographs and elevation models; and finally, the possibility of organization of different features types in a same information layer. Another advantage of the raster models is that they allow to visualize the spatial distribution of a given attribute, comparing its distribution with other attributes (Mendonça Santos et al., 2000). As disadvantages, Chou (1997) mentions the data redundancy, the confusion in cell resolution and the difficulty of assigning cell values.

There is little work on continuous modeling of soil and landscape, particularly in the tropics. Against this background, this work aimed the study of the relationship between soils and landscapes of the highland “Mares de Morros” domain in the Northwest Rio de Janeiro State, emcompassing the municipalities of “Varre Sai”, “Porciúncula” and “Natividade”, trying to recognize the arrangement and distribution of geomorphological and pedological components with the help of geoprocessing techniques, in order to characterize the geomorphometric pattern of each defined soilscape unit.

32.2 Material and Methods

The study area covers approximately 16.470 ha, contained in the topographical chart of “Varre Sai” sheet (IBGE, 1991). The digital cartographic base contains contour lines with vertical 20 m equidistance and drainage network mapped at 1:50.000 scale, using the UTM projection system.

For the development of this study, the following softwares were used: ARC/INFO version 8.2 (ESRI, 1994) and ArcView GIS version 3.2a (ESRI, 1996a and 1996b);

ERDAS IMAGINE version 8.5 (ERDAS Systems, 2001); and Microsoft Excel–2000 (Microsoft Corporation).

The soilscapes units represent groups of the soil survey map units of the Rio de Janeiro State (EMBRAPA SOLOS, 2003). Thus, five soilscape units were defined: Oxisols (P1), Ultisols (P2), Aquent Entisols (P3), Inceptisols (P4) and Rock Outcrops (P5). The geomorphometric parameters used to elaborate the soilscapes patterns were obtained from a digital model using the raster format, described as follows: Digital Elevation Model (DEM), derived from primary elevation data (contour lines and surface-specific points); drainage network and limit, through interpolation; slope (D), generated from the DEM, aspect (A), expressed as azimuth; curvature surfaces or curvature planes (CP) and curvature profile (PC), where positives values are related to convex surface, negative are related to concave surface, and close to zero are defined as a flat surface; superficial flow direction (SFD) in relation to descending gradient; superficial flow accumulation (SFA), based on the number of cells that flow for a certain cell; relative elevation of sub-basins (RESB), that describes the variation of altimetric quota of the contribution basin of each segment of the drainage network (Carvalho Junior, 2005); and hidrography Euclidian distance (DEH), that describes the relationship between the distance of each cell and the drainage. The relative areas for each soilscape can be observed in Table 32.1.

The DEM was generated from the contour lines and surface-specific points over the entire area, obtained from the topographical chart "Varre Sai" (IBGE, 1991), to which drainage lines and a polygonal feature of interpolation limit, were added. The data processing was performed using the ARC/INFO software using the "TOPOGRIDTOOLS" option that composes an interpolation method drawn to create hydrologically consistent digital elevation models. Before the interpolation process, a preliminary data pre-processing was performed, aiming to correct the orientation, simplify the hidrography and graphically adjust the contour lines, characterizing the line of the bottom valley in relation to the drainage lines.

The DEM resolution definition was in agreement with proposition of Hutchinson & Gallant (2000), associating the visual comparison between the original contour lines with those generated from the DEM, as well as the analysis of the total amount of spurious depressions of each resolution. According to Carvalho Junior (2005), the DEM evaluation of this area showed best results using a cell size of 20 or 30 m.

Since soil distribution can be spatially predicted from geographic position using a variety of techniques (McBratney et al., 2003; Section 31.4), the soilscapes were

Table 32.1 Area values and percentage for each soilscape units in the study area

Soilscapes	Area (Ha)	%
P1– Oxisols	9.141	55,5
P2– Ultisols	4.658	28,3
P3– Aquent Entisols	1.080	6,6
P4 – Inceptisols	1.517	9,2
P5 – Rock outcrops	75	0,5
Total	16.470	100

individualized by statistical parameters (maximum, minimum, medium and standard deviations values) relative to the considered attributes (Elevation, D, THE, CP, PC, SFS, ESA, ARSB and DEH). It was observed that the considered attributes had, in some cases, a distribution close to the normality, while others did not follow a normal distribution.

32.3 Results and discussion

The DEM's derived contour lines with 20 and 30 m resolution were those closer to the primary contour lines data (Carvalho Junior, 2005), with a number of spurious depressions of 143 and 224, respectively. Thus, we defined the best resolution for the DEM grid with a 20 m of cell size for this work (Fig. 32.1). This DEM had the following statistical characteristics: maximum value of 1.135,9 m, minimum of 197,7 m, average of 628,2 m and standard deviation of 154,4 m.

The spatial analysis of the geomorphometric pattern of the soilscapes studied showed the following statistical values (Table 32.2).

In relation to the distribution of attributes in soilscape P1, we verified that the attribute A has a trend of higher frequency of north facing slopes (315–45°), representing 28% of the total. The Eastern (45–135°) and Southern (135–225°) facing slopes, have 26% each, and Western facing slopes (225–315°) represent 20% of the total area.

With reference to the ESA attribute, despite of a broad range of variation (from 0 to 192.000), most data (96,4%) area within the range of 0–100 for soilscape P1.

In soilscape 2, the distribution of the attribute A showed a dominance of Southern-facing slopes (direction 135–225°), with 31% of the total, followed by West, East and North directions, in decreasing order. With reference to ESA, similarly to soilscape 1, despite the large variability (range of 0–235,288), most values (97,6%) are within the range of 0–500.

In soilscape P3, 30,8% of its area is Northern facing slopes, followed by Eastern (28,3%) and South and West with less significance. The ESA attribute was also very variable (range: 0–241.766), but most values (87.2%) fall within the range of

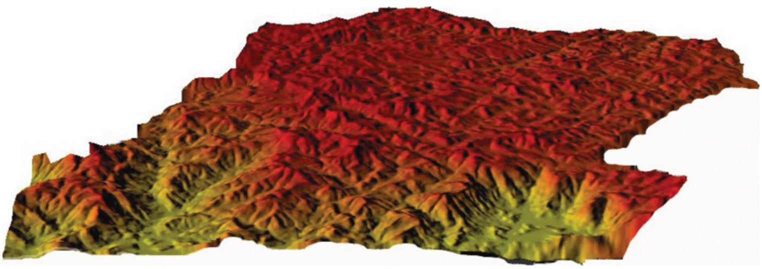


Fig. 32.1 3D visualization with 3x of vertical exaggeration of hydrologic correct DEM with resolution of 20 m (See also Plate 43 in the Colour Plate Section)

Table 32.2 Values maximum, minimum, medium and standard deviations of the geomorphometric attributes of the soilscapes

Attributes	Maximum	Minimum	average	Standard deviation
Soilscape P1 (Oxisols)				
Elevation (m)	938,8	503,6	706,0	62,0
D–slope (%)	124,9	0,0	27,9	16,5
PC–curvature profile	6,3	−4, 7	0,021	0,5
CP–Curvature planes	4,6	−4, 5	0,037	0,3
SFS– superficial flow direction	128,0	1,0	35,5	44,4
ESA– superficial flow accumulation	192.326,0	0,0	213,8	4.215,0
ARSB– relative elevation of sub-basins (m)	505,8	0,0	149,8	54,0
DEH– Euclidian distance (m)	1.381,6	0,0	125,1	94,1
A–Aspect (degrees)	360,0	0,0	169,7	105,8
Soilscape P2 (Ultisols)				
Elevation (m)	1.136,0	197,7	481,5	159,2
D–slope (%)	131,4	0,0	34,6	18,0
PC–curvature Profile	6,7	−3, 9	0,037	0,5
CP–Curvature planes	4,9	−3, 6	0,028	0,4
SFS– superficial flow direction	128.000,0	1,0	25,4	37,2
ESA– superficial flow accumulation	235.288,0	0,0	334,3	5.428,2
ARSB– relative elevation of sub-basins (m)	626,9	0,0	262,5	111,6
DEH– Euclidian distance (m)	863,5	0,0	131,1	98,9
A – Aspect (degrees)	360,0	0,0	184,7	94,8
Soilscape P3 (Aquent Entisols)				
Elevation (m)	795,1	197,7	579,7	160,0
D–slope (%)	83,1	0,0	10,9	10,7
PC–curvature Profile	5,1	−2, 4	0,176	0,4
CP–Curvature planes	3,2	−3, 3	0,007	0,2
SFS– superficial flow direction	128,0	1,0	33,5	42,6
ESA– superficial flow accumulation	241.766,0	0,0	3.145,3	17.748,8
ARSB– relative elevation of sub-basins (m)	626,9	0,0	139,4	71,2
DEH– Euclidian distance (m)	506,4	0,0	51,9	47,8
A–Aspect (degrees)	360,0	0,0	165,0	109,9
Soilscape P4 (Inceptisols)				
Elevation (m)	1.017,9	435,8	747,5	104,2
D–slope (%)	111,0	0,2	39,5	17,2
PC–curvature Profile	4,6	−2, 7	−0, 018	0,5
CP–Curvature planes	2,8	−3, 3	0,030	0,4
SFS– superficial flow direction	128,0	1,0	27,7	39,4
ESA– superficial flow accumulation	14.724,0	0,0	23,6	184,1
ARSB– relative elevation of sub-basins (m)	549,8	59,7	260,6	65,9
DEH– Euclidian distance (m)	440,0	0,0	132,9	84,0
A–Aspect (degrees)	360,0	0,0	182,3	95,8

Table 32.2 (continued)

Attributes	Maximum	Minimum	average	Standard deviation
Soilscape P5 (Rock outcrops)				
Elevation (m)	785,1	284,6	567,0	102,7
D–slope (%)	118,7	0,6	53,9	20,0
Soilscape P5 (Rock outcrops)				
PC–curvature Profile	4,0	−3, 2	−0, 207	0,6
CP–Curvature planes	5,0	−2, 5	0,201	0,5
SFS– superficial flow direction	128,0	1,0	20,8	31,5
ESA– superficial flow accumulation	914,0	0,0	5,5	36,2
ARSB– relative elevation of sub-basins (m)	626,9	121,9	380,7	99,6
DEH– Euclidian distance (m)	640,0	0,0	221,3	119,4
A–Aspect (degrees)	360,0	0,0	197,5	87,9

0–100. With reference to the superficial accumulated flow, 10,3% of its area had values greater than 600, due to a predominant hydromorphic landscape, with large floodplains.

For soilscape P4, the A attribute showed 32,2% of the area with Southern facing slopes, followed by the West facing with 25,3%, East (21.9%) and North (20.7%). The ESA attribute had much less variability than the former soilscapes (range 0–14.724), and most values (97,3%) fall within the range of 0–100. Values greater than 100 represent only 2,7%, due to the highly dissected, mountainous nature of the landscape, with scarps and steep slopes dominating.

In the P5 landscape, 40,0% of the area are Western-facing slopes, followed by Southern (27,5%), East (20.5%) and North (12.7%) facing slopes. The relative values of ESA showed less variability compared with all others (range: 0–914), and almost all values fall within the range of 0–100. This is consistent with the highly mountainous and deeply dissected landscape, typical of this unit.

The grouped analysis of the geomorphometric attributes of the soilscapes allowed the evaluation of their main differences. Table 32.3 presents the mean values and standard deviation of the geomorphometric attributes for each soilscape.

Regarding the mean elevation values of the soilscapes, it was observed that they occupy topographically different positions, except for the soilscapes P3 and P5 (respectively Aquent Entisols and Rock Outcrops). However, P3 soilscape showed a greater altimetric standard deviation, indicating its widespread occurrence.

The smallest slope average occurs in soilscape P3 (Aquent Entisols), whereas the largest is in P5 (Rock Outcrops), consistent with the typical subdue relief of hydromorphic soils, and highland position of most rock outcrops. On the other hand, soilscape P5 showed the steepest mean slope, in areas where erosive processes are dominant.

The attributes PC and CP showed higher mean values, either positive or negative, in soilscape P5, thus indicating a landscape of strong concavity or convexity, suggesting a high degree of dissection and drainage incision. The largest mean value of ESA is associated with soilscape P3, as expected, in view of high values for cells of this

Table 32.3 Average and standard deviation of the geomorphometric attributes for each soilscape of the study area

Soilscapes	Elevation (m)		D (%)		PC		CP	
	mean	sd	mean	sd	mean	sd	mean	sd
P1	706,0	62,0	27,9	16,5	0,021	0,5	0,037	0,3
P2	481,5	159,2	34,6	18,0	0,037	0,5	0,028	0,4
P3	579,7	160,0	10,9	10,7	0,176	0,4	0,007	0,2
P4	747,5	104,2	39,5	17,2	−0, 018	0,5	0,030	0,4
P5	567,0	102,7	53,9	20,0	−0, 207	0,6	0,201	0,5

Soilscapes	SFS		ESA		ARSB (m)		DEH (m)		A (°)	
	mean	sd	mean	Sd	mean	sd	mean	sd	mean	sd
P1	35,5	44,4	213,8	4.215,0	149,8	54,0	125,1	94,1	169,7	105,8
P2	25,4	37,2	334,3	5.428,2	262,5	111,6	131,1	98,9	184,7	94,8
P3	33,5	42,6	3.145,3	17.748	139,4	71,2	51,9	47,8	165,0	109,9
P4	27,7	39,4	23,6	184,1	260,6	65,9	132,9	84,0	182,3	95,8
P5	20,8	31,5	5,5	36,2	380,7	99,6	221,3	119,4	197,5	87,9

Where: mean = average and sd = standard deviation

attribute indicate areas of high water flow concentration that identify drainage channels or rivers. On the other hand, cells of this attribute with accumulated flow close to 0 indicate areas of high topography that identify the gentle sloping watershed, such as in the cases of soilscapes P4 and P5, which possess the lowest mean values.

The mean value of the ARSB attribute for each soilscape reflects the equilibrium between pedogenetic and morphogenetic processes. In this sense, the lowest values occurs in soilscapes P3 and P1 (respectively Aquent Entisols and Oxisols), where pedogenesis appears to be more active than morphogenesis. On the other hand, the largest values of this attribute, in decreasing order from soilscapes P5, P2–P4, indicate that morphogenesis and erosion acts with greater intensity in these soilscapes.

The differences between the soilscapes can be graphically observed in Fig. 32.2.

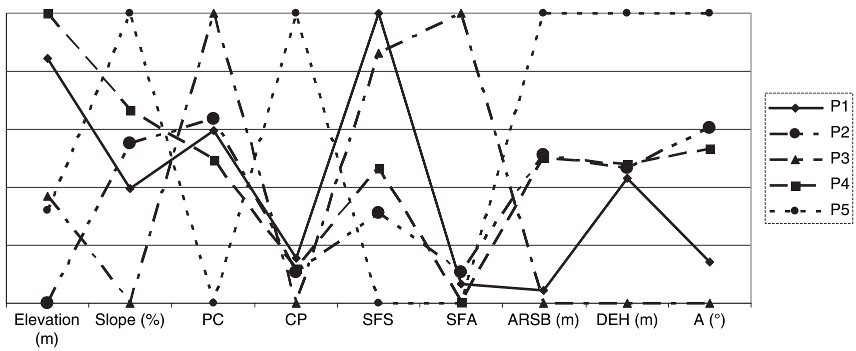


Fig. 32.2 Geomorphometric patterns of the five soilscapes (the mean values for each variable were re-scaled between 0 and 1, for the proportion of the class with maximum mean value being 1 and with minimum mean value being 0, with the maximum indicated by the top of the figure)

Despite the calculated mean value represents the trend of behavior for populations that show a normal distribution pattern, which is not the case of all geomorphometric attributes used in this study, we can expected that the variables showed in Fig. 32.2 contribute to enhance the differentiation between the soilscapes, hence serving as input variables for a digital mapping with use of supervised classifiers, like done by Carvalho Junior (2005) with an artificial neural network.

32.4 Conclusions

- The more accurate the digital elevation model (DEM), the greatest representativeness will be attained for derived attributes. This, in turn, is directly dependent upon the primary source of elevation data that define the spatial grid resolution;
- The derived attributes of DEM can contribute, separately or together, to understand the soil-landscape relationship of dissected tropical areas of "Mares de Morros";
- The techniques of digital soil mapping, from the attributes described in this work, allowed an easier differentiation of the defined soilscape, and help to produce a digital classification through techniques of supervised classification;
- This approach is very useful for applying to environmental surveys in Brazil, where availability of detailed soil data is limited.

References

Blaszczynski, J. S. Landform characterization with geographic information Systems. Photogrammetric Engineering and Remote Sensing. 63(2): 183–191, 1997.

Burrough, P. A. Principles of geographical information systems for land resources assessment. New York, Oxford University Press Inc, 194p. 1986.

Carvalho Junior, W. Classificação Supervisionada de Pedopaisagens do Domínio dos Mares de Morros Utilizando Redes Neurais Artificiais. Viçosa, UFV. 143p. 2005. (Doctorat Thesis)

Chou, Y. H. Exploring spatial analysis in geographic information systems. Santa Fe, OnWord Press, 478p. 1997.

EMBRAPA SOLOS. Mapa de Solos do Estado do Rio de Janeiro. Escala 1:50.000. 2003.

Environmental System Research Institute. ESRI. ARC/INFO, Command references. New York, p.281. 1994.

Environmental Systems Research Institute. ESRI. ArcView GIS. The Geographic Information System for everyone. New York, 1996a.

Environmental Systems Research Institute. ESRI. ArcView Spatial Analyst: advanced spatial analysis using raster and vetor data. New York, 1996b.

ERDAS IMAGINE 8.5. ERDAS IMAGINE INC. Atlanta, GA. 638p. 2001.

Hutchinson, M. F. and Gallant, J. C. Digital elevation models and representation of terrain shape. In: Wilson, J. P. and Gallant, J. C. (Eds). Terrain Analysis: Principles and applications. John Wiley & Sons, Inc. 2000.

Instituto Brasileiro de Geografia e Estatistica (IBGE). Diretoria de Geociências. Departamento de Cartografia. Varre-Sai. Folha SF.24-V-A-IV-3. Rio de Janeiro, Escala 1:50.000. 1991.

McBratney, A. B., Mendonça Santos, M.L. and Minasny, B. On digital soil mapping. Geoderma 117, 3–52, 2003.

Mendonça Santos, M. L., Guenat, C., Bouzelboudjen, M. and Golay, F. Three-dimensional GIS cartography applied to the study of the spatial variation of soil horizons in a Swiss floodplain. Geoderma 97, 351–366, 2000.

Wood, J. The geomorphological characterization of digital elevation models. PhD Tesis. City University. London. Disponível em: http://www.soi.city.ac.uk/~jwo/phd. 185p. 2000.

Chapter 33
Digital Soil Mapping of Soil Properties in Honduras Using Readily Available Biophysical Datasets and Gaussian Processes

Juan Pablo Gonzalez, Andy Jarvis, Simon E. Cook, Thomas Oberthür, Mauricio Rincon-Romero, J. Andrew Bagnell and M. Bernardine Dias

Abstract Creating detailed soil maps is an expensive and time consuming task that most developing nations cannot afford. In recent years, there has been a significant shift towards digital representation of soil maps and environmental variables and the associated activity of predictive soil mapping, where statistical analysis is used to create predictive models of soil properties. Predictive soil mapping requires less human intervention than traditional soil mapping techniques, and relies more on computers to create models that can predict variation of soil properties. This paper reports on a multi-disciplinary collaborative project applying advanced data-mining techniques to predictive soil modelling for Honduras. Gaussian process models are applied to map continuous soil variables of texture and pH in Honduras at a spatial resolution of 1 km, using 2472 sites with soil sample data and 32 terrain, climate, vegetation and geology related variables. Using split sample validation, 45% of variability in soil pH was explained, 17% in clay content and 24% in sand content. The principle variables that the models selected were climate related. Gaussian process models are shown to be powerful approaches to digital soil mapping, especially when multiple explanatory variables are available. The reported work leverages the knowledge of the soil science and computer science communities, and creates a model that contributes to the state of the art for predictive soil mapping.

33.1 Introduction

Statistical Soil Modeling is the development of statistical soil models for large areas based on soil samples and digital maps of environmental variables. It is also known in the literature as predictive soil mapping. Recent scientific advances in

J.P. Gonzalez
Robotics Institute, Carnegie-Mellon University, Pittsburgh, Pennsylvania, USA
e-mail: jpgonzal@cmu.edu

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

soil-landscape modeling have demonstrated the power of predictive modeling of soil characteristics (including texture, moisture, pH, and some nutrients) at high resolution. These advances are built on statistically defined relationships between observable features of the landscape as well as improved understanding of processes that control soil formation. At the same time, significant advances have been made in the availability of high resolution data on many of the driving mechanisms of soil variability, especially terrain, climate and land-cover.

There is a significant amount of research in predictive soil mapping. For a thorough review of existing approaches to predictive soil mapping see references within this book as well as Nachtergaele (1996), Scull et al. (2003) and Heuvelink and Webster (2001). However, most of the work in predictive soil mapping has been done for temperate zones, corresponding to North America, Europe and Australia. This is due in part to limited spatial data infrastructures, as well as a scarcity of funding for basic generation of data and information. Very little research has been done in developing appropriate predictive soil mapping techniques for the tropics. The tropics have different climate patterns than temperate zones, and different processes behind soil formation rendering some predictive soil mapping models developed for North America, Europe or Australia less applicable.

The world is currently witnessing a growing demand for technological innovation to empower developing communities (Sachs, 2002). Inspired by the current demand for advanced technology relevant to developing communities, this paper focuses on the topic of applying Machine Learning techniques to the problem of soil mapping in the tropics. In recent years, there has been a significant shift towards digital representation of soil maps and environmental variables that has created the field of predictive soil mapping (Scull et al., 2003). In predictive soil mapping, statistical analysis is used to create predictive models of soil properties, thus requiring less human intervention than traditional soil mapping techniques, and relying more on computers to create models and predict soil properties. This technique is highly relevant for improving soil information in the tropics to respond to the demands for soil data to improve natural resource management and aid communities to better manage their resources and respond to global changes.

The goal of this project is to develop statistical soil models for Honduras, and create a model that matches or advances the state of the art for predictive soil mapping, with relevance to tropical countries. Specifically, the objective was to model and predict variations in topsoil pH content, clay content and sand content for the whole country at the highest spatial resolution possible. This research is developed within the context of limited data infrastructures that many tropical countries experience, and focuses on using widely available spatial data in the development of the models. Honduras was selected as a case study site representative of many tropical countries. Honduras is a small tropical country (112 000 km^2), but in spite of its small size, Honduras has coastal and mountainous areas, elevations from 0 to 2870 meters, and temperatures from 10 to 30°C.

33.1.1 Traditional Soil Maps

Currently, 68% of the countries of the world have soil maps at 1:1 000 000 or finer. However, these countries only represent 31% of the world's land surface. Most of the remaining 69% corresponds to developing countries (Nachtergaele, 1996). Since 1981 the world has a global soil map at a scale of 1:5 million. The maps, published by FAO and UNESCO, were based on soil surveys conducted in the 1930s to 1970s. This map provides worldwide coverage at 1:5 000 000 and has been converted into Soil Taxonomy (Soil Survey Staff, 1975), which classifies the soils in 12 main categories (*soil orders*) with subcategories. For many developing countries this is the only current source of soil information.

In Honduras, there is a partial map of soils at 1:250 000 scale produced in 1962 with vast areas of the country (> 80%) not classified (Selvaradjou et al., 2005) and two region specific agro-ecological soil maps with very basic soil information (see http://eusoils.jrc.it/esdb_archive/EuDASM/latinamerica/lists/chn.htm). The only alternative is the FAO world map, shown in Fig. 33.1. The FAO soil map of the world is a valuable tool because of its coverage, but it has significant drawbacks: it was made with information and technology of the 1960s; since which time there have been significant changes in spatial information technologies such as GPS, remote sensing and geographic information systems (GIS). Another limitation, which is shared with traditional soil survey techniques, is the classification of soils as distinct categories. As noted almost 40 years ago by Webster (1968), this makes substantial assumptions about the conformity of soil variation to categorical classification, that can lead to errors of interpretation. Such errors may not be evident to users of the information (see Chapter 3).

A further problem of soil classifications is that although they capture some of the general characteristics of the soil at some scales, attempts to interpret the soil map in terms of a specific property will tend to fail since soil attributes do not cluster perfectly: a cut on the basis of one attribute may split the variance of another

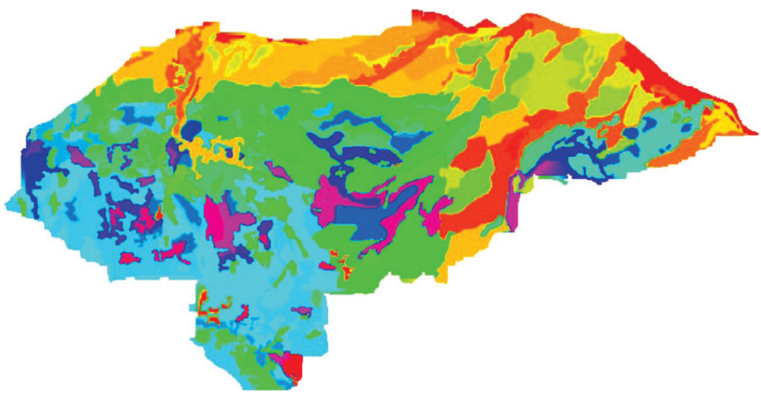


Fig. 33.1 FAO soil map for Honduras (See also Plate 44 in the Colour Plate Section)

attribute near its peak. The failure of traditional soil survey techniques to produce accurate results at smaller scales significantly limits the soil information available to programs that attempt to implement community-based management of resources.

Furthermore, traditional soil maps depend on subjective expert opinion which varies significantly depending on the person creating the maps and the soil classification used (Hudson, 1992). The maps are therefore predominantly qualitative, and depend on poorly specified predictive models – based on tacit knowledge – that are not updatable.

33.1.2 Existing Approaches to Predictive Soil Mapping

There have been a number of approaches to predictive soil mapping, differing both in terms of statistical technique and auxilary data used in the mapping. Many existing approaches to predictive soil mapping use a derivative of Kriging (Krige, 1951; Matheron, 1962). Ordinary Kriging is a form of weighted local spatial interpolation that uses a Gaussian model for the data. Its main drawbacks are the fact that it does not use knowledge of soil materials or processes, and that it requires a large number of closely-spaced samples in order to produce satisfactory results. There are extensions to this method that allow the use of ancillary data, but they are difficult to extend to more than one ancillary variable although methods for this do exist.

Some of the most promising approaches to predictive soil mapping are expert systems and regression trees (Corner et al., 2002). Expert systems use expert knowledge to establish rule-based relationships between environment and soil properties. Often they do not use soil data to determine soil-landscape relationships, but some approaches do. Regression Trees are decision trees with linear models in the leaves. They create a piecewise linear representation of the predicted variable. Using this method Henderson et al. (2005) obtained the very good results, which are able to explain more than 50% of the variance of several soil properties such as pH, clay content and sand content.

33.2 Methods

33.2.1 Input Datasets

The input soil data was collated by CIAT and consists of 2670 soil profiles taken during the 1990s distributed throughout the country. Each soil sample site contains data on texture, pH, organic carbon, organic matter, nitrogen, exchangeable aluminium, and electric conductivity for 4 different horizons, although data was incomplete for a number of sites and for a number of variables. Of the 2670 sites, 2472 had complete information for texture of the topsoil (taken at a depth up to 20–30 cm), and 2451 had data in pH of topsoil. Other variables were missing considerable amounts of data, and hence were not analysed here.

Available for training and prediction were 32 terrain, climate, vegetation and geology related variables. The emphasis was made to select variables potentially available for any tropical country, rather than to rely on data generated specifically for this purpose that. Terrain variables were generated from two different DEMs; SRTM 3 arc second (approx. 90 m resolution) available for the globe from the CSI-CGIAR (Jarvis et al., 2006), and a 50 m Honduras DEM derived from 1:100 000 cartographic sheets hereon referred to as TOPO. Both DEMs were used due to concerns that 90 m spatial resolution was too coarse to capture local soil variation present in the input soil data, although SRTM data presents a significant opportunity for predictive soil mapping given its global coverage. Vegetation data was derived from the SPOT Vegetation products, available globally at 1 km spatial resolution, and the climate variables were generated from the WorldClim climate database (Hijmans et al., 2005), also available for the globe at 1 km spatial resolution. MODIS vegetation data (EVI and NDVI) could considerably improve the vegetation variables, but the data was not available for this study. The geology variable was derived from digitizing a 1:500 000 map sheet, and geological classes were ordered into ages through expert consultation to ensure that the variable was continuous rather than categorical (Gaussian Process models can only use continuous variables). The full list of variables and their respective spatial resolutions is shown in Table 33.1.

33.2.2 Gaussian Processes for Predictive Soil Mapping

We chose the approach of Gaussian Processes, a powerful, non-parametric regression technique with solid probabilistic foundations. The main advantages of Gaussian Processes over other approaches is that they provide well defined confidence intervals, which are very important for soil scientists to assess the quality of the model; and that they allow the use of spatial interpolation and numerous ancillary features to create the model. Kriging can be considered a special case of Gaussian Processes in which only spatial interpolation is used and no ancillary features are included in the model.

Gaussian Processes can be seen as a generalization of Gaussian distributions to function space, which is of infinite dimension. Even though they are not new, they have regained relevance as a replacement for supervised neural networks (Gibbs, 1997; MacKay, 1997). Gaussian Processes are equivalent to several other mathematical approaches including neural networks with infinite number of hidden units, radial basis functions with infinite number of basis functions, least squares support vector machines and kernel ridge regression.

33.2.2.1 Covariance Function

The idea with Gaussian processes is to put a prior in the probability of the interpolating function given the data. Since this prior is Gaussian, a Gaussian Process is defined by its covariance function. The covariance function and its hyperparameters define the family of functions that can be chosen by the Gaussian Process for

Table 33.1 Explanatory variables used in the analysis

Variable type	Variable	Data source	Spatial resolution	Method
Terrain	Elevation, slope	SRTM and TOPO DEM	90 m and 50 m	Landserf
	Topmodel Wetness Index	SRTM and TOPO DEM	90 m and 50 m	Beven and Kirkby (1979)
	Sediment Transport Index	SRTM and TOPO DEM	90 m and 50 m	Moore et al. (1993), produced using ILWIS
	Stream Power Index	SRTM and TOPO DEM	90 m and 50 m	Moore et al. (1993), produced using ILWIS
	Slope Position	SRTM and TOPO DEM	90 m and 50 m	Moore et al. (1993), produced using ILWIS
	Mean curvature (3×3 and 15×15 window)	SRTM and TOPO DEM	90 m and 50 m	Landserf
	Profile curvature (3×3 and 15×15 window)	SRTM and TOPO DEM	90 m and 50 m	Landserf
	Landform features (3×3 and 15×15 window)	SRTM and TOPO DEM	90 m and 50 m	Landserf
Vegetation	Mean NDVI	SPOT Vegetation	1 km	Average
	Inter-annual NDVI variation 1998–2005 (%)	SPOT Vegetation	1 km	Coefficient variability (%)
	Inter-annual NDVI variation 1998–2005 (%)	SPOT Vegetation	1 km	Coefficient variability (%)

Table 33.1 (continued)

Variable type	Variable	Data source	Spatial resolution	Method
Climate	Mean Annual Temperature	WorldClim	1 km	Busby (1991)
	Mean Diurnal Range	WorldClim	1 km	Busby (1991)
	Isothermality (P2/P7)	WorldClim	1 km	Busby (1991)
	Temperature Seasonality	WorldClim	1 km	Coefficient Variability (%)
	Max Temp of Warmest Month	WorldClim	1 km	Busby (1991)
	Min Temp of Warmest Month	WorldClim	1 km	Busby (1991)
	Temperature Annual Range	WorldClim	1 km	Busby (1991)
	Mean temp of wettest quarter	WorldClim	1 km	Busby (1991)
	Mean temp of driest quarter	WorldClim	1 km	Busby (1991)
	Mean temp of warmest quarter	WorldClim	1 km	Busby (1991)
	Mean temp of coldest quarter	WorldClim	1 km	Busby (1991)
	Annual Precipitation	WorldClim	1 km	Coefficient Variability (%)
	Precipitation of Wettest Period	WorldClim	1 km	Busby (1991)
	Precipitation of Driest Period	WorldClim	1 km	Busby (1991)
	Precipitation Seasonality(Coefficient of Variation)	WorldClim	1 km	Coefficient Variability (%)
Geology	Age of parent material		1 km	N/A

interpolating the data. The covariance function selected was the squared covariance with a linear term as shown below:

$$C(\mathbf{x}_i, \mathbf{x}_j) = \theta_1 \exp\left[-\frac{1}{2}\sum_{l=1}^{L}\frac{(x_i^{(l)} - x_j^{(l)})^2}{r_l^2}\right] + \theta_2 + \theta_3\delta_{ij} + \sum_{l=1}^{L}\sigma_w^2 x_i^{(l)} x_j^{(l)}$$

where

L	number of inputs
l	l^{th} input
θ_1	vertical scale
r_l	length scale
θ_2	bias
θ_3	output noise
σ_w	linear term

33.2.2.2 Learning the Hyperparameters and Selecting Variables

The covariance function depends on a set of hyperparameters that need to be determined. The best way to determine the hyperparameters of a Gaussian Process is to learn them from the data by maximizing the likelihood of a prediction given the training data and the parameters. This approach has a regularizing effect as well, therefore reducing the likelihood of having a model that overfits the data.

However, because Gaussian Processes can use a large number of ancillary variables, a regularization step is also required to limit the number of variables used in the model. In order to keep training time low and to further prevent overfitting we use the following variable selection approach: an training set of 20% of available soil samples is used to create an initial model, and its performance is evaluated on 60% of the available samples (validation set). When several variables had similar R^2 values, expert opinion selected the variable considered most important to include in the model. We continued adding variables until the R^2 score of the model stopped improving. Once the most important variables were determined, a new model was trained with the combined 80% of the samples. In order to obtain an independent estimate of the performance of the model, the model is tested against the remaining 20% of the samples (independent test set). With this approach it takes approximately 27 h to select variables and create each model. This process only takes place once, unless new variables become available and they need to be added to the model.

33.2.2.3 Prediction

Once a model is chosen, the next step is to use that model to generate soil maps for an area of interest. In order to do this, features from digital maps of the area are used as the inputs to the model, therefore creating a predicted map for a soil component. We generated maps for pH, sand content, and clay content in the topsoil of Honduras. Even though the prediction stage of Gaussian Processes is much faster than the training stage, prediction is required for *all* points, therefore the process is very computationally intensive. With the current implementation, using a Pentium 4 @1.8 GHz, it takes 21 ms to generate the prediction for one location. The time required to generate a map depends on the size of the map and its resolution. For Honduras (112, 000 km^2), it takes 40 minutes to generate a map with 1 km grid size, 3.4 days with 90 m grid size and 30 days with 30 m grid size. If we were to generate a map of Africa it would take 7.2 days, 2.4 years and 22 years respectively. However, this assumes that all the calculations take place on a single computer, which is not likely to be the case. If multiple computers are available, each one could process a much smaller area therefore reducing the total time required proportionally to the number of computers available.

33.3 Results

33.3.1 Accuracy of Current Techniques

In order to understand the significance of the results achieved, it is important to be aware of the accuracy of current techniques for soil mapping. Measurements of soil characteristics can have a variability of 20% or more between laboratories (Nachtergaele, 1996) and many quantitative prediction methods explain less than 10% of variation. Henderson et al. (2005) explained up to 50% of the variance of pH in soil in Australia and are the motivating force behind the current effort for predictive soil mapping at CIAT.

33.3.2 Topsoil pH

The pH in the topsoil produced the best results the statistical validation. Two different models were created: one that includes the x and y location of the samples as variables (i.e. uses spatial interpolation), and one that does not. The model that uses spatial interpolation performed better, but the one that does not gives better insight into the driving factors for pH determination.

The variables found to be relevant for the model with spatial interpolation were x and y (spatial location of the sample) and P5 (maximum temperature of warmest month). The R^2 for this model is 0.454 (for the test data), that is, the model explains approximately 45% of the variance in the data. From a computer science or engineering perspective, this number seems very low. However, for soil prediction and from a soil science perspective, it is acceptable. The performance of the model for the training set (80%) and the test set (20%) are shown (Fig. 33.2). The resultant map of pH for Honduras with 1 km spatial resolution (Fig. 33.3) demonstrates the heterogeneity of pH across the country, with only some areas of relatively homogenous pH coinciding with specific classes of the FAO soil map (Fig. 33.1). The prediction

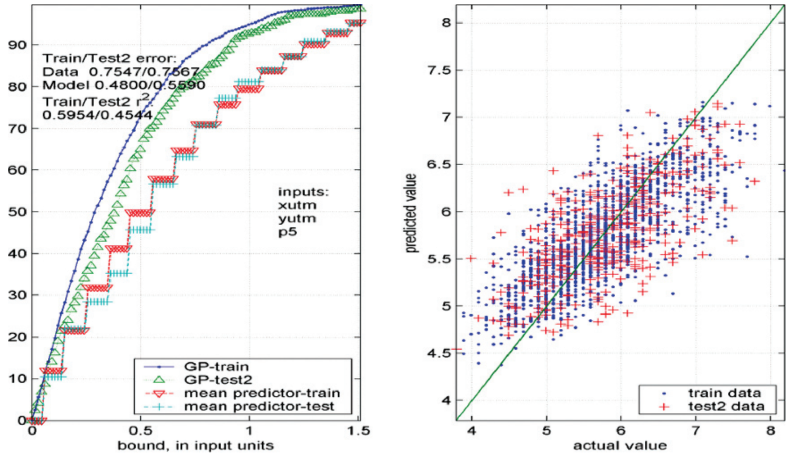


Fig. 33.2 Model performance for pH in topsoil. The figure on the left shows the comparative performance of the model vs. a mean predictor. The *x* coordinate is the bound, in pH units, and the *y* coordinate is the percentage of the predictions that fit within the predicted value +/− the bound. For example, 95% of the predictions will fall within 1 pH unit of the predictions for the training set. This number is slightly lower for the independent test set (92%) and much lower for a mean predictor (80%). The figure on the right shows actual values versus predicted values. In an ideal case, both would be the same (*solid, green line*), but in practice there will always be dispersion around the *y* axis. The more dispersion, the worse the model is (See also Plate 45 in the Color Plate Section)

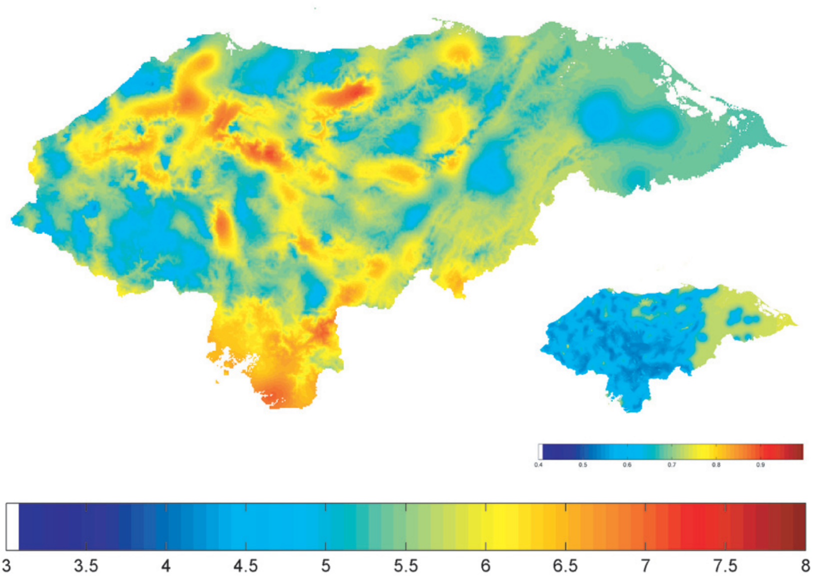


Fig. 33.3 Predicted map of pH in topsoil and 67% confidence interval (See also Plate 46 in the Color Plate Section)

has a 67% confidence interval of about 0.5 pH units, although this is greater in the eastern part of the country where less soil samples were present.

When no spatial interpolation is used the variables selected by the model are P5 (Maximum temperature of warmest month), P2 (Mean diurnal temperature range), P16 (Precipitation of wettest quarter), and geology class of parent material. The R^2 for this model is 0.3652 (for the test data) (Fig. 33.4), which is significantly lower than for the model using spatial interpolation, but can still be considered useful especially under circumstances where irregular or lower densities of soil profile data is available. The resultant map of pH using this model (Fig. 33.5) is similar to the map when spatial interpolation was used, but the 67% confidence interval increases to approximately 0.6. This is still satisfactory given the inherent errors in laboratories, and based on alternative sources of information.

33.3.3 Sand and Clay Content in Topsoil

The models for sand and clay content performed poorly compared to the topsoil pH. While the results using spatial interpolation were acceptable and still comparable to some existing approaches, these results had more limited predictive value. The R^2 for sand was 0.235 (with spatial interpolation) and 0.1032 (without spatial interpolation). For clay, R^2 was 0.167 (with spatial interpolation) and 0.140 (without spatial interpolation).

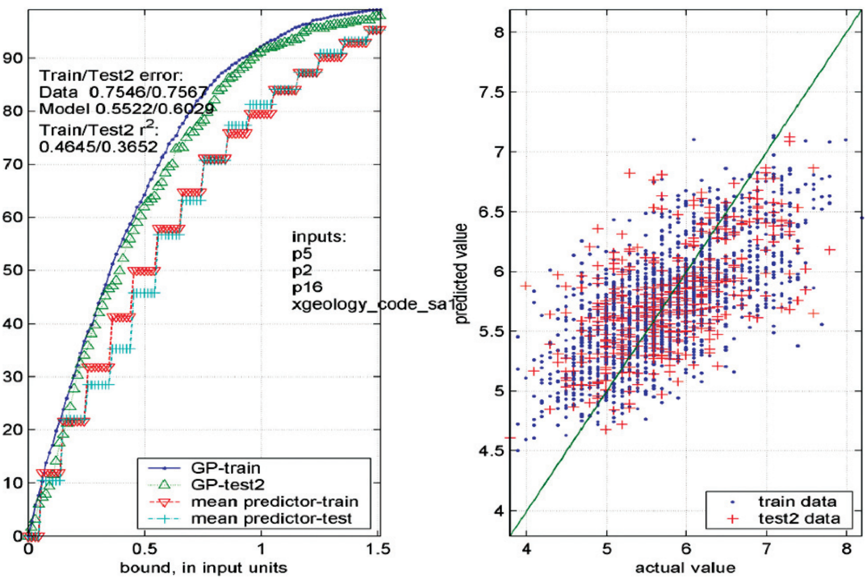


Fig. 33.4 Model performance for pH in topsoil without spatial interpolation (See also Plate 47 in the Color Plate Section)

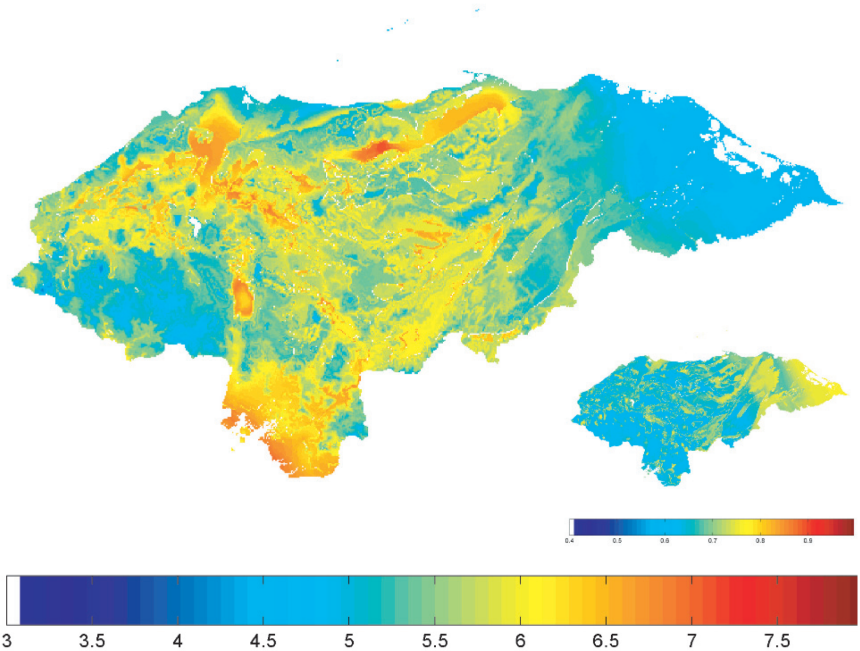


Fig. 33.5 Predicted map of pH in topsoil and 67% confidence interval, without using spatial interpolation (See also Plate 48 in the Color Plate Section)

There are several possible causes for the reduced performance of the sand and clay models. One of the most plausible explanations is that the clay and sand content are not as spatially correlated as pH, therefore requiring higher resolution input variables to accurately predict variation.

33.4 Conclusions and Future Work

Gaussian processes have proven to be a powerful technique for predictive soil mapping, successfully predicting 17–45% of variability in pH, sand and clay content in Honduras. They produce quantitative predictions with solid confidence intervals, combine pedogenic factors with spatial interpolation, allow for complete coverage of an area and enable continued improvement.

The map of pH variation for Honduras (Fig. 33.1) indicates the dominance of climate as a predictive variable at national scale. The weaker influence of terrain and, to a lesser extent geology, seems surprising, given the plentiful evidence of the capacity of terrain variables to predict soil variation (e.g. Gessler et al., 2000). It seems reasonable to explain this as a result of the scale-dependent power of terrain, relative to that of climate (Burrough, 1983). This can be explained as follows: Over a small area that typifies such studies of terrain-influence, climate variation is very small, and unable to influence soil variation. As the area expands, climate influences soil formation more discernibly, and tends to dominate the power of terrain and available geological information. The weaker influence of geology is difficult to explain, and may reflect a confounding effect of map unit interpretation. The surprisingly weak influence of terrain may reflect the inability to provide variables at sufficient resolution (90 m.) to reflect soil formation processes, or more powerful terrain indices are required. Further work should compare directly Gaussian Process models with more established techniques for soil mapping such as kriging or regression trees.

Nevertheless, the digital soil map of pH and clay and sand content created using Gaussian Processes provides a step forward in terms of information resources on soil variation for Honduras. The map is now being used to generate models of species distributions for important crop wild relative species (see Jarvis et al., 2005 for background), and for assessing suitability of crops. It could also be used in a range of applications which require high resolution soil property data, including weather insurance and studies of the impacts of global climate change (see Chapter 3).

Acknowledgments The authors would like to thank Manuela Veloso for her support of the v-unit initiative and her input at different stages in the development of this project.

References

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 43–69.

Burrough, P.A., 1983. Multiscale sources of spatial variation in soil. 2. A non- Brownian fractal model and its application in soil survey. Journal of Soil Science 34, 599–620.

Busby, J.R., 1991. BIOCLIM – A bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds), Cost Effective Biological Surveys and Data Analysis, pp. 64–68. CSIRO, Melbourne.

Corner, R.J., Hickey, R.J., Cook, S.E., 2002. Knowledge Based Soil Attribute Mapping In GIS: The Expector Method. Transactions in GIS 6, 383–402.

Gessler, P.E., Chadwick, O.A., Chamran, L., Althouse, L., Holmes, K., 2000. Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Science Society American Journal 64, 2046–2056.

Gibbs, M.N., 1997. Bayesian Gaussian Processes for Regression and Classification. Ph.D. Thesis, University of Cambridge.

Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P, 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398.

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Hudson, B.D., 1992. The soil survey as paradigm-based science. Soil Science Society of America Journal 56, 836–841.

Jarvis, A., Yeaman, S., Guarino, L., Tohme, J., 2005. The role of geographic analysis in locating, understanding and using plant genetic diversity. In: Zimmer, E.A., Roalson, E.H. (Eds), Methods in Enzymology, Vol. 395, Molecular Evolution: Producing the Biochemical Evidence Part B, Elsevier, New York, USA, 279–298.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2006. Hole-filled SRTM for the globe Version 3, available from the CGIAR-CSI SRTM 90 m Database: http://srtm.csi.cgiar.org.
Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52, 119–139.
MacKay, D.J.C., 1997. Gaussian Processes: A Replacement for Supervised Neural Networks. Lecture notes for a tutorial at NIPS.
Matheron, G., 1962. Traite de geostatistique appliquee. Tome 1, Editions Technip, Paris, 334pp.
Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Science Society of America Journal 57, 443–452.
Nachtergaele, F.O., 1996. From the Soil Map of the World to the Global Soil and Terrain Database, AGLS Working Paper. FAO. Rome.
Sachs, J., 2002. Science, Technology & Poverty: Five Ways to Mobilize Development in Low-income Countries, IAEA Bulletin, Making a Difference, 44, 1.
Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography 27, 171–197.
Selvaradjou, S.K., Montanarella, L., Spaargaren, O., Dent, D., 2005. European Digital Archive of Soil Maps (EuDASM) – Soil Maps of Latin America and Carribean Islands (DVD-Rom version). EUR 21822 EN. Office of the Official Publications of the European Comunities, Luxembourg.
Soil Survey Staff, 1975. Keys to Soil Taxonomy USDA – Soil Conservation Service. 7th ed., Washington D.C.
Webster, R., 1968. Fundamental objections to the 7th Approximation. European Journal of Soil Science 19, 354–366.

Chapter 34
Digital Mapping of Soil Classes in Rio de Janeiro State, Brazil: Data, Modelling and Prediction

M.L. Mendonça-Santos, H.G. Santos, R.O. Dart and J.G. Pares

Abstract A soil database for Rio de Janeiro State was collated in Access, for a project on quantifying the magnitude, spatial distribution and organic carbon in the soils of Rio de Janeiro State (Projeto Carbono_RJ). The main activities were the search, selection, analysis and review of the data for each soil profile already described in the study area, the georeferencing of each soil profile (when spatial co-ordinates were not available) and the input of new soil profiles into a new interface. The Rio de Janeiro soil dataset now contains 731 soil profiles, 2744 soil horizons, and 48 soil attributes usually described at the soil survey process. From this soil dataset, only 431 soil profiles that were adequately geo-located have been used in this application. The dataset contains limited data for bulk density and hydraulic soil properties, among others. From this dataset, quantitative modelling and digital soil mapping have been completed experimentally at 90 m resolution, using soil data and predictor variables, such as satellite images, lithology, a prior soil map and a DEM and its derivates. This dataset, which is one of the more complete soil datasets in Brazil, is being used as a testbed for learning and teaching DSM, using a variety of methods based on the *scorpan* model (McBratney et al., 2003). In the first instance, the soil dataset was used to predict soil classes at the Order level of the Brazilian Soil Classification System – SiBCS (Embrapa, 2006). Five models were built and their results were compared and mapped.

34.1 Introduction

Digital Soil Mapping has been defined by Lagacherie and McBratney (2007) as "the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables". The main

M.L. Mendonça-Santos
Researcher at EMBRAPA –Brazilian Agricultural Research Corporation, The National Centre of Soil Research- CNPS, Rua Jardim Botânico 1024, 22.460-000, Rio de Janeiro, RJ, Brazil
e-mail: loumendonca@cnps.embrapa.br

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

use of the this approach is to replace the polygon-based soil maps of the past with digital maps of soil properties and classes and their associated uncertainties for areas previously mapped, or for new areas. These maps are stored and manipulated in digital form in a GIS environment, creating the possibility of vast arrays of data for analysis and interpretation at any time.

Prediction of soil classes and properties in digital soil mapping relies on finding relationships between soil and the predictor variables of soil-forming factors and processes. The rationale is based on Jenny's equation, which was asserted to list the factors responsible for soil formation, rather than a quantitative formulation,

$$\mathrm{S} = f(c, o, r, p, t, \ldots),$$

where S stands for soil, c represents climate, o organisms including humans, r relief, p parent material and t time.

McBratney et al., (2003) have generalised and formalised a Jenny-like formulation not with the aim of explaining the variables responsible for the soil-forming process but rather for empirical quantitative descriptions of the relationships between soil and the other spatially-referenced factors (or environmental co-variates) which are used here as soil spatial prediction functions. Seven factors are considered:

s: soil, other properties of the soil at a point;
c: climate, climatic properties of the environment at a point;
o: organisms, vegetation or fauna or human activity;
r: topography, landscape attributes;
p: parent material, lithology;
a: age, the time factor;
n: space, spatial position.

Soil can be considered as a factor because soil can be predicted from its properties, or soil properties from its class or other properties. The *scorpan* model can be written as follows:

$$Sc = \mathrm{f}(scorpan) \text{ or } Sa = \mathrm{f}(scorpan)$$

where *Sc* is a set of soil classes and *Sa* is a soil attribute and s refers to soil information either from a prior map, or from remote or proximal sensing or expert knowledge. Implicit in this are the spatial coordinates *x, y* and an approximate or vague time coordinate $\sim t$. This time coordinate can be expressed as "at about some time *t*". Each factor will be represented by a set of one or more continuous or categorical variables, e.g., *r* by the elevation, slope or other DEM attribute. The sources of data, the methods to estimate f (. . .), as well as the steps to perform the *scorpan* framework are presented and discussed in review presented by McBratney et al., (2003).

The Rio de Janeiro dataset described below, that represents one of the more organised Brazilian soil datasets, is being used as a testbed for teaching and learning

DSM techniques. In this chapter, we use the Rio de Janeiro dataset and digital soil mapping procedures to predict soil classes at the Order level of the Brazilian Soil Classification System (Embrapa, 2006), using regression/classification trees as the predictive modeling framework. Five different models were built and compared, in order to access the best approach.

34.2 Material and Methods

34.2.1 The Study Area

The study area is Rio de Janeiro State, located between longitudes 41 °W and 45 °W and latitudes 20 °30′S and 23 °30′S with an area of about 44 000 km^2 (Fig. 34.1). It corresponds to 89 of the 1:50 000 topographic sheets of IBGE (Brazilian Institute of Geography and Statistics). The area is characterised by eight large landscape types (Fig. 34.2), namely, Coastal Plains, North-Northwest Fluminense, Rio Paraíba do Sul Middle Valley, Mountainous Area, Upper Itabapoana River Plateau, Serra dos Orgãos, Bocaina and Mantiqueira (Lumbreras et al., 2003; Rio de Janeiro, 2001), which are described as follows:

34.2.1.1 Coastal Plains

This is the most heterogeneous physiographic domain. It embraces all of the extensive flooded areas such as swamps, lowlands, beach margins, sand dunes and even isolated hills and massive mountainous alignments, up to 1 000 metres in elevation.

In the Coastal Plains a natural vegetation of subdeciduous tropical forest and sandbank vegetation prevails, close to the coast, with an Aw climate type, tropical dry, with 3–6 months of drought. It is an area characterised by high temperatures, moderate rainfall and high evaporation rate where Gleissolos Melânicos or Gleissolos Háplicos and Gleissolos Tiomórficos (Aquent Entisols), Planossolos Hidromórficos (Alfisols), Espodossolos Cárbicos or Ferrocárbicos (Spodsols) are found. In this area, the most intense urban and industrial expansion has occurred. Argissolos Vermelhos and Vermelho-Amarelos (Ultisols), Neossolos Quartzarênicos (Quartzipsaments), Organossolos (Histosols) and Cambissolos Háplicos (Inceptisols), also occur. On the edge of the Coastal Plain is a tabular flat relief surface, where deep well drained soils prevail (Yellow Latossolos and Yellow Argissolos – Oxisols and Ultisols).

34.2.1.2 North – Northwest Fluminense

It is an extensive area dominated by high and low hills, where Ultisols and Alfisols, with moderate or high natural fertility are found. There are also more developed and leached soils (low fertility Red-yellow Latossolos and Red-yellow Argissolos), characterised by the thick C horizons, that extend to great depths. Gleissolos Háplicos

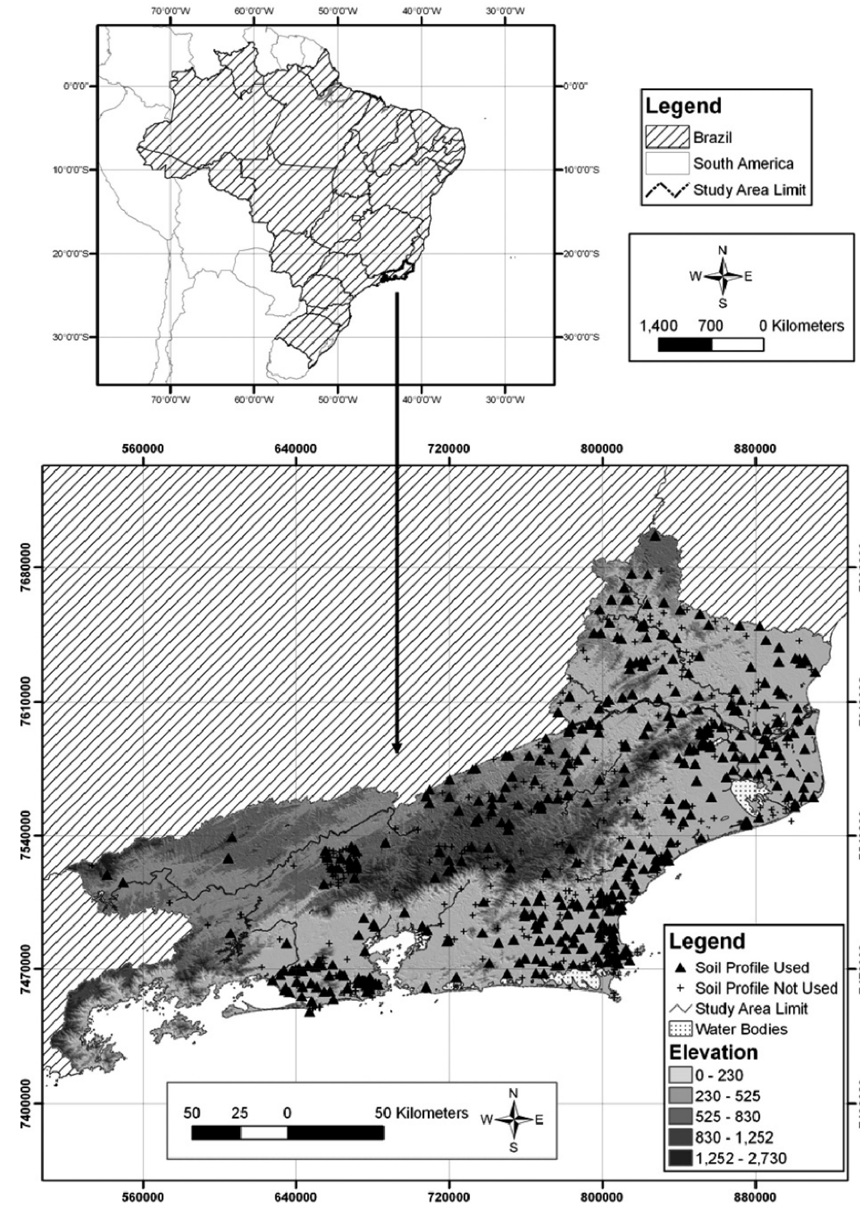


Fig. 34.1 The study area location and the soil profile distribution on the elevation map, extracted from the SRTM DEM (Jarvis et al., 2006) at 90 m pixel resolution

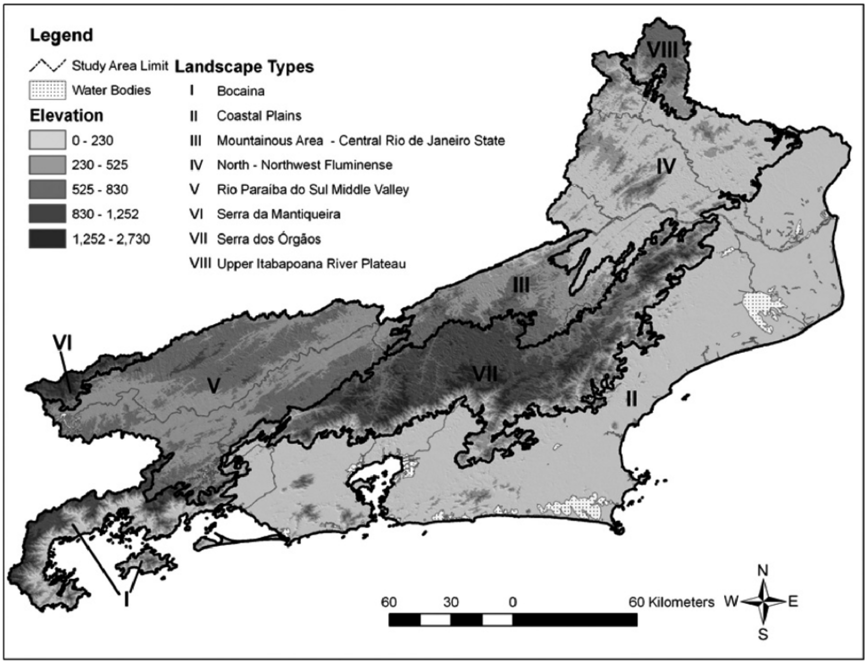


Fig. 34.2 Landscape types of Rio de Janeiro state

and Planossolos Hidromóficos (Aquent Entisols and Alfisols occur in the restricted and discontinuous fluvial plains in the valley bottoms, with eventually Neossolos Flúvicos (Fluvent Entisols), occasionally with presence of toxic levels of salts and/or sodium in subsurface.

34.2.1.3 Rio Paraíba do Sul Middle Valley

This region consists of an extensive depression between the Sea Range and the scarps of the Mantiqueira mountains. Here we find Yellow Latossols derived from Tertiary sedimentary deposits characterized by a flat tabular relief. Between these hills we find the fluvial plains of the Paraíba do Sul river, where Neossolos Flúvicos and Cambissolos Háplicos (Fluvent Entisols and Inceptisols) of high natural fertility occur. The soils close to the Paraíba do Sul river channels are commonly Red-yellow and Red Argissolos, usually not very thick and of good natural fertility (high base saturation). They are quite dissected and eroded, in such a way that in some places the gullies constitute a severe limitation to land use. Low nutrient status Red-yellow Latossolos and Red-yellow Argissolos (Oxisols and Ultisols) prevail. Cambissolos Háplicos and Red-yellow Argissolos (Inceptisols and Ultisols) of high susceptibility to erosion are also common in this area.

34.2.1.4 Mountainous Area – Central Rio de Janeiro State

Most of the Mountainous Area is inserted in the steep relief of Serra dos Órgãos, north of the Sea Range aligned to the coastline of Rio de Janeiro State. In the Mountainous Area the soils are not very thick (Cambisols/Inceptisols) or are very leached (Ferralsols/Oxisols). Ultisols/Acrisols are found on cooler, drier slopes.

In this area, the humid climate and the lower average temperatures in the mountains favours the accumulation of organic matter, creating soils with thick humic A horizons.

Cambisols (Inceptisols) prevail in scarps and steep slopes at higher altitudes, with shallow soils associated with rock outcrops. These soils also occur in areas where more gentle relief is observed, constituting flat terraces (at around 900 m) where deposits of pre-weathered gneiss are thicker, resulting in very deep Oxisols. The humic character on the surface of these soils is related to the mild climate where the decomposition of the organic matter is slower, forming quite thick A horizons.

34.2.1.5 Serra dos Órgãos – Part of the Central Rio de Janeiro Mountainous Area

The Sea Range crosses the whole Rio de Janeiro State in a WSW-ENE direction, accompanying the structural direction of the geological substratum. In its southerly portion, on the boundary with São Paulo State, it comes very close to the sea, constituting a peculiar environment, considered to be a different environmental domain, denominated by the Mountains of Bocaina – South Rio de Janeiro coast, described ahead. In the mountains and mountainous scarps are found Cambissolos Háplicos (Inceptisols) and due to the very high slope gradients and to the widespread occurrence of rock outcrops, Neossolos Litólicos (Lithic Entisols) also occur and, to a less extent, Red-yellow Latossolos (Oxisols), in general not very thick as well as Red and Red-yellow Argissolos Eutróficos (Ultisols and Alfisols).

34.2.1.6 Bocaina – South Rio de Janeiro Mountainous Coast

This domain includes a mountainous group represented by part of the Sea Range, called the mountains of Bocaina, that extends from Itaguaí County, Rio de Janeiro, to the boundary with the State of São Paulo. It comes very close to the sea, delineating a coast line cut out by rocky walls, intermingled with narrow fluvio-marine deposits. In these alluvial plains occur Neosolos Flúvicos and Gleissolos Háplicos (Fluvent and Aquic Entisols), while in the sandbanks Espodossolos Cárbicos or Ferrocárbicos (Spodsols) are found. Shallow soils like Cambissolos Háplicos and Neossolos Litólicos (Inceptisols and Lithic Entisols) and, in positions of less uneven relief, Red-yellow Latossolos (Oxisols). In these areas Cambissolos Húmicos (Humic Inceptisols) are commonly found and more rarely, rather shallow Red-yellow Latossolos.

34.2.1.7 Upper Itabapoana River Plateau

This plateau rises to 700 m and has a more humid and mild climate than the extensive adjacent depression, and a more preserved forest cover. The prevailing hilly relief where Red-Yellow Latossolos occurs is used primarily for pasture and secondarily for growing coffee. Strongly undulating terrain is also common and the dominant soils are Red-Yellow Latossolos and, in a smaller proportion, Red-Yellow Argissolos and Cambissolos (Oxisols, Ultisols and Inceptisols) and rock outcrops.

34.2.1.8 Serra da Mantiqueira

This is a group of mountains with similar characteristics to the coastal Sea Range, separated from it by the large depression of the Paraíba do Sul River Middle Valley, between the states of Minas Gerais and São Paulo. Its most outstanding feature is the alkaline rock topography of Itatiaia, where we find the highest peak in the State, rising to 2 787 m above sea level. The local climate and vegetation cover are typically of higher altitudes, where Neossolos Litólicos and Cambissolos Húmicos (Lithic Entisols and Humic Inceptisols) are mostly found. In surrounding areas, in relatively lower relief positions and uneven topography, we find Cambissolos Háplicos (Inceptisols), rock outcrops and relatively thin Red-Yellow Latossolos. In the lower slopes, Red-Yellow Latossolos (Oxisols) and Red-Yellow Argissolos (Ultisols) are found.

34.2.2 Soil and Ancillary Datasets

A new soil database was built in Access and Delphi, in order to facilitate the search, selection, analysis and review of the data for each soil profile already described in the State. This also allowed the georeferencing of soil profiles (when spatial coordinates were not available) and the addition of new data from the soil profiles and other measurements.

From the original soil dataset (731 soil profiles), only 431 soil profiles could be used in this application (Fig. 34.1), since the others could not be adequately geo-located. As illustrated in Fig. 34.1, the existing soil profiles are unequally distributed in the study area, with only a few in the southern part of the state. Soil attributes (both morphological and chemical), vegetation and landform information were recorded by different researchers who have performed soil surveys in the area over time. The previous work was synthesized and updated by Carvalho Filho et al., (2003). Additional soil profiles were described and sampled during the so-called "Projeto Carbono_RJ" (Mendonça-Santos et al., 2005) and added to the database.

The Brazilian Soil Classification System – SiBCS (Embrapa Solos, 2006) was used to allocate the soil profiles at Order level. The 431 soil profiles used in this application pertain to 9 from the 13 soil Orders of the SiBCS (Fig. 34.3). Detailed information on the soil sampling procedure and physico-chemical methods of analyses are given in Embrapa Solos (1997).

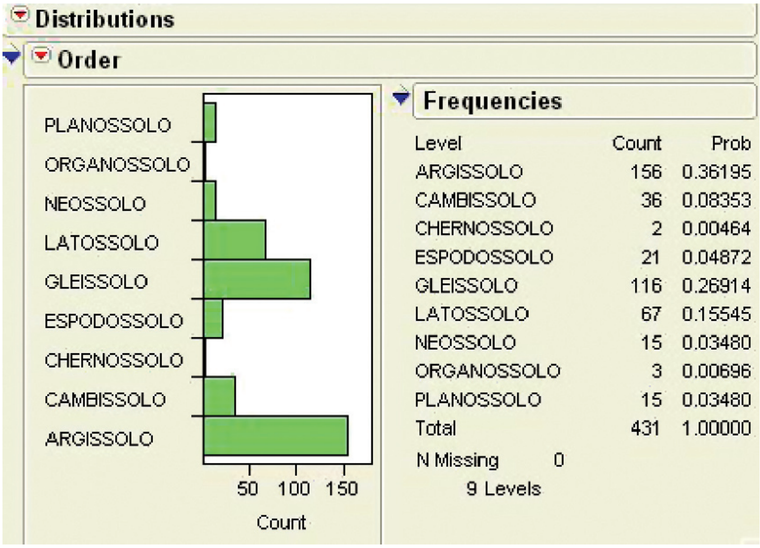


Fig. 34.3 Soil Order and frequency of the 431 soil profiles used in the modeling process

In this application the following covariates were used as predictor variables to build the spatial soil class prediction models: Geocover^{TM} mosaic (bands 7, 4 and 2 in RGB), freely distributed by NASA (https://zulu.ssc.nasa.gov/mrsid/); the NDVI index (using band 2 instead of 3); Land Use/Land Cover (LULC) map of Rio de Janeiro State, derived from Landsat ETM+ (Mendonça-Santos et al., 2003) which use in digital soil mapping is further discussed in Chapter 16; the Lithology class map (Rio de Janeiro, 2001) and SRTM DEM 90 m, obtained from the CGIAR database at http://srtm.csi.cgiar.org (Jarvis et al., 2006).

The SRTM DEM 90 m had no data values in the lowest zones as shown in Fig. 34.4. In order to fill those gaps, the empty DEM zones were interpolated using Vesper software (Minasny et al., 2002). The new DEM 90 m was then used as input to the LandMapR software (MacMillan, 2003), to obtain the DEM derivates to be used in predictive models, which suitability is further discussed in Section 10.4.

The soil dataset was complemented with the covariates of environmental factors for each soil data point, extracted using ERDAS Imagine software (Leica Geosystems, 2003). An ancillary dataset representing the whole study area was interpolated onto a 90-m grid corresponding to the SRTM DEM, and populated with environmental and soil variables. Exploratory statistical analysis was performed on soil data in JMP software (SAS Institute, 2005). The modelling and prediction of soil classes was done by a regression/classification tree, using See5 software (RuleQuest Research, 2003). The output results were imported and mapped in ArcGIS (ESRI, 2004).

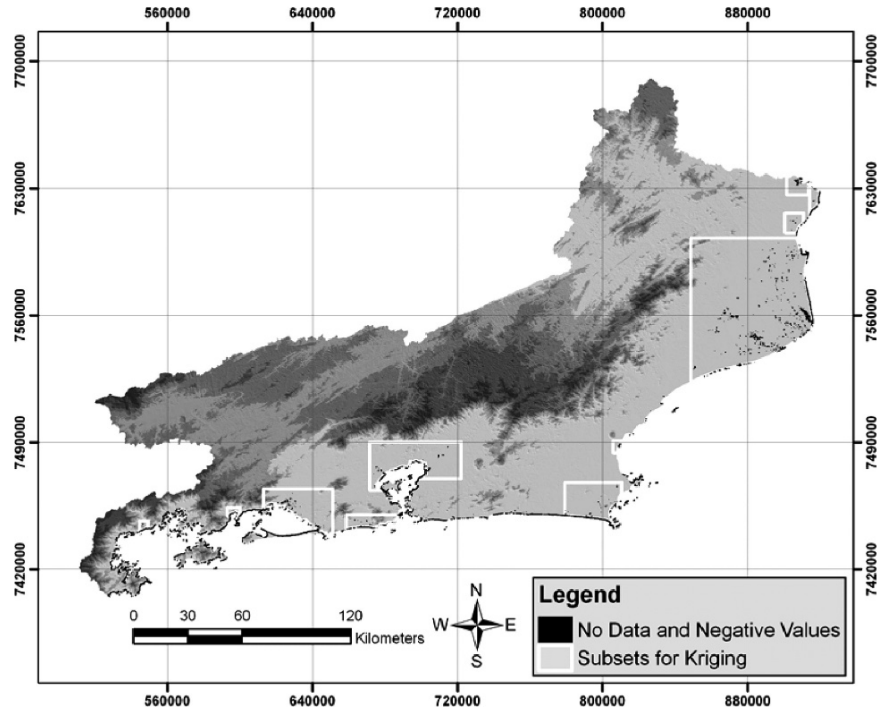


Fig. 34.4 SRTM DEM 90 m showing the no data and negative value areas, with the subsets where kriging was performed

34.2.3 Modelling and Prediction of Soil Classes

To predict soil classes a classification/decision tree algorithm (program See 5.0, Rulequest Research, 2003) was used. The data were partitioned into prediction ($^3/_4$ of the data) and validation ($^1/_4$ of the data) sets. The program provides an error assessment for both subsets.

A tree structure was generated by partitioning the data recursively into a number of groups, each division being chosen to maximise some measure of difference in the response variable in the resulting groups, as discussed in Section 10.5.1. Table 34.1 illustrates the main data and information used to build every soil class prediction model, varying the predictor variables (M1, M2, M3, M4 and M5) (at Order level). Information was obtained in almost all factors of *scorpan* model as shown in Table 34.1, including a preexisting polygon based soil class map (Carvalho Filho et al., 2003) and a Lithology class map (Rio de Janeiro, 2001).

In order to compare the performance of the models accounting for the number of parameters used, Akaike's Information Criterion (AIC) (Akaike, 1973) was used as a quality index:

$$\text{AIC} = -2\,loglike + 2\,m$$

Table 34.1 Predictor variables, predictive model, predicted variable and soil dataset information for each model built (M1, M2, M3, M4 and M5)

Models	Predicted Model	Preditor variables for the *scorpan* model	Predicted variables	Soil profiles	Pred_dataset (3/4)	Valid_dataset (1/4)
M1	Classification Tree	***s*** (Soil classes polygon map); ***o*** (Landsat ETM + 7 Images, bands 7, 4, 2 and NDVI); ***r*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE); ***p*** (Lithology classes polygon map)	Soil Class at Order level, according to Embrapa (2006)	431	324	107
M2		***o*** (Landsat ETM + 7 Images, bands 7, 4, 2, NDVI and LULC map); ***r*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE), ***p*** (Lithology classes polygon map)		431	324	107
M3		***o*** (Landsat ETM + 7 Images, bands 7, 4, 2 and NDVI); ***r*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE), ***p*** (Lithology classes polygon map)		431	324	107
M4		***o*** (Landsat ETM + 7 Images, bands 7, 4, 2 and NDVI), ***r*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE)		431	324	107
M5		***r*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE)		431	324	107

where *loglike* is the log-likelihood of the prediction, and m is the number of parameters used in the model. This index is a compromise between the goodness of fit and the parsimony of the model. The best model is the one that has the smallest AIC.

The log-likelihood for class prediction ($k = 1, \ldots, K$) is calculated as follows (Hastie et al., 2001):

$$loglike = \sum_{k=1}^{K} \log \hat{p}_k$$

where $\hat{p}_k$ is the proportion correctly classified as class k:

$$\hat{p}_k = \frac{1}{N_k} \sum I(y_i = k),$$

and $N =$ total number of data (soil profiles)

34.3 Results and Discussion

The main difference between the models is the set of predictor variables used to build them. In model 5 (**M5**) only data on the ***r*** factor of *scorpan* model were used to build the model (DEM derivates: ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE). In **M4**, satellite images and NDVI index (Landsat ETM+7 Images, bands 7, 4, 3 and NDVI), were used in addition to DEM derivates (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE), as data sources for the ***o*** and ***r*** factors of the *scorpan* model. Model 3 (**M3**) was built using data source for ***o*** (Landsat ETM + 7 Images, bands 7, 4, 3 and NDVI); ***r*** (ELEV, ASPECT, PLAN, PROF, QWETI, SLOPE) and ***p*** (Lithology classes polygon map) factors of the *scorpan* model. Model 2 (**M2**) is similar to M3, but has one more source of information for the ***o*** factor: the LULC class map. **M1** differs from M3 in only one important aspect: it includes soil information from a prior polygon-based soil map.

Results for the prediction of SiBCS – Soil Order and the performance measure (AIC) of each model are given in Table 34.2.

In a general way, all models were able to give a good prediction of soil classes at the highest categorical level (Order) of SiBCS, showing reasonable error values, what means that soil classes in the study area could be predicted from environmental covariates, that can be easily, cheaply or even freely acquired, like Geocover™ images and the SRTM DEM 90 m (models M4 and M5).

As can be observed, M5 (with only DEM derivates) had the worse performance (highest AIC) and was able to predict only 7 from the original 9 soil classes represented in the soil dataset (Fig. 34.3). A surprising performance was shown by M4, with the bands and NDVI from the Geocover™ mosaic, in addition to the DEM derivates, giving the second smallest AIC.

Table 34.2 Results of the prediction of soil classes (at Order level of SiBCS), using a classification tree

Models	Predicted soil classes	m (number of parameters used)	Prediction dataset error (%)	Validation dataset error (%)	AIC
M1	(**9 classes**): ARGISSOLO, CAMBISSOLO, CHERNOSSOLO, ESPODOSSOLO, GLEISSOLO, LATOSSOLO, NEOSSOLO, ORGANOSSOLO, PLANOSSOLO	32	2.4	3.6	**1616.3**
M2	(**7 classes**): ARGISSOLO, CAMBISSOLO, ESPODOSSOLO, GLEISSOLO, LATOSSOLO, NEOSSOLO, PLANOSSOLO	30	17.1	45.9	2774.0
M3	(**8 classes**): ARGISSOLO, CAMBISSOLO, ESPODOSSOLO, GLEISSOLO, LATOSSOLO, NEOSSOLO, ORGANOSSOLO, PLANOSSOLO	21	11.7	44.1	2681.7
M4	(**8 classes**): ARGISSOLO, CAMBISSOLO, ESPODOSSOLO, GLEISSOLO, LATOSSOLO, NEOSSOLO, ORGANOSSOLO, PLANOSSOLO	10	12.3	45.0	2669.0
M5	(**7 classes**): ARGISSOLO, CAMBISSOLO, ESPODOSSOLO, GLEISSOLO, LATOSSOLO, NEOSSOLO, PLANOSSOLO	6	21	45.0	3249.7

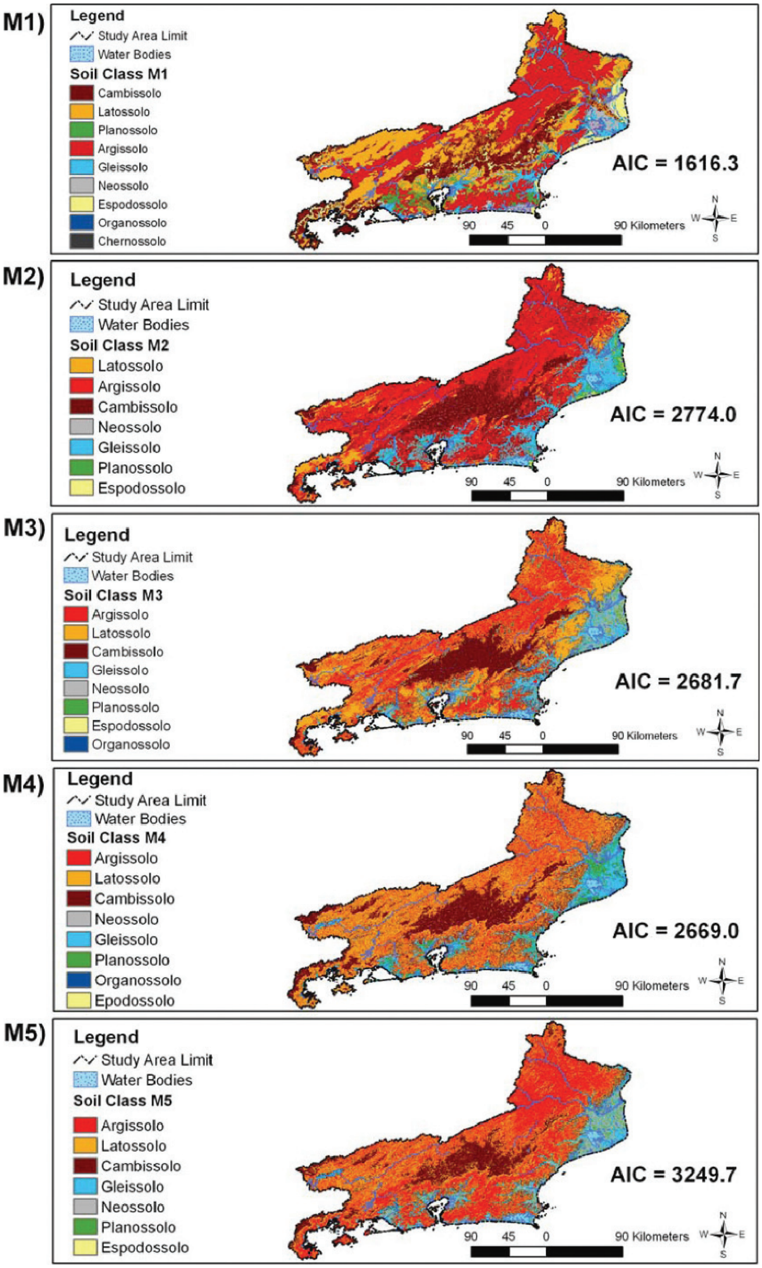


Fig. 34.5 Resulting maps of predicted soil classes from the five models (See also Plate 50 in the Colour Plate Section)

The best model was M1, which was able to predict all the 9 soil classes included in the model, with a very good performance (the smallest AIC) as well as their spatial distribution all over the study area (soil classes appear where expected), in a fairly generalised soil map (Oder level of SiBCS). In fact, M1 was an exceptional model due to the use of the existing soil information (polygon soil map) to estimate soil class. The use of the *s* factor of the *scorpan* model, was first proposed by McBratney et al., (2003), and which is perfectly logical, but has not been previously tested as a source to predict soil classes.

Figure 34.5 illustrates the resulting maps for each model. Model M2 has performed a poor error, predicting the "Gleissolo" class for almost all landscape type VII (Serra dos Órgãos) with the highest elevation (Fig. 34.2). In fact, this soil type could occur in this landscape type, but only in very small areas restricted to the valley bottoms, close to the drainage network, water channels and rivers. Another misclassification was that in M2 M3 and M5 have not or poorly predicted the "Espodossolos" class, in comparison with M1 (the best model) or M4 (the second best model).

Fair consistency is observed between the landscape types and soil Orders throughout the map, mainly as provided in M1 (Fig. 34.6), in accordance with the expert knowledge acquired from field work.

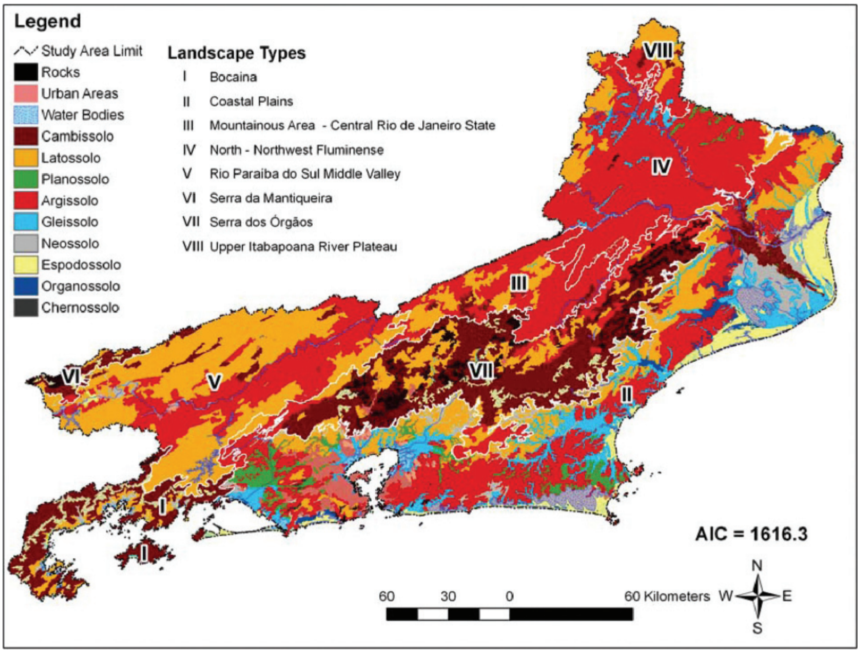


Fig. 34.6 Digital soil class map (at order level of SiBCS) as predicted by the best model (M1) with landscape types (See also Plate 49 in the Colour Plate Section)

34.4 Conclusions and Remarks

Five different models (M1...M5) were built and tested using DSM procedures, in order to predict soil classes. Even though the available soil profiles did not have an optimal distribution in the study area, it was possible to predict soil classes and their spatial distribution through the study area at higher categorical level using 431 soil profiles and soil and landscape covariates. This work demonstrates that the soil-landscape relationship can be predicted in a quantitative and efficient way, using available information on data sources and existing methodologies in the *s.c.o.r.p.a.n.* procedure. This approach is very promising as a procedure to predict soil (classes and/or properties) in areas with lack of soil information, as is the case in much of Brazil. This kind of information can be directly used, support other environmental and planning studies or can help on deciding the needs for new sampling efforts for more detailed studies.

Acknowledgments This work was conducted under the scope of the Research Projects E-26/171.360/2001 "Modelagem da magnitude e distribuicão espacial do carbono orgânico nos solos do Estado do Rio de Janeiro, usando técnicas quantitativas, SIG e Base de Dados" and E-170.023/2001 "Aplicacão de técnicas quantitativas digitais para otimizar o mapeamento de solos para fins de planejamento e gestão ambiental", both funded by FAPERJ – Carlos Chagas Filho Foundation for Research Support of Rio de Janeiro State, that we wish to acknowledge for supporting our research on DSM since the beginning. We also thankfully acknowledge Embrapa, FAPERJ, CNPq (The National Council for Scientific and Technological Development) and CPRM (Brazilian Geological Service) for providing infrastructure and funding for the organisation of the 2nd Global Workshop on Digital Soil Mapping in Rio de Janeiro, which contributed to join people, ideas and experience on this exciting subject.

Great thanks to Alex. McBratney and Budiman Minasny for their support of our DSM efforts and to Bob MacMillan for permission to use his LandMapR® Software. Special thanks to J.F. Lumbreras for the soil polygon map and help with correcting of soil profile locations and descriptions.

References

Akaike, H. 1973. Information theory and an extension of maximum likelihood principle. In: B.N. Petrov and F. Csaki (Eds.), Second International Symposium on Information Theory. Akademia, Kiado, Budapest, pp. 267–281.

Embrapa Solos. Centro Nacional de Pesquisa de Solos. 2006. Sistema Brasileiro de Classificação de Solos. (SiBCS) 2nd ed. 306p.

Carvalho Filho, A., Lumbreras, J.F., Wittern, K.P., Lemos, A.L., Santos, R.D. dos, Calderano Filho, B., Calderano, S.B., Oliveira, R.P., Aglio, M.L.D., Souza, J.S. De, Chaffin, C.E. 2003. Mapa de reconhecimento de baixa intensidade dos solos do estado do Rio de Janeiro. Rio de Janeiro: Embrapa Solos, 2003. 1 mapa, color. Escala 1:250.000. Available at http://www.cnps.embrapa.br/solosbr/sigweb.html.

EMBRAPA. Centro Nacional de Pesquisa de Solos. 1997. Manual de métodos de análises de solo. 2nd ed. Rio de Janeiro, 212 p.

ESRI. 2004. Geoprocessing in ArcGIS, version 9.1. Redlands, CA, Environmental Systems Research Institute, Inc.

Hastie, T., Tibshirani, R., Friedman, J. 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer. 533p.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E. 2006. Hole-filled SRTM for the globe. Version 3. Available from the CGIAR-CSI SRTM 90 m Database: http://srtm.csi.cgiar.org.

Lagacherie,P., McBratney, A.B. 2007. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In; Lagacherie, P., McBratney, A.B. and Voltz, M. (Eds.), Digital soil mapping: an introductory perspective. Developments in soil science vol. 31. Elsevier, Amsterdam, pp. 3–22, 600p.

Leica Geosystems. 2003. ERDAS Tour Guide, version 8.5.GIS & Mapping, LLC, Atlanta, GA. 638p.

Lumbreras, J.F., Naime, U.J., Carvalho Filho, A., Wittern, K.P., Shinzato, E., Dantas, M.E., Palmieri, F., Fidalgo, E.C.C., Calderano, S.B., Medina, A.I.M., Pimentel, J., Chagas, C.S. 2003. Zoneamento Agroecológico do Estado do Rio de Janeiro. Embrapa Solos. Boletim de Pesquisa e Desenvolvimento n.33. Emdrapa Solos, Rio de Janeiro. 113p.

MacMillan, R.A. 2003. LandMapR□ Software Toolkit- C + + Version 2003. User Manual. LandMapper Environmental Solutions Inc., Edmonton, AB. 110pp.

McBratney, A.B., Mendonça-Santos, M.L., Minasny, B. 2003. On digital soil mapping. Geoderma 117, 3–52.

Mendonça-Santos, M.L., Jacques, P.D., Coelho, M.R., Pimentel, J., Santos, H.G. dos, Almeida, P., Barbosa, E.L.de A., Costa, T.C. e C. Da, Ávila, R.M., Shinzato, E., Branco, P.C.M.P. de A. 2003. Mapeamento do Uso Atual e Cobertura Vegetal dos Solos do Estado do Rio de Janeiro. Boletim de Pesquisa e Desenvolvimento n. 22. Embrapa Solos, Rio de Janeiro. 44p.

Mendonça-Santos, M.L., Santos, H.G. dos, Coelho, M.R., Pares, J.G. 2005. Caracterização de paisagens e solos representativos do Estado do Rio de Janeiro para fins de estimativa de estoques de carbono no solo. Boletim de Pesquisa e Desenvolvimento n. 66. Embrapa Solos, Rio de Janeiro. 80p.

Minasny, B., McBratney, A.B., Whelan, B.M., 2002. VESPER version 1.6. Australian Centre for Precision Agriculture. McMillan Building A05, The University of Sydney, NSW.

RuleQuest Research. 2003. See5/C5.0 version 1.20. RuleQuest Research Pty Ltd., Sydney, Australia. http://www.rulequest.com/see5-info.html.

Rio de Janeiro. 2001. Geologia, geomorfologia, geoquímica, geofísica, recursos minerais, economia mineral, hidrogeologia, estudos de chuvas intensas, aptidão agrícola, uso e cobertura do solo, inventário de escorregamentos, diagnóstico geoambiental. Companhia de Pesquisa de Recursos Minerais – CPRM/EMBRAPA SOLOS/DRM-RJ. 1CD-rom.

SAS Institute. 2005. JMP Start Statistics, version 6. Cary, North Carolina.

Part V
Priorities in Digital Soil Mapping

Chapter 35
Synthesis and Priorities for Future Work in Digital Soil Mapping

F. Carré and J.L. Boettinger

35.1 Introduction

This chapter is a brief synthesis of the contents of the previous chapters. We aim to analyze the important points presented and provide a perspective of digital soil mapping research and development efforts needed in the future.

Three important steps in the digital soil mapping process are explored in this book: (1) the input data for soil and other environmental covariates, (2) the inference systems for the prediction of the distribution of soil types and classes, and (3) protocols and their demand for the delivery and exchange of digital soil mapping outputs. Figure 35.1 represents the different issues presented in the book.

35.2 The Input Data on Soil and Other Environmental Covariates

We can distinguish two different strategies of digital soil mapping with sparse soil data infrastructures: (i) with legacy soil data (existing soil maps and soil observations), (ii) in the absence of legacy soil data. Two different approaches have been developed:

1. When legacy soil data in the form of soil maps are available, the quality of legacy data must be evaluated in order to assess the accuracy of the digital soil mapping output. Although this work is a logical first step in digital soil mapping, in this book there is one case (Chapter 25) where different legacy soil data already exist. Once the quality of the data is evaluated, it can be corrected through the use of ancillary soil data, usually derived from Digital Elevation Models (DEM). In Chapter 6 it is called ‘the rescuing of legacy data’.

F. Carré
European Commission, Directorate General Joint Research Centre, Land Management & Natural Hazards Unit, TP 280, 21020 Ispra (VA), Italy
e-mail: Florence.Carre@jrc.it

A.E. Hartemink et al. (eds.), *Digital Soil Mapping with Limited Data*,

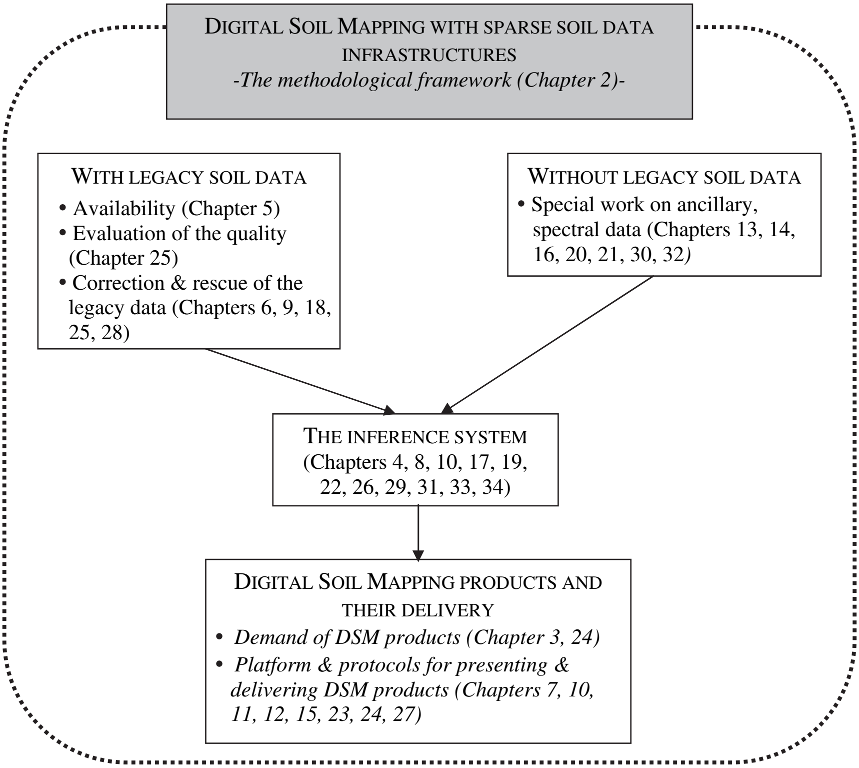


Fig. 35.1 Scheme for an operational framework of DSM

Chapter 5 related the quantity of legacy soil data (generally traditional soil maps) with socio-economic indicators like the Gross Domestic Product (GDP), reflecting also the history of a country. Although the relations are not direct, it is clear that the national coverage of detailed soil maps increases with increasing GDP.

Legacy soil data in the form of soil observations can also be useful for digital soil mapping but only if the point data are properly georeferenced (e.g., Chapter 25). Other soil observations can be useful for inferring soil properties from classes and for soil assessment (e.g., Chapters 23, 27), and may be considered further in the delivery of digital soil mapping products in the section below.

Legacy soil data represent crucial information for digital soil mapping. From these soil maps, soil scientists can learn about the pedological context as well as some insight in the spatial variability. Since legacy data are also the history of soil mapping, several conclusions can be drawn on their relevance and their usage. This final point is emphasized in the demand and delivery of digital soil mapping outputs section below.

2. When no legacy data exist, the work is focused on utilizing reliable and affordable technologies and methodologies for deriving 'scorpan' factors or soil

covariates. An innovative technique was presented in Chapter 13 on Diffuse Reflectance Spectroscopy (DRS), which allows for measuring soil properties, particularly the concentrations of organic carbon, carbonates, clay, sand, water, and metals in specific environments. The focus is on calibrating the relationships between real soil property measurements and the visible-near infrared and mid-infrared reflectances, with the ultimate goal of building a spectral library of soils.

Satellite-derived spectral data such as MODIS, Landsat, and ASTER are being used for modeling environmental covariates. For example, Chapter 30 presents a sophisticated methodology for deriving soilscape patterns through object-oriented landscape and landuse classification of satellite images, post-classification processing and fusion with other environmental factors. The inclusion of the soil covariate of landuse/field pattern is quite innovative in the digital soil mapping methodology, and has a strong impact on soil functioning. Moreover, the analysis of satellite images through time reduces the bias caused by one-time measurement of transient properties (e.g., soil water content, crop cover).

When no legacy data exist – particularly no georeferenced measurements of soil properties– it is important to optimize the sampling of future measurements. There are several approaches existing in the literature and Chapter 20 proposes a fuzzification of the environmental covariates to stratify the area in few classes where field sampling can be focused to establish the relationships between soil and environmental conditions.

35.3 The Inference System for Digital Soil Mapping with Limited Soil Data Infrastructure

The soil inference systems presented in the different chapters largely depend on the existence and quantity of soil measurements, as described in Chapter 2. If no legacy data exist, a top-down approach is necessary for having a first recourse on the sampling from the covariates. If legacy data exist, a bottom-up approach has to be used for understanding and modeling the pedological context of the area, and to map soil variability (for more information on these two approaches see Chapter 17).

With the increase in the soil measurements database, the tendency for non-spatial statistical models will be to use combined or ensemble models, which come from data-mining techniques. For instance, Chapter 33 describes the usage of Gaussian models for mapping topsoil pH, clay and sand content in Honduras. This technique allows for stratifying the dataset and for finding the most suitable models for predicting a variable to apply on each subset. The overall accuracy of the prediction is then increased with a first stratification of the dataset.

The major techniques found in the different chapters dealing with soil inference systems are explained in McBratney et al. (2003).

35.4 The Demand and Proposed Protocol for Digital Soil Mapping Outputs Delivery and Exchange

The analysis of the demand for digital soil mapping outputs (see Chapters 3 and 24) is highly relevant since it allows for:

1. adapting digital soil mapping outputs to specific user needs,
2. putting constraints on protocols and data input before deriving the inference systems,
3. emphasizing communication and collaborations with other scientists, more oriented to resource management for making digital soil assessment,
4. building legislation and international agreements for a better soil conservation policy (through a better communication on soil).

The primary need of soil data concerns the spatial distribution of soil properties, particularly at medium to high resolution since the users need data for land management (see Chapter 10). As stated in Chapter 24, there is now less emphasis on soil classification maps although such information remains of interest for soil scientists to understand pedogenesis. Since soil properties maps are linked to actual soil measurements, field and laboratory measurements should be standardized to allow for the assessment of the accuracy of the related soil map with an independent dataset. The standardization includes sampling protocol, measurements and accuracy assessments. Moreover, in order to facilitate data exchange between soil mapping institutes and soil map users, protocols should be set up concerning web-communication. This involves work on platforms of data representation and downloads and structure of metadata with explicit quality analysis/quality control methodologies. A clear and interesting example for Australia is presented in Chapter 24 and is a good starting point to increase the potential investments for future similar work.

35.5 Lessons Learnt and Future Priorities

The chapters dealing with limited spatial soil data infrastructure make a bridge between the concept of digital soil mapping (which had been presented in the first digital soil mapping workshop in Montpellier, France) and the operation of the digital soil mapping process (e.g., Chapter 34). They show that digital soil mapping is not yet operational from an application point of view. This can be explained by the fact that the world can be stratified into two parts leading to different methodologies: a part where a lot of legacy soil data (particularly traditional soil maps and georeferenced point data) exist, and a part where there are no legacy soil data. The digital soil mapping techniques which are involved cannot be the same, leading to different outputs. The different points to focus on for short-term research on digital soil mapping are the following:

1. Evaluating and using legacy data in digital soil mapping.
2. Exploring new sampling schemes and environmental covariates in digital soil mapping.
3. Using integrated sensors or other new technologies for inferring soil properties or status.
4. Innovative inference systems (new methodologies for predicting soil classes and properties, and estimating uncertainties).
5. Using digital soil mapping products and their uncertainties for soil assessment and environmental applications.
6. Protocol and capacity building for making digital soil mapping operational.

These different points will be the issues addressed at the next digital soil mapping workshop which will be held in Utah in October 2008.

Reference

McBratney, A., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Colour Plate Section

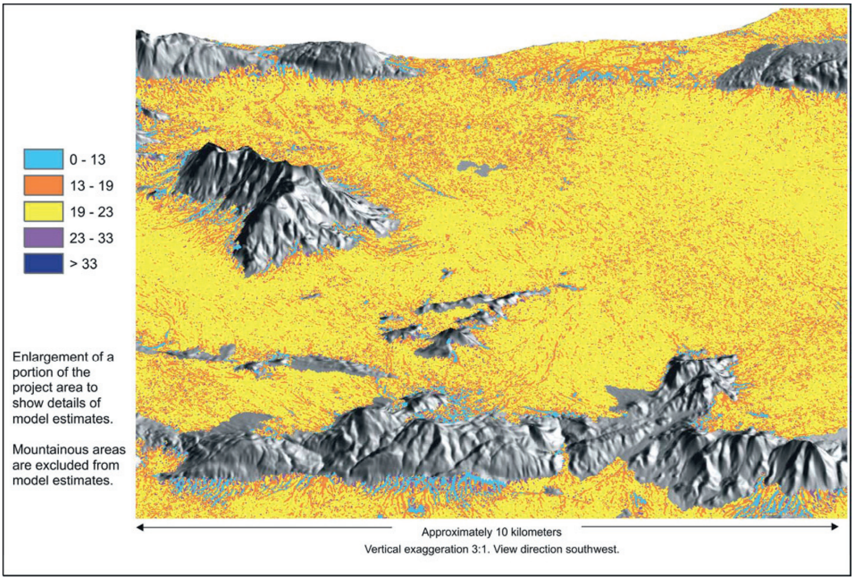


Plate 1 Estimated depth (cm) to top of zone of accumulation of secondary carbonates (See also Figure 4.1 on page 48)

(a)

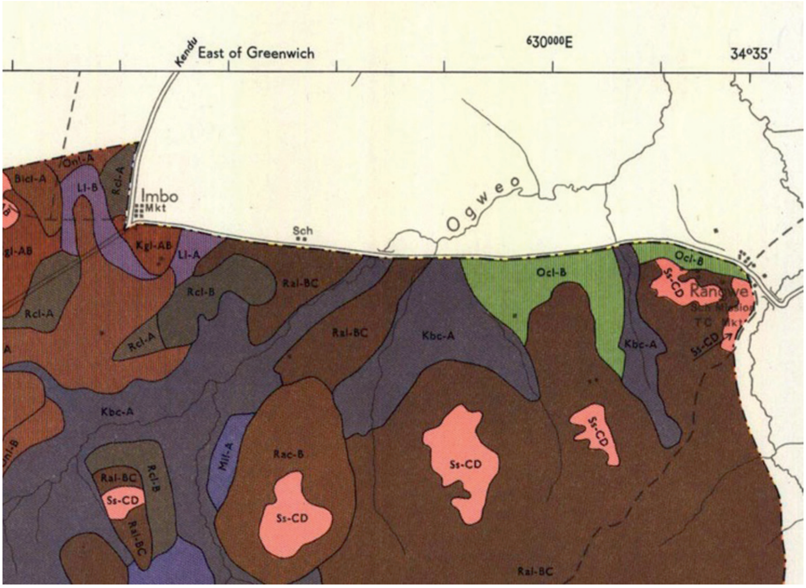


Plate 2 Portion of East Konyango (Kenya) soil map (See also Figure 6.1A on page 76)

(b)

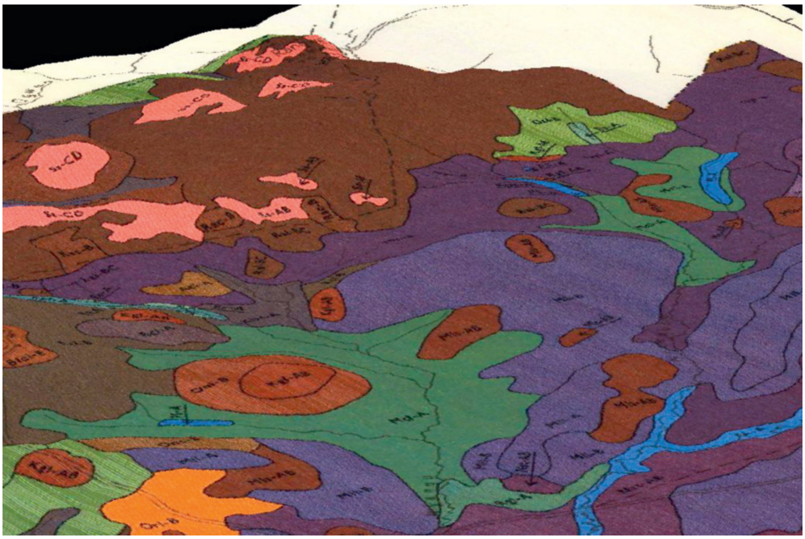


Plate 3 Perspective view, soil map draped on SRTM elevation model (See also Figure 6.1B on page 76)

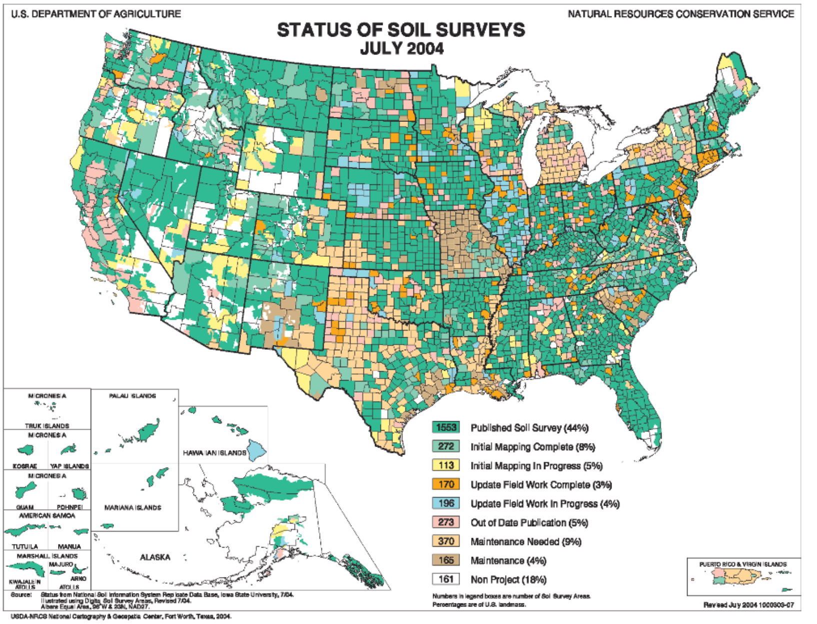


Plate 4 Status of soil surveys in the United States. (approximately 1:2,000,000) (See also Figure 7.1 on page 84)

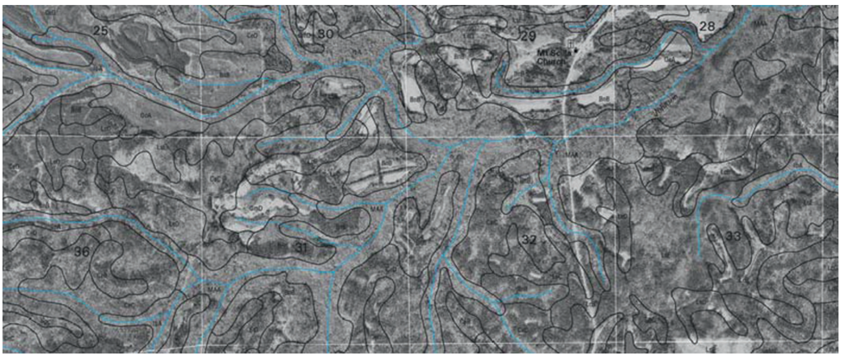


Plate 5 Individual soil mapping units as portrayed in the Soil Survey of Barbour County, Alabama (Trayvick, 1995) (See also Figure 7.2 on page 85)

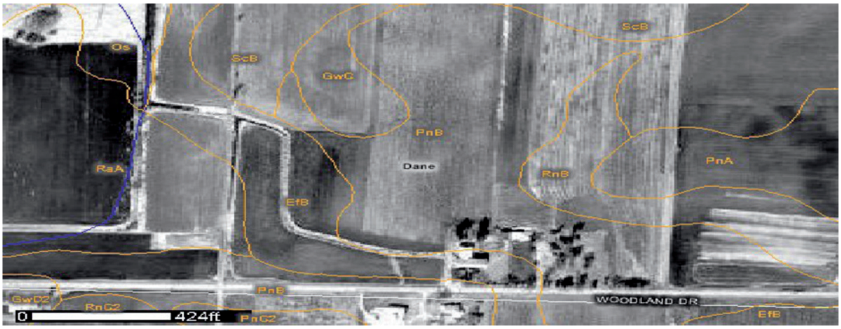


Plate 6 Portion of the Dane County, Wisconsin soil survey (Glocker and Patzer, 1978) produced from the Web Soil Survey (See also Figure 7.3 on page 85)

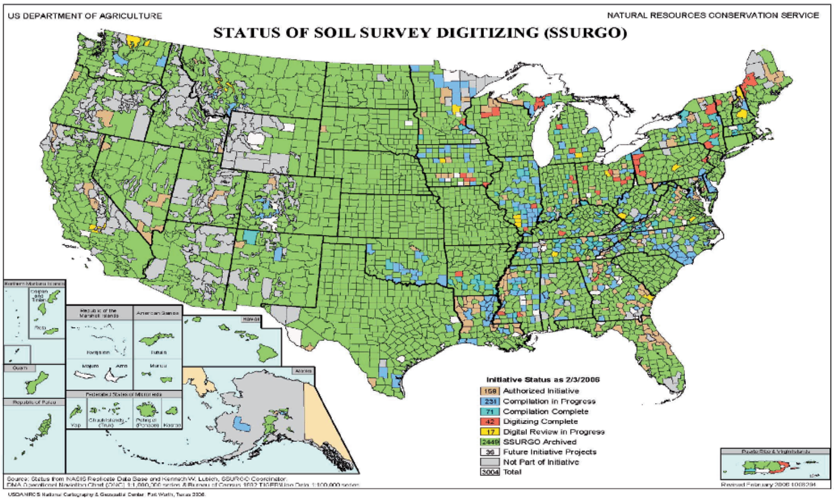


Plate 7 Status of SSURGO digitizing in the United States. (approximately 1:2,000,000) (See also Figure 7.4 on page 86)

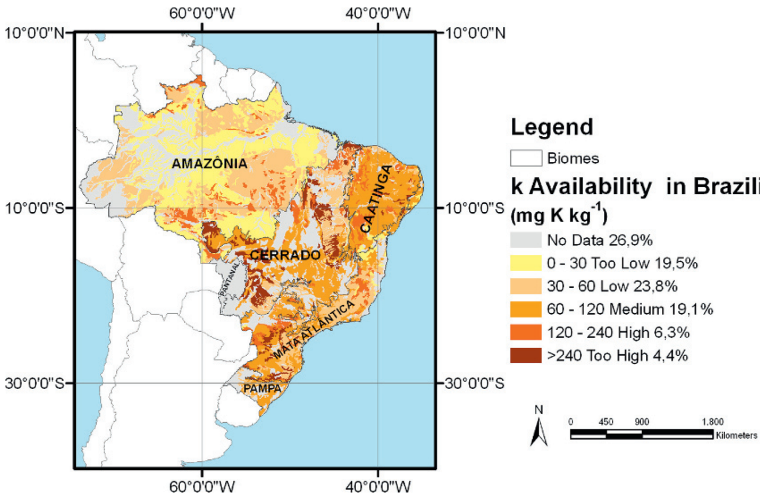


Plate 8 K availability map from calculated K soil profiles with associated spatial information (See also Figure 8.6 on page 100)

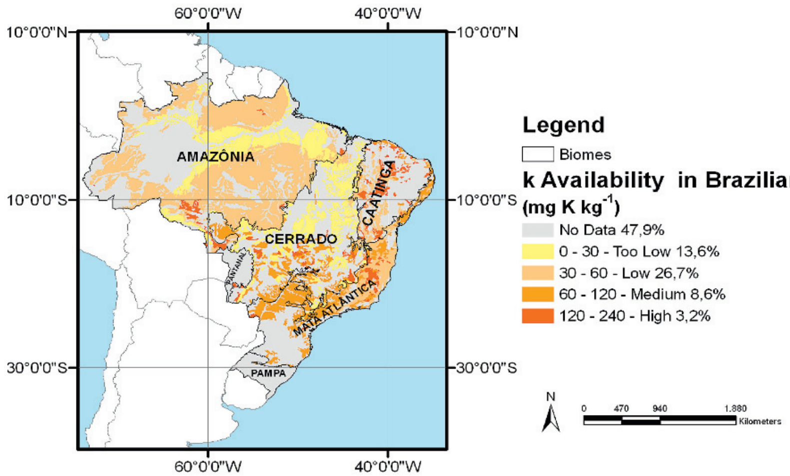


Plate 9 K availability from descriptive statistics application to calculated K soil profiles in a third-order classification to Acrisols, Luvisols e Ferralsols (See also Figure 8.7 on page 100)

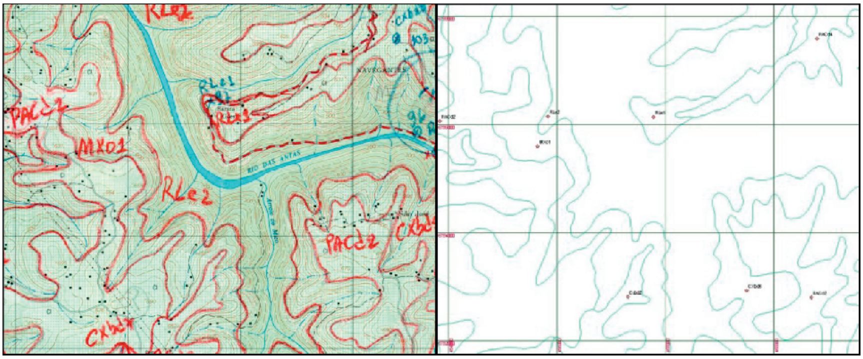


Plate 10 The topographic map sheet resulting from the field work and the mapping units after digitized (See also Figure 9.3 on page 108)

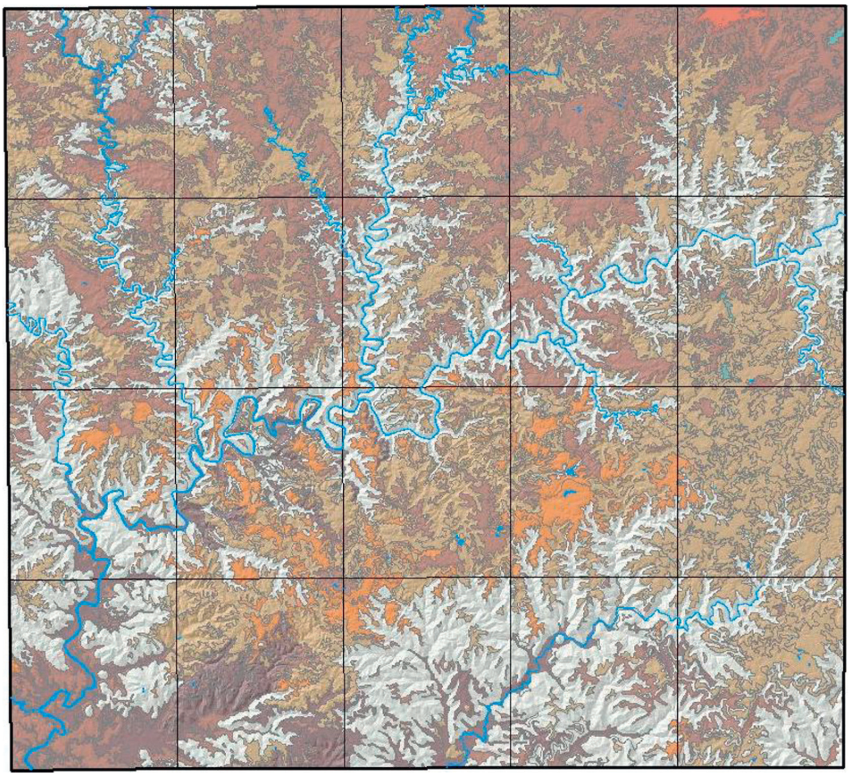


Plate 11 Continuous georeferenced digitized soil map of the Serra Gaúcha region (See also Figure 9.4 on page 109)

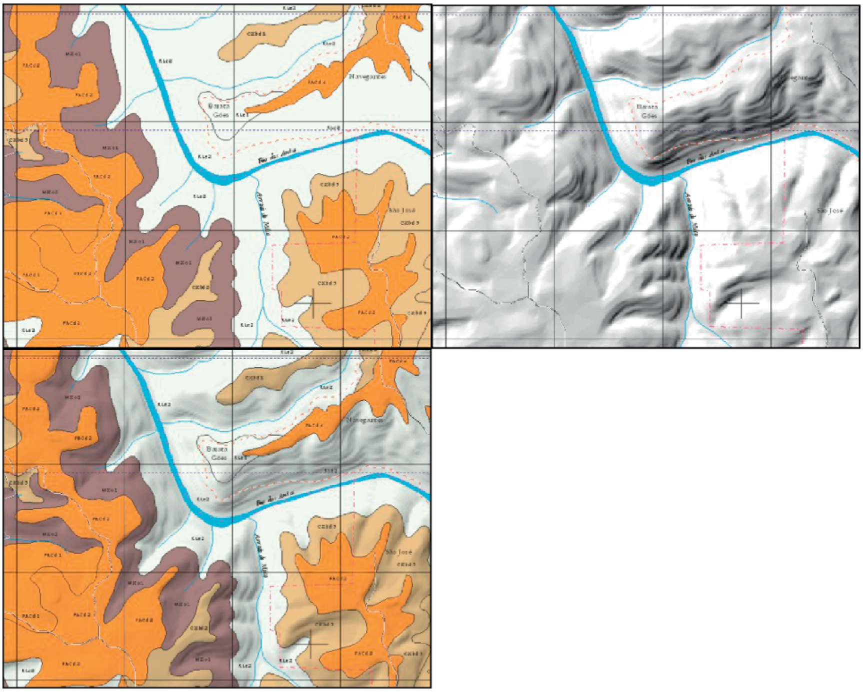
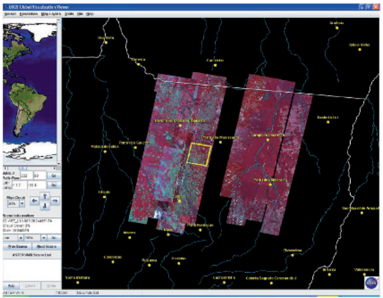
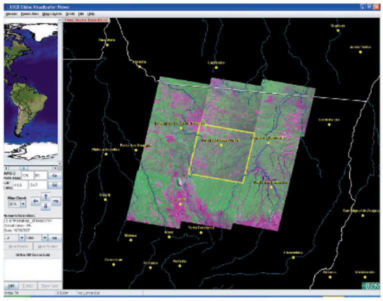


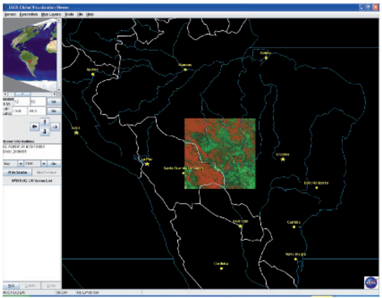
Plate 12 Soil map with conventional soil information, analytical DEM hill shading and fusion of the hill shading with the conventional soil map (See also Figure 9.5 on page 110)



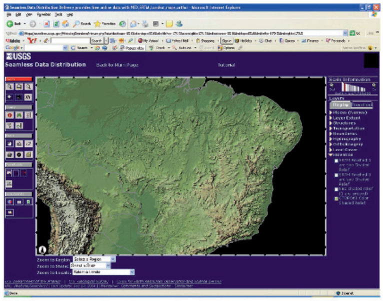
a) Free ASTER Image data for an area in Brazil



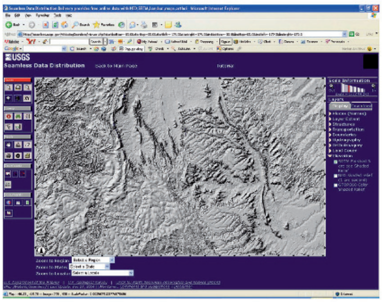
b) Free LandSat Image data for an area in Brazil



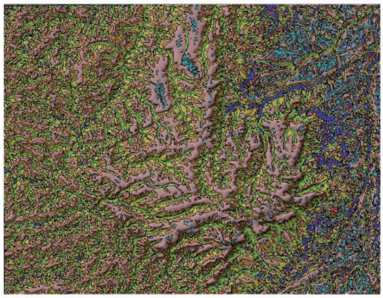
c) Free MODIS Image data for an area in Brazil



d) Free SRTM DEM data for all of Brazil



e) Free SRTM DEM data for an area of Brazil



f) A Simple landform classification of SRTM data

Plate 13 Illustration of access to and use of free spatial data that is widely available for most areas, even those considered to have sparse spatial data availability (See also Figure 10.2 on page 119)

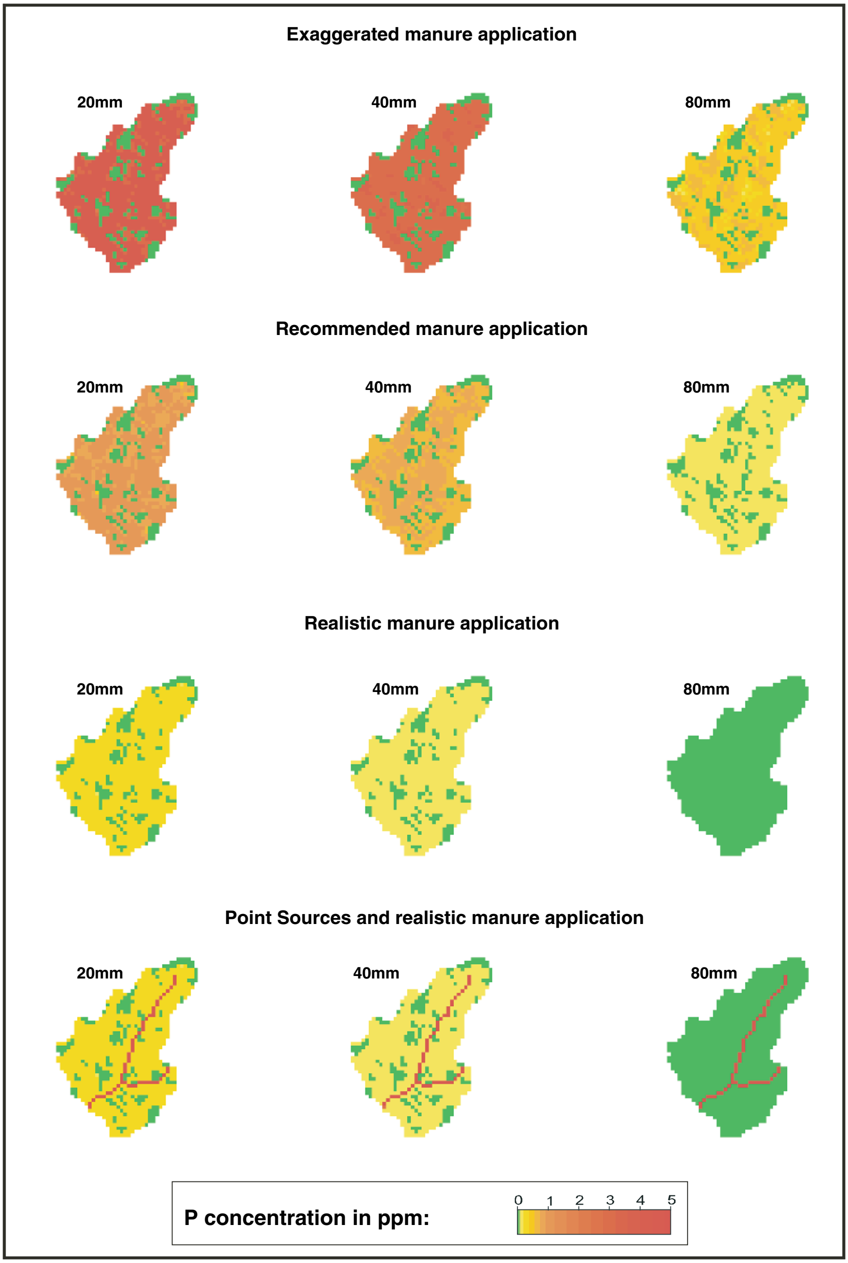


Plate 14 P concentration (ppm) in runoff for four scenarios and three different storm sizes (See also Figure 12.2 on page 158)

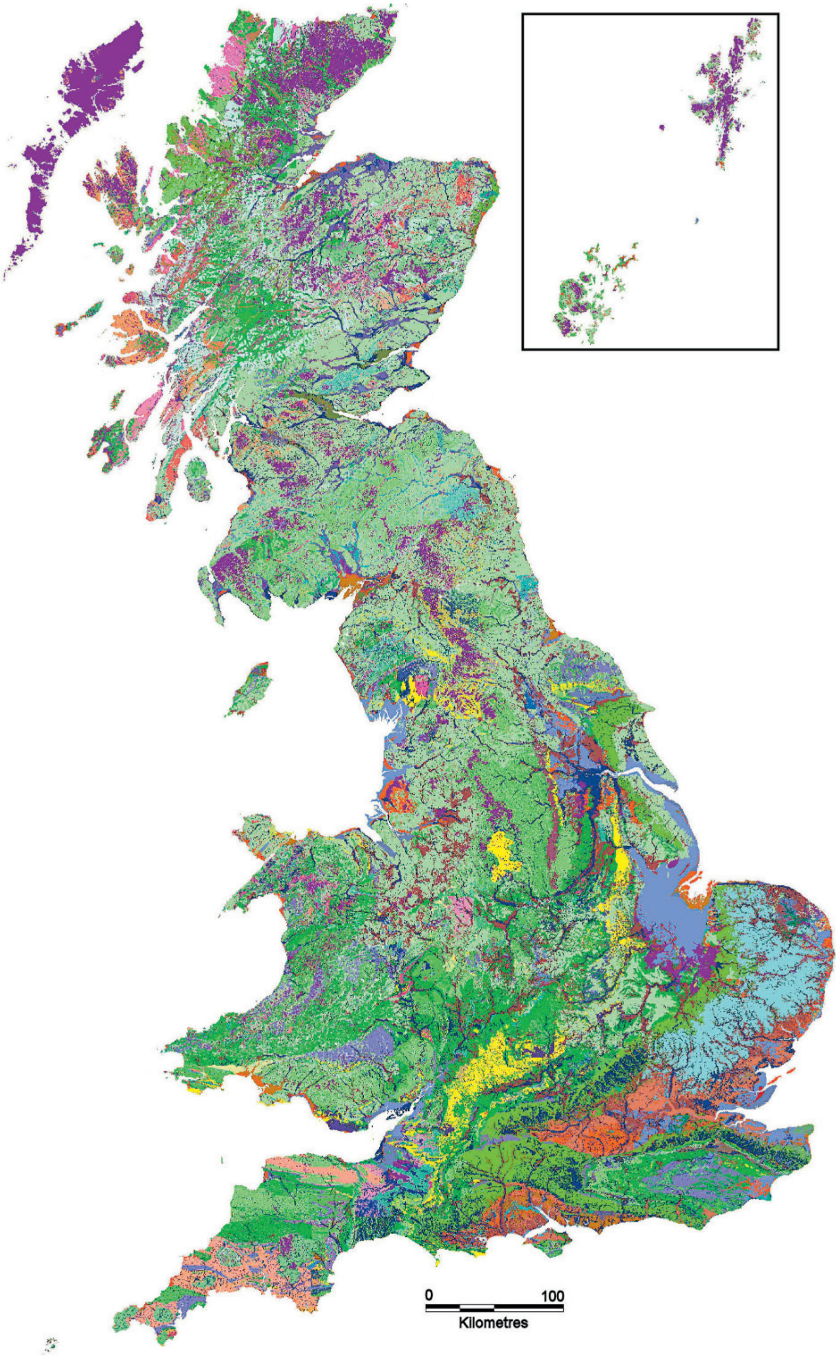


Plate 15 Parent Material Map for Britain (version 0.1) (See also Figure 14.2 on page 176)

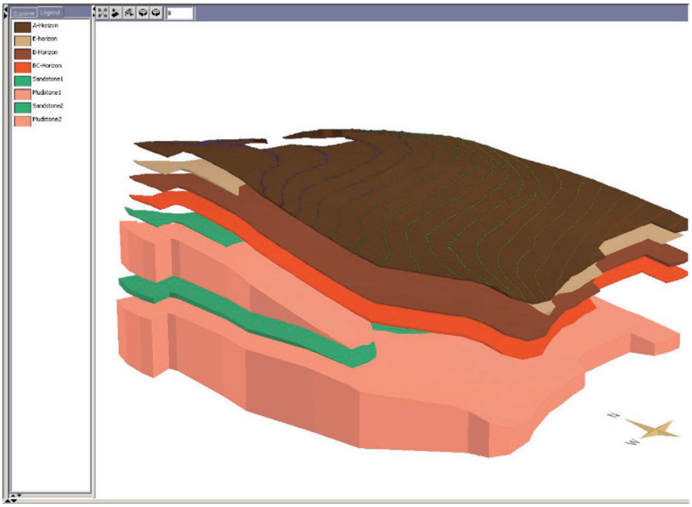


Plate 16 Final 3D soil-geology model for the SW of the Brakenhurst site (500 × 500 m, 8x vertical exaggeration). The model shows the topsoil A-Horizon in dark brown, partially overlying the E-Horizon of a surface water gley in beige. The brown horizon is an amalgamation of all subsoil B-Horizons, which are underlain be the red B/C horizon which constitutes mainly weathered and soliflucted bedrock. Bedrock sandstone is shown in green, and mudstone in pink (See also Figure 15.4 on page 189)

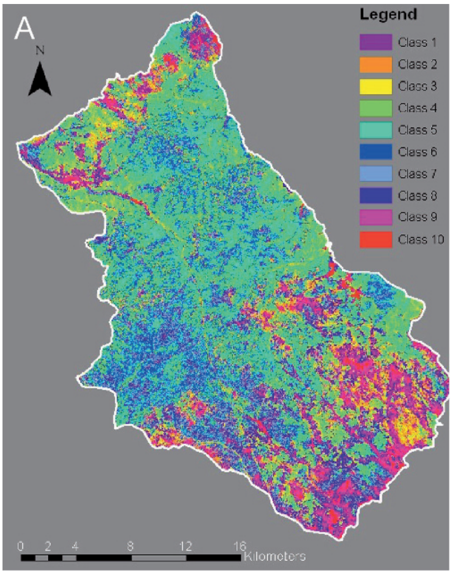


Plate 17 (**A**) Map illustrating a 10-class unsupervised classification of a raster data layer stack containing the soil enhancement ratios of Landsat spectral band ratios 3/2, 3/7, and 5/7; slope; compound topographic index; fractional vegetation cover derived from the NDVI in the Green River Basin of Wyoming, USA.

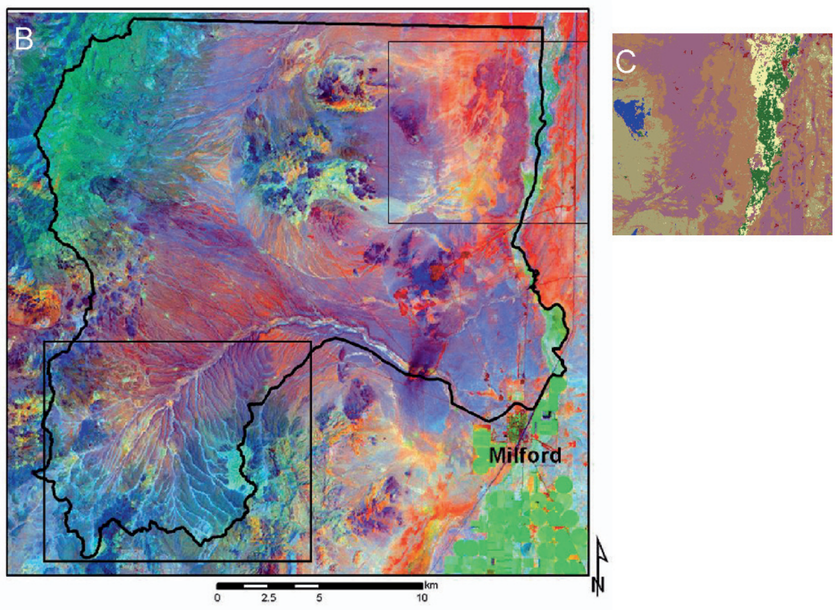


Plate 17 (Continued) (**B**) Image of the first three components of the principal components analysis (PCA) of the Landsat 7 ETM+ image of a Basin and Range landscape in southwestern Utah. The box at lower left indicates the approximate location of the area shown in Fig. 16.2. The PCA image distinguishes areas of different parent materials (e.g., sedimentary vs. igneous rocks as indicated in Fig. 16.2), and different vegetation density (e.g., high vegetation density is represented by *green* areas at tops of mountains in *upper left* and in irrigated fields in *lower right*, in contrast to the *red* to *purple* areas on the arid alluvial fans with low vegetation density). (**C**) Map illustrating a supervised classification of the PCA, focusing on the area indicated by the *upper right* box in B. Training areas were selected in the field for the supervised classification. Each color relates to a predicted soil map unit relating to soil class (e.g., loamy-skeletal Typic Haplocalcids) and dominant vegetation (e.g., *Artemisia tridentata* ssp. *wyomingensis* community) (See also Figure 16.3 on page 199)

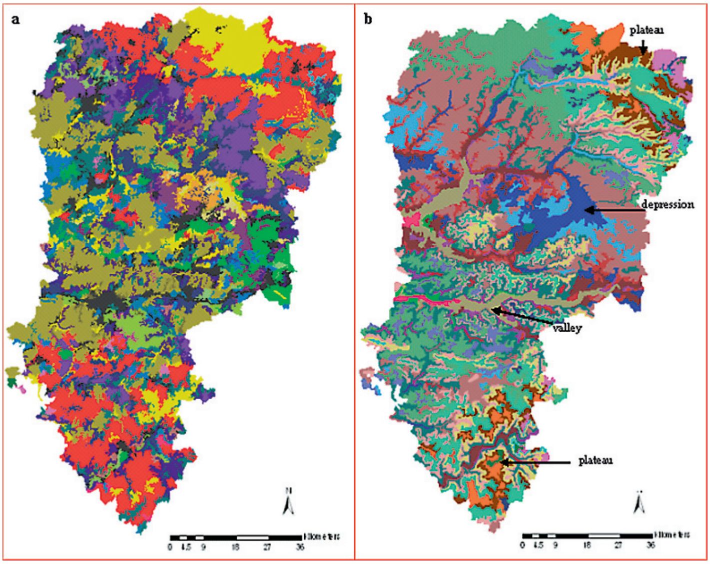


Plate 18 Maps resulting from the bottom-up (**a**) and the top-down (**b**) approaches (See also Figure 17.2 on page 208)

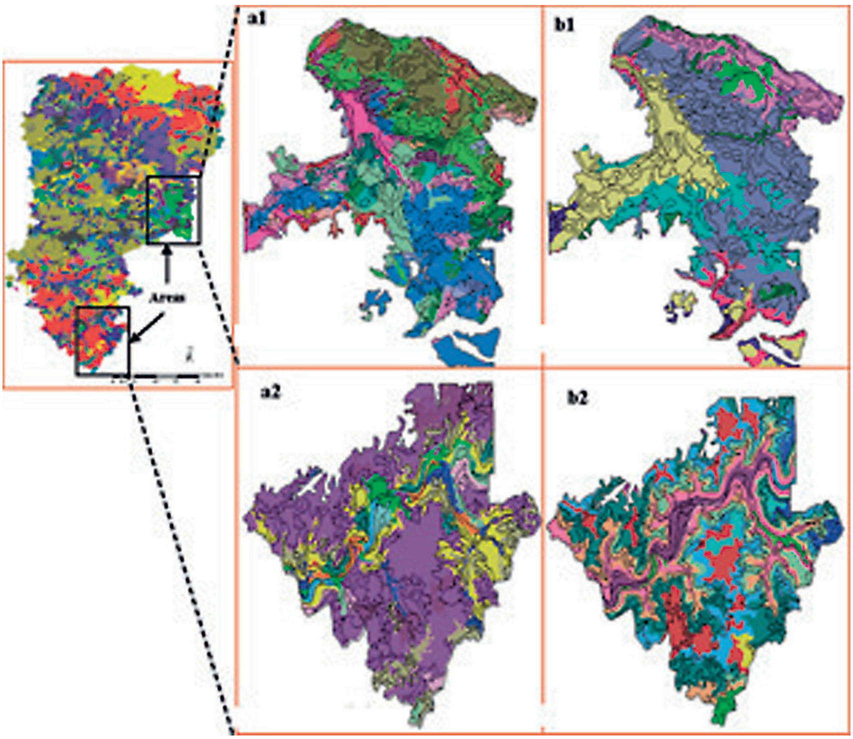


Plate 19 Comparison between contours of the manually drafted 1:250,000 map (in *black*) to the results of the bottom-up (**a**) and top-down (**b**) approaches for the Eastern (1) and Southern (2) parts (See also Figure 17.3 on page 209)

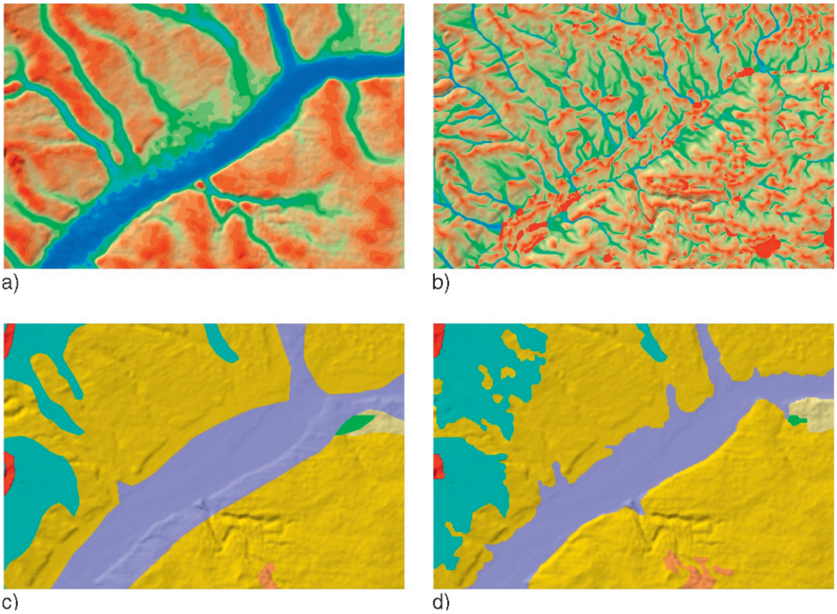
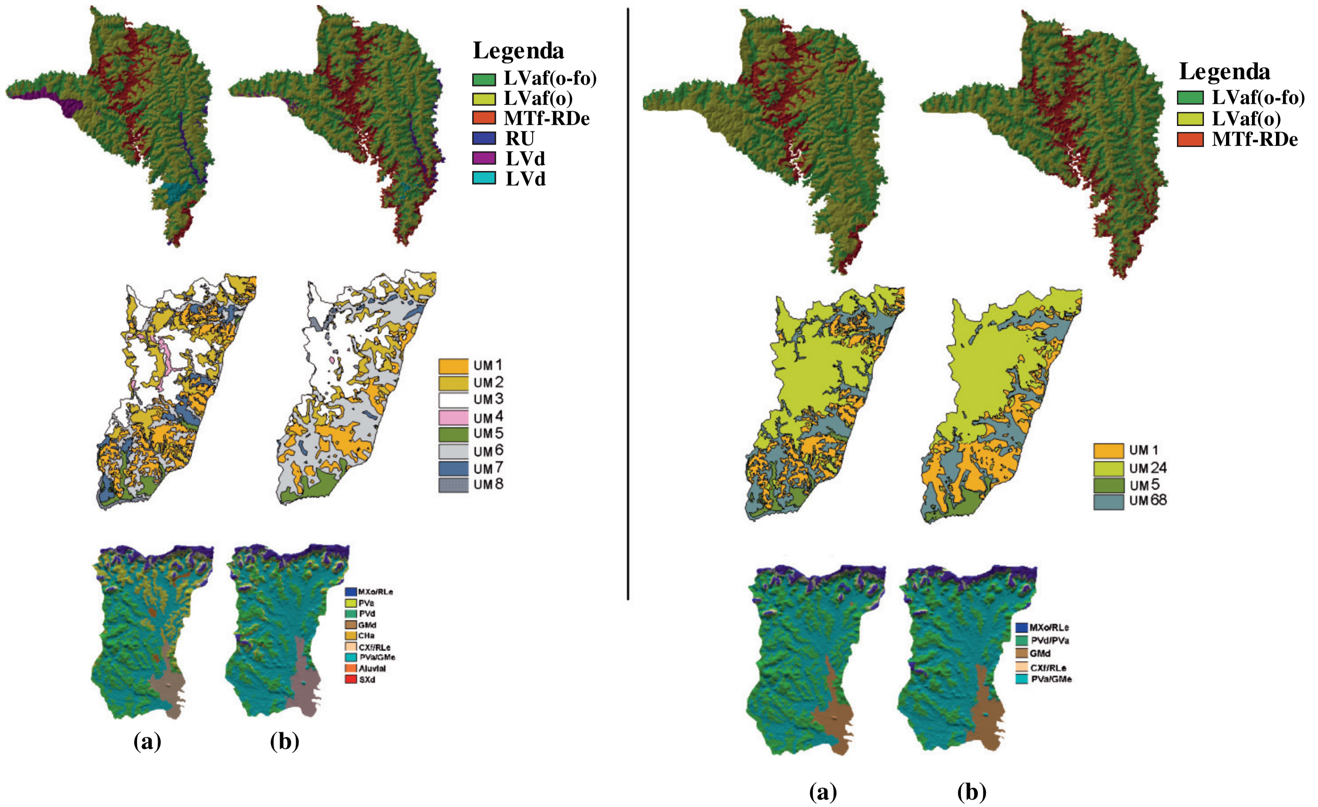


Plate 20 Comparison of flow-accumulation based on a Monte Carlo simulation using a single-flow approach (**a**) and a multiple-flow approach (**b**) for a section of the geological map of the Republic of Niger (**c** = original, **d** = corrected) (See also Figure 18.2 on page 216)



[t]

Plate 21 Original soil maps (**a**) and respective estimated soil maps (**b**) using the original soil survey legend (left side of plate) and taxonomic simplifications of the original soil survey legend for (right side of plate) Ibirubá Municipality (up of plate), Sentinela do Sul Municipality (middle of plate), and Arroio Portão Watershed (bottom ○) (See also Figure 19.1 on page 228)

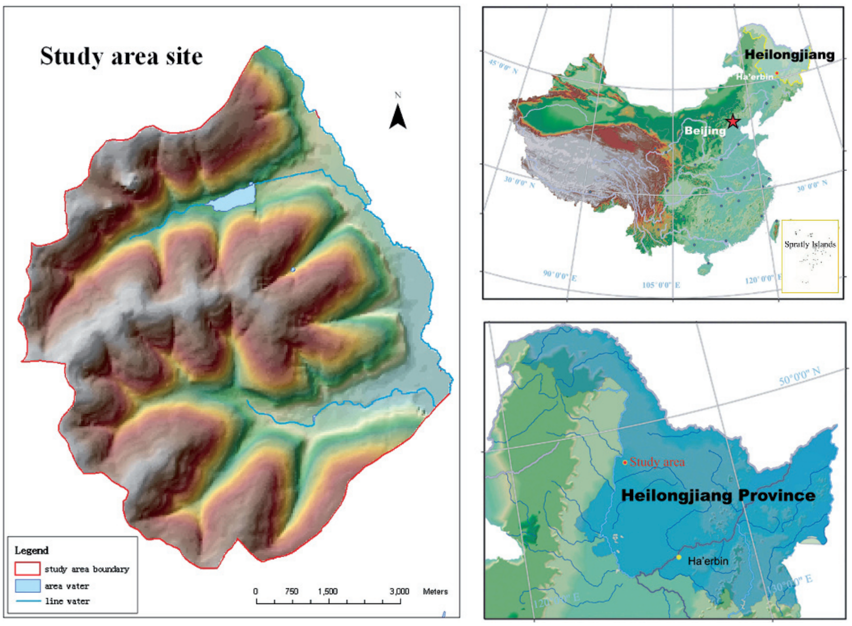


Plate 22 Location of study site: Heshan Farm, Nenjiang county, Heilongjiang Province, China (See also Figure 20.1 on page 238)

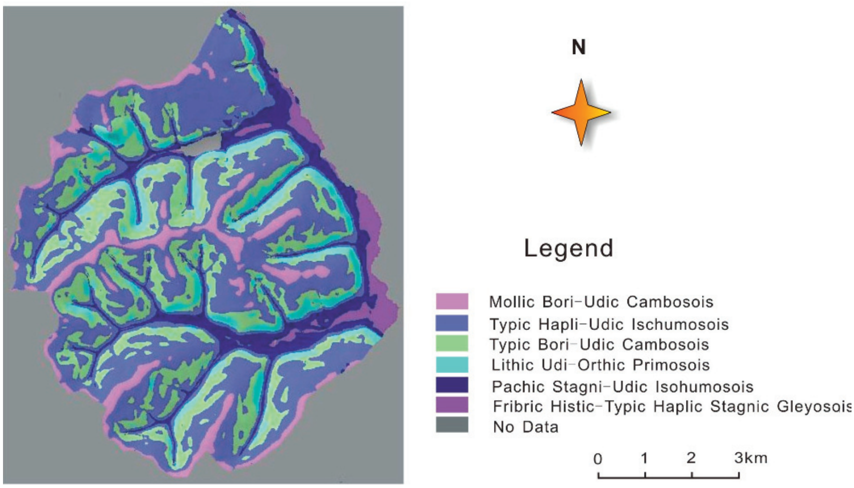


Plate 23 Soil map produced from SoLIM using the soil-landscape model constructed using the FCM-based method (See also Figure 20.4 on page 242)

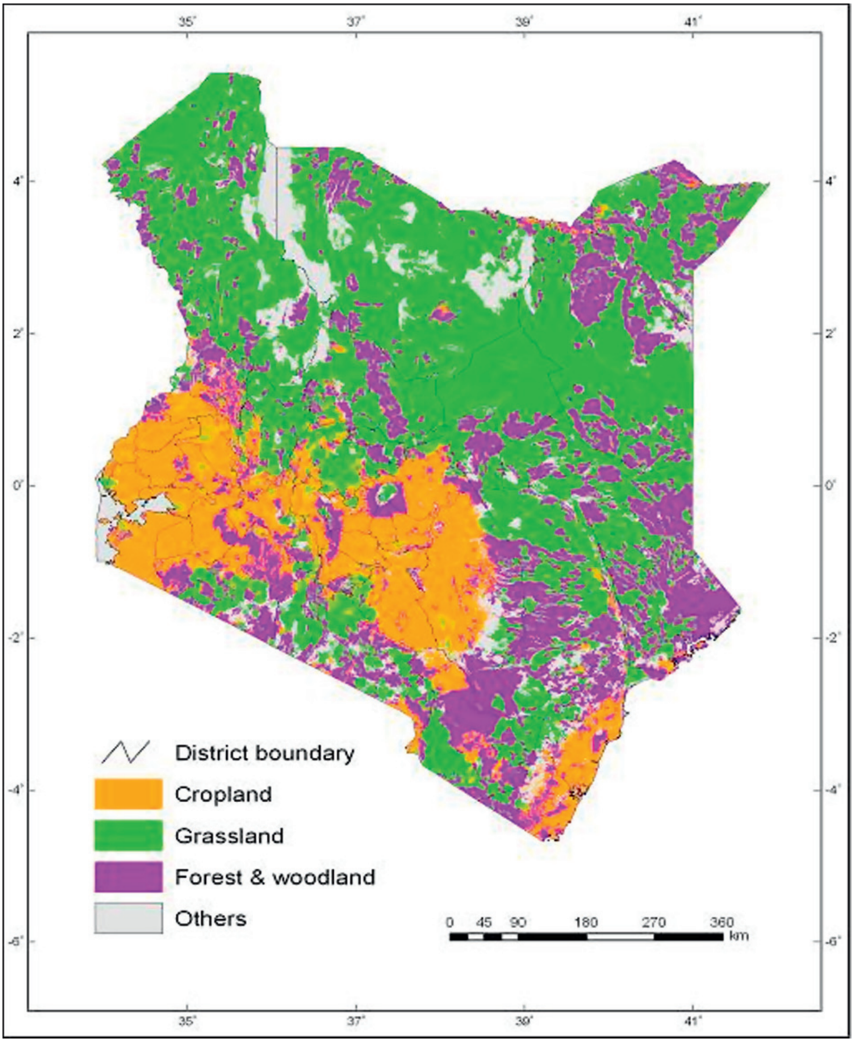


Plate 24 Kenya, dominant land use types (FAO 2005) (See also Figure 21.1 on page 249)

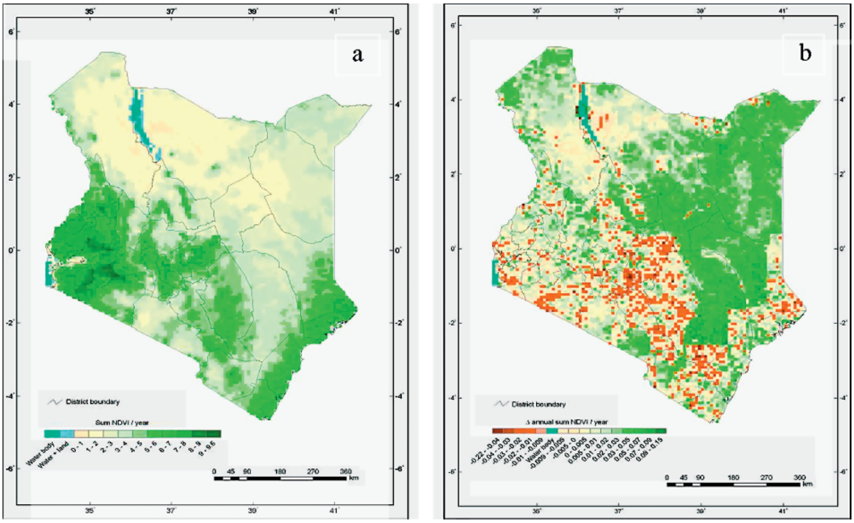


Plate 25 Spatial pattern (**a**) and temporal trend (**b**) of biomass 1981–2003 (See also Figure 21.2 on page 252)

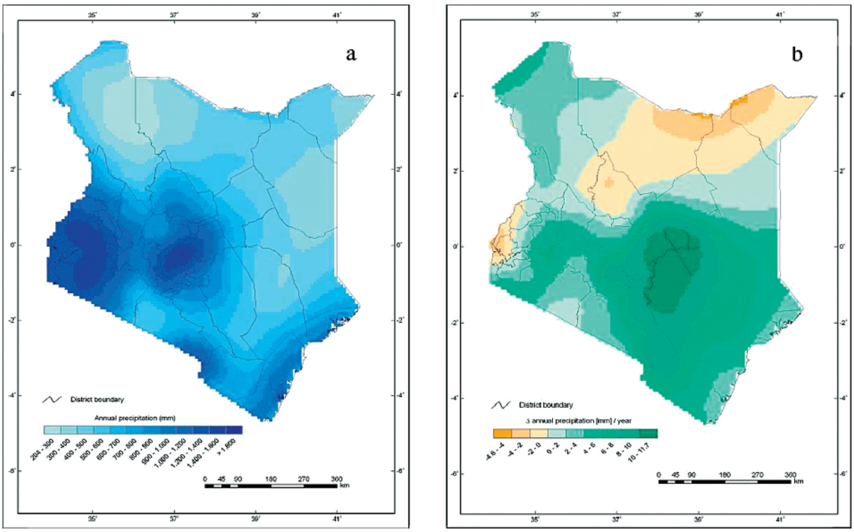


Plate 26 Spatial pattern (**a**) and temporal trend (**b**) of annual rainfall 1980–2002 (See also Figure 21.4 on page 253)

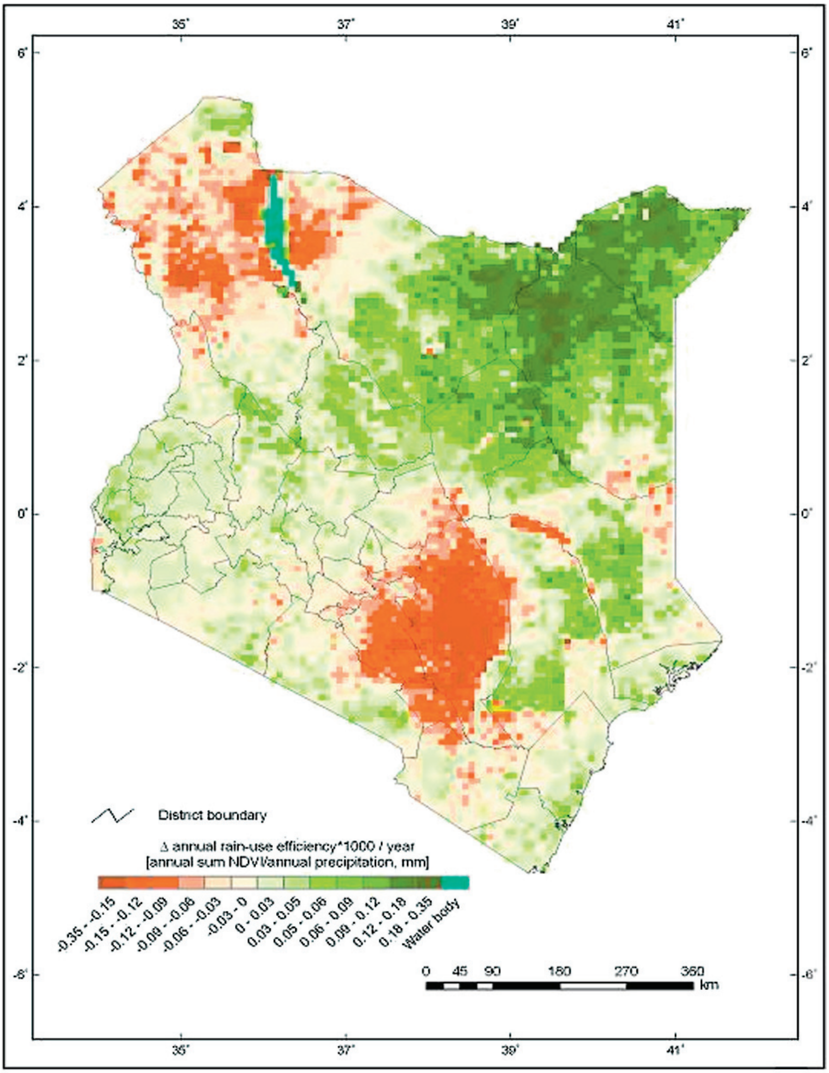


Plate 27 Trend of rain-use efficiency 1981–2002 (See also Figure 21.6 on page 254)

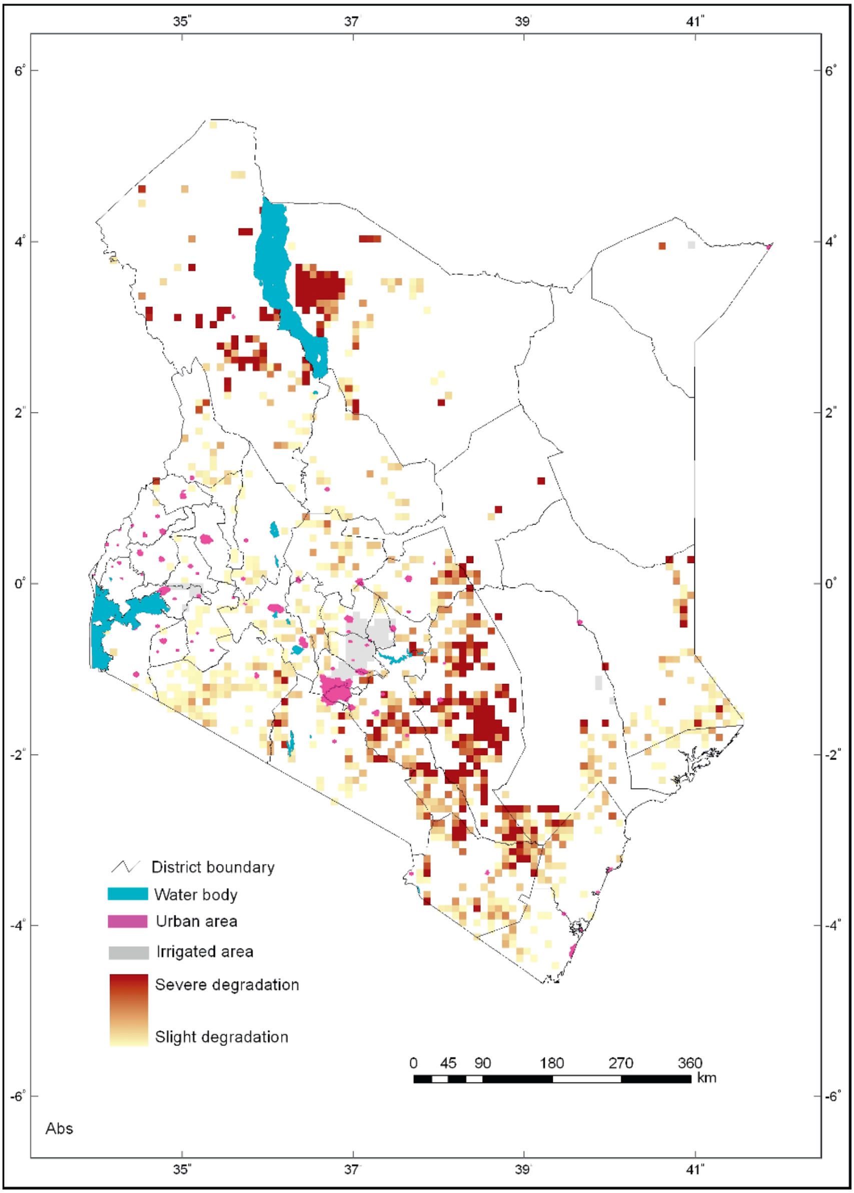


Plate 28 Kenya: black spots of land degradation between 1981 and 2003 (See also Figure 21.7 on page 256)

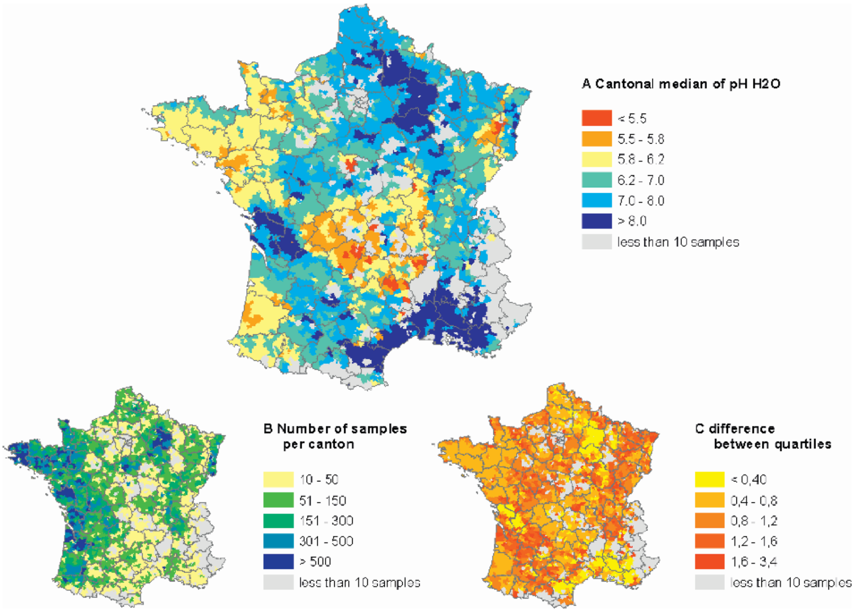


Plate 29 Maps of cantonal statistics of pH in water of cultivated topsoil for the period 1995–1999: **(A)** cantonal median value; **(B)** number of samples per canton; **(C)** inter-quartile value (See also Figure 23.1 on page 276)

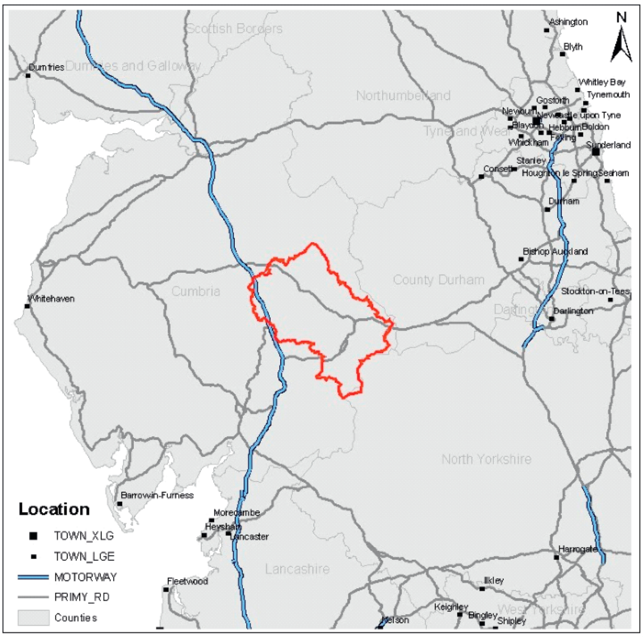


Plate 30 Location map (See also Figure 25.1 on page 293)

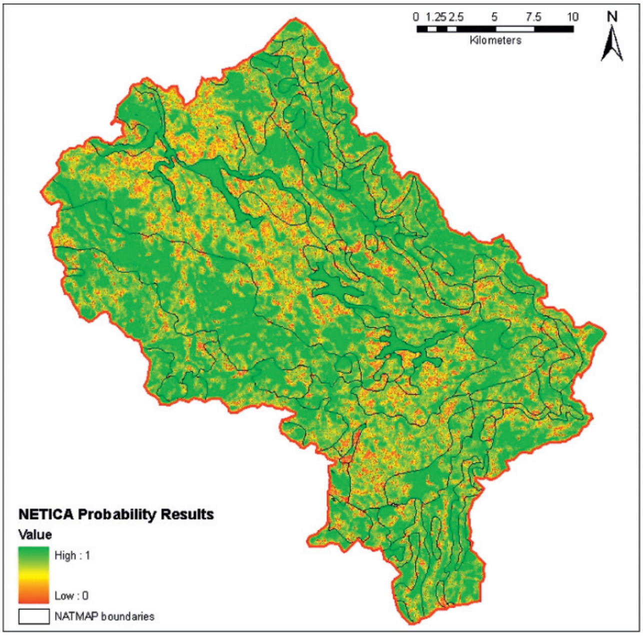


Plate 31 Model quality assessment (See also Figure 25.2 on page 295)

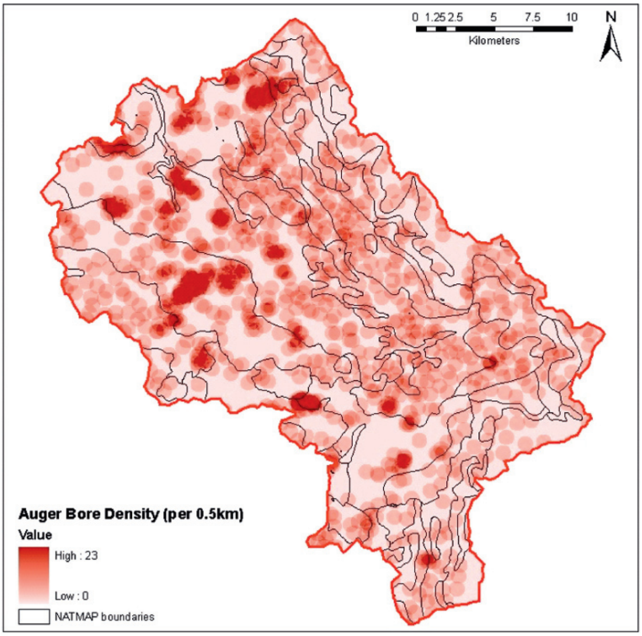


Plate 32 Spatial assessment (See also Figure 25.3 on page 298)

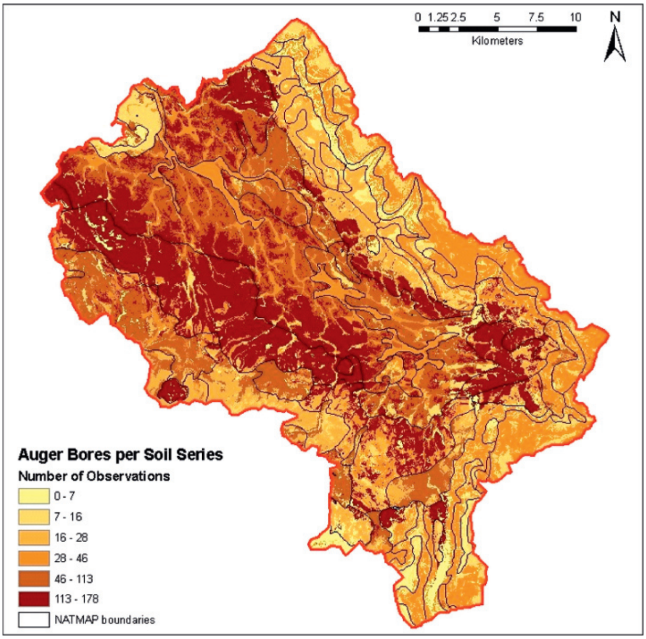


Plate 33 Thematic assessment (See also Figure 25.4 on page 299)

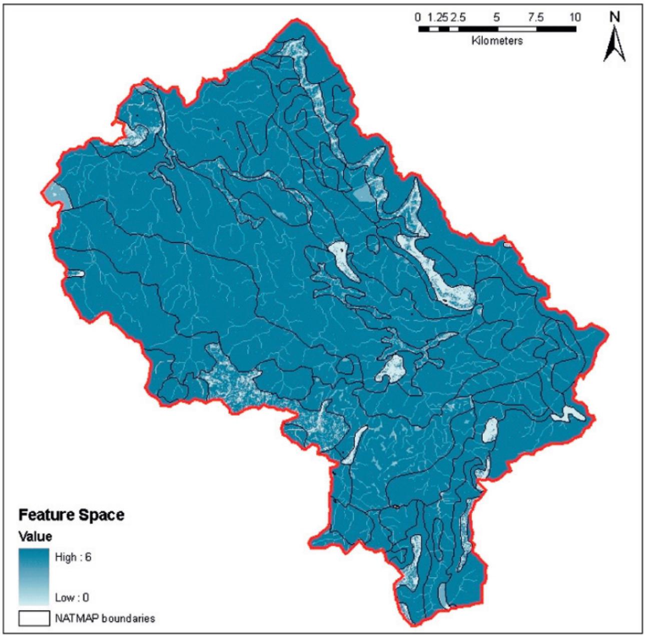


Plate 34 Feature space assessment (See also Figure 25.5 on page 299)

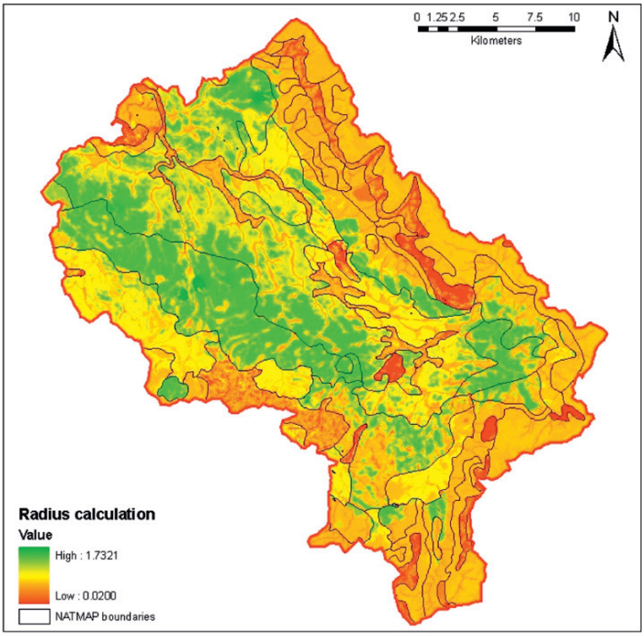


Plate 35 Combined quality assessment (See also Figure 25.6 on page 300)

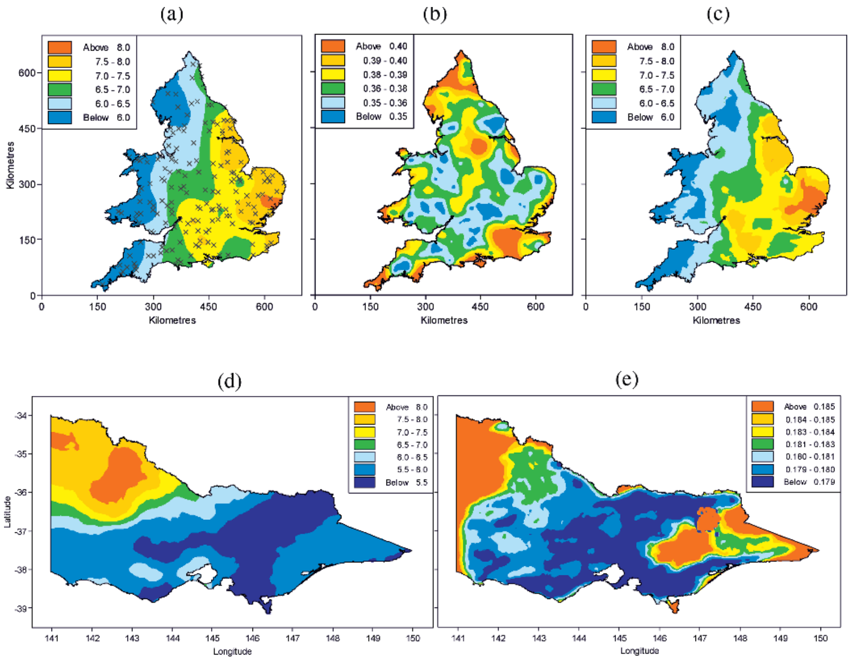


Plate 36 England and Wales maps (**a**) kriged predictions of pH 1991 using the data averaged within a 5 km radius, sample locations are plotted as crosses; (**b**) kriged estimation variances of pH 1991 using the data averaged within a 5 km radius; and (**c**) kriged predictions of raw pH 1991. Maps of kriged predictions in Victoria, Australia of (**d**) pH; and (**e**) kriged estimation variances of pH (See also Figure 27.3 on page 316)

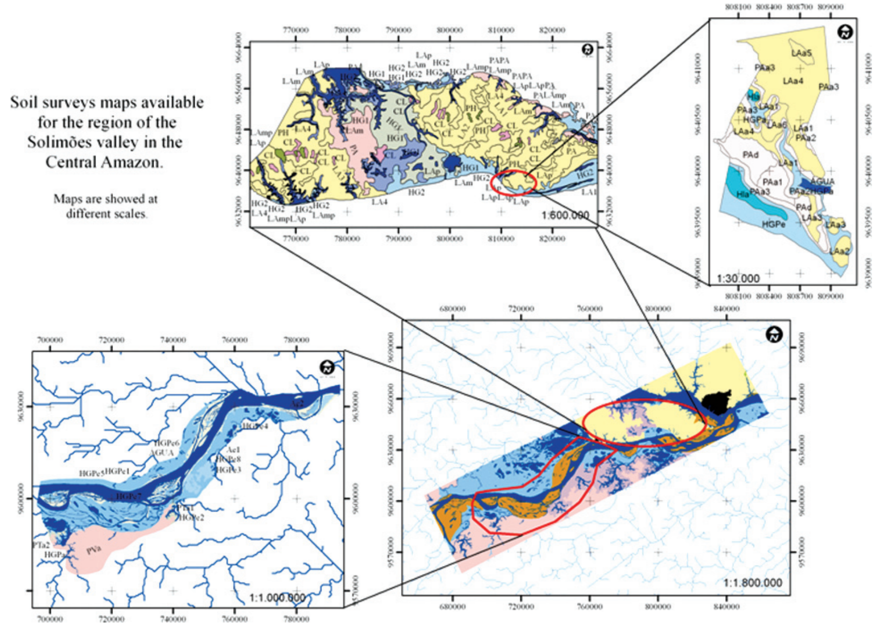


Plate 37 Location of the different soil survey maps available for the Central Amazon. The maps have different site sampling intensity ranging from compatible with a semi-detailed legend (Research Station of Caldeirão – to 1:10,000) to an exploratory soil survey (SIPAM Digital Soil Data Base – 1:250,000) (See also Figure 29.2 on page 329)

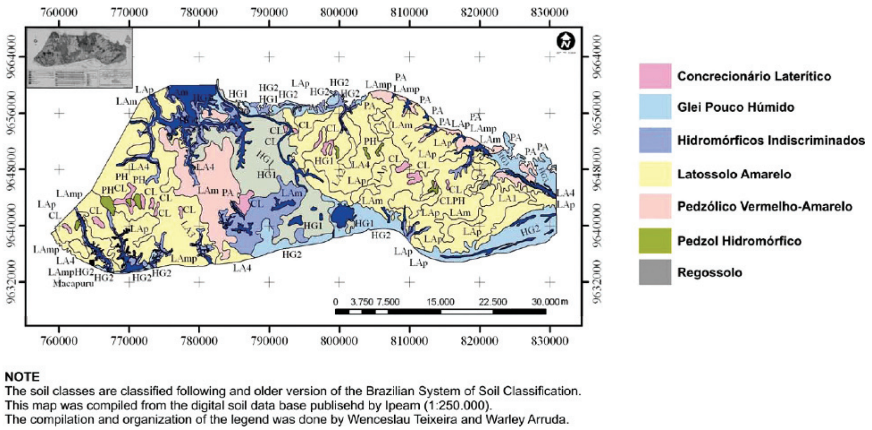
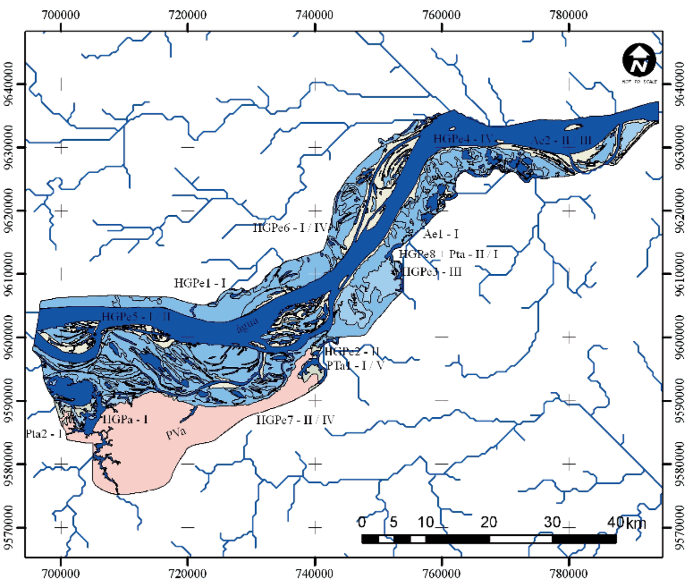


Plate 38 Renewed reconnaissance soil map from Cacau-Pirêra to Manacapuru (Roadway AM 070) – Published by IPEAM (1970). The original map of the reconnaissance soil map from Cacau-Pirêra to Manacapuru (Roadway AM 070) is showed in the *small box left* side (See also Figure 29.3 on page 331)

Reconnessaince soil map of aluvial soils near the city of Manacapuru - AM



- Glei Pouco Húmico Álico
- Glei Pouco Húmico Eutrófico
- Plintossolo Álico
- Podzólico Vermelho-Amarelo Álico
- Solos Aluviais Eutróficos

NOTE

The soil classes are presented as in the printed map and follow an older version of the Brazilian System of Soil Classification.
Those soils were also classified following a duration of flooding period (inundação).
This scheme were divided: I) flooding has a duration less than one month; II) flooding between 1 and 3 months every year, III) flooding between 3 and 6 months and IV) flooded period is higher than 6 months.

Plate 39 Renewed reconnaissance soil map of the soil near the *border* of Solimões in the city of Manacapuru (CETEC, 1986) (See also Figure 29.4 on page 332)

(a)

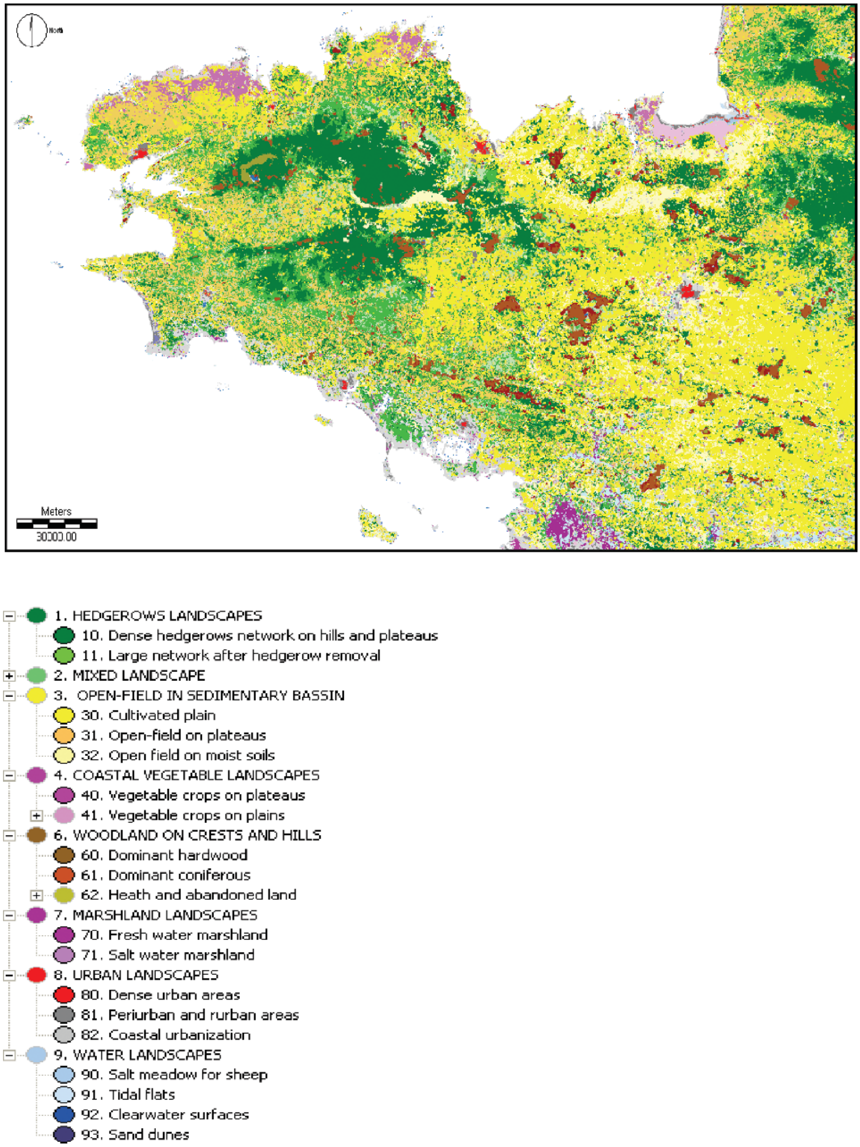


Plate 40 Landscape classification realized by means of MODIS satellite images at regional scale (Phase 1) (See also Figure 30.2a on page 343)

(b)

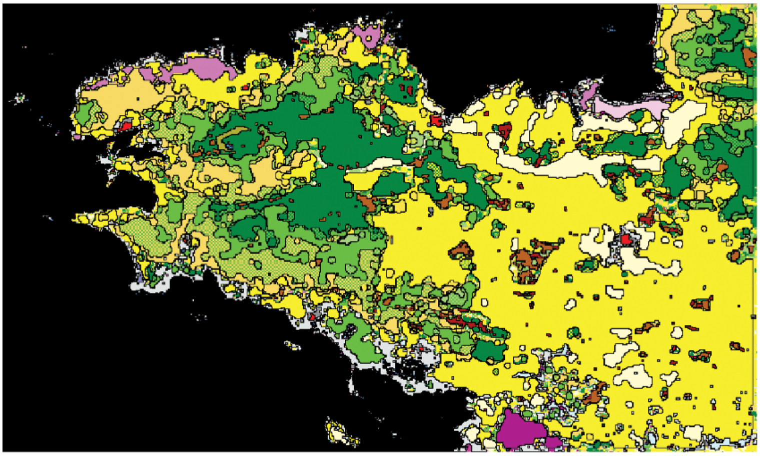


Plate 41 Post-processing techniques applied to landscape classification (Phase 2) (See also Figure 30.2b on page 344)

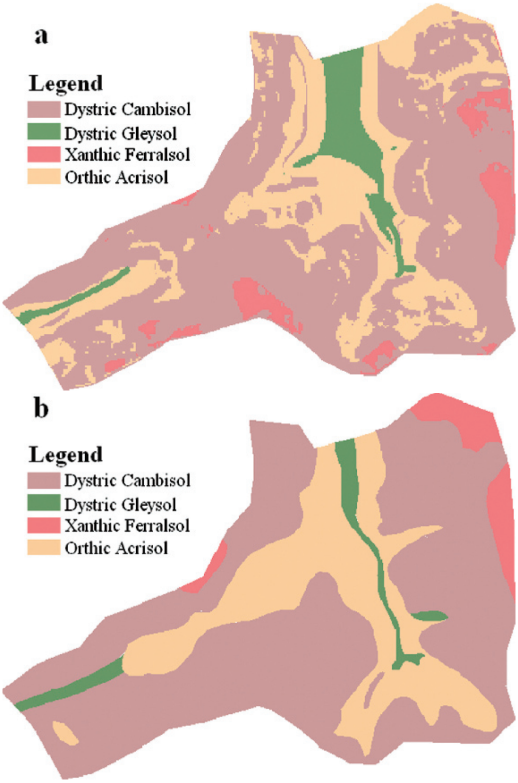


Plate 42 Preliminary digital soil map derived from DEM (**a**), and final soil map elaborated by traditional soil mapping (**b**) of CAPTA-Frutas, Jundiaí, SP, Brazil) (See also Figure 31.5 on page 354)

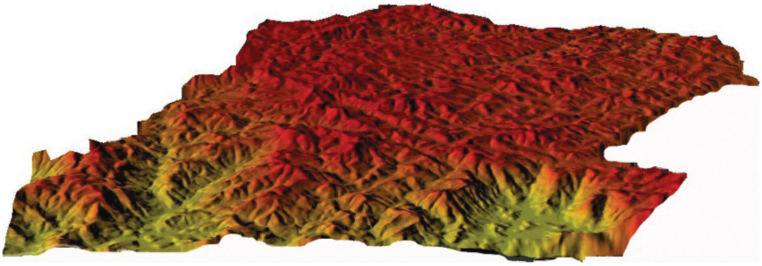


Plate 43 3D visualization with 3x of vertical exaggeration of hydrologic correct DEM with resolution of 20 m (See also Figure 32.1 on page 360)

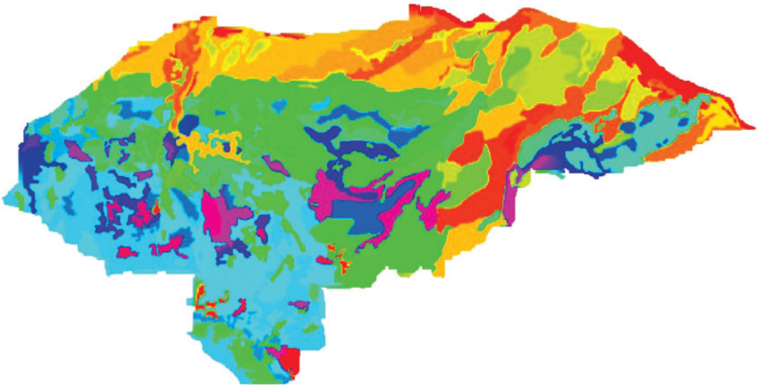


Plate 44 FAO soil map for Honduras (See also Figure 33.1 on page 369)

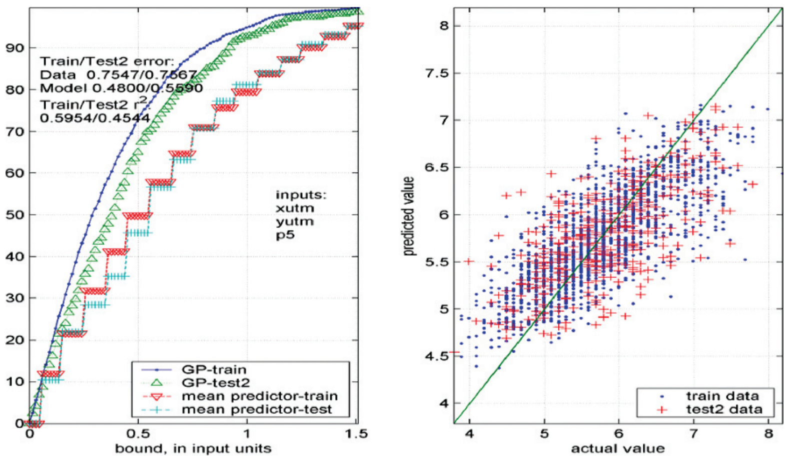


Plate 45 Model performance for pH in topsoil. The figure on the left shows the comparative performance of the model vs. a mean predictor. The *x* coordinate is the bound, in pH units, and the *y* coordinate is the percentage of the predictions that fit within the predicted value +/− the bound. For example, 95% of the predictions will fall within 1 pH unit of the predictions for the training set. This number is slightly lower for the independent test set (92%) and much lower for a mean predictor (80%). The figure on the right shows actual values versus predicted values. In an ideal case, both would be the same (*solid, green line*), but in practice there will always be dispersion around the *y* axis. The more dispersion, the worse the model is (See also Figure 33.2 on page 376)

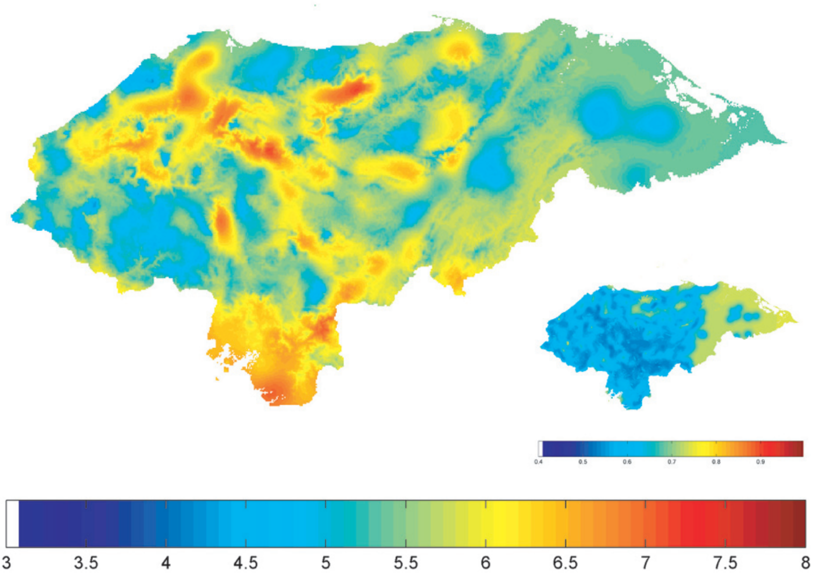


Plate 46 Predicted map of pH in topsoil and 67% confidence interval (See also Figure 33.3 on page 376)

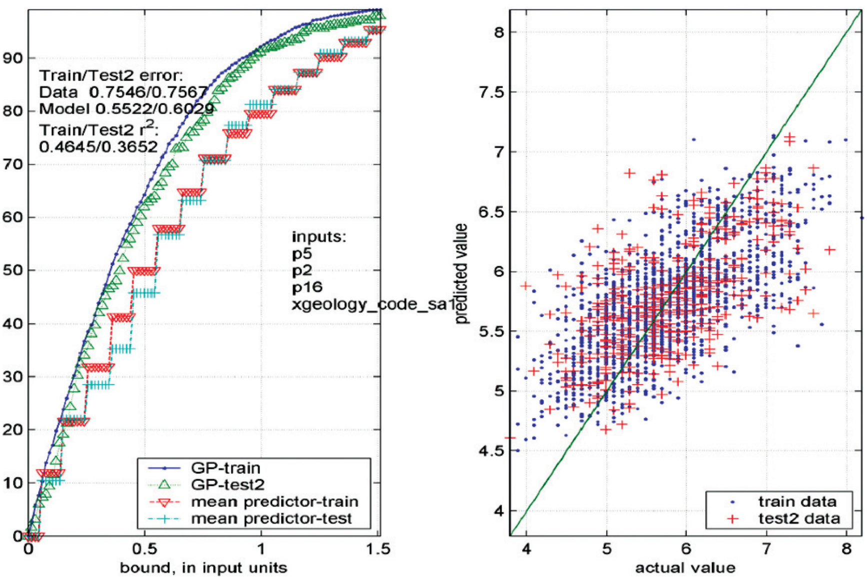


Plate 47 Model performance for pH in topsoil without spatial interpolation (See also Figure 33.4 on page 377)

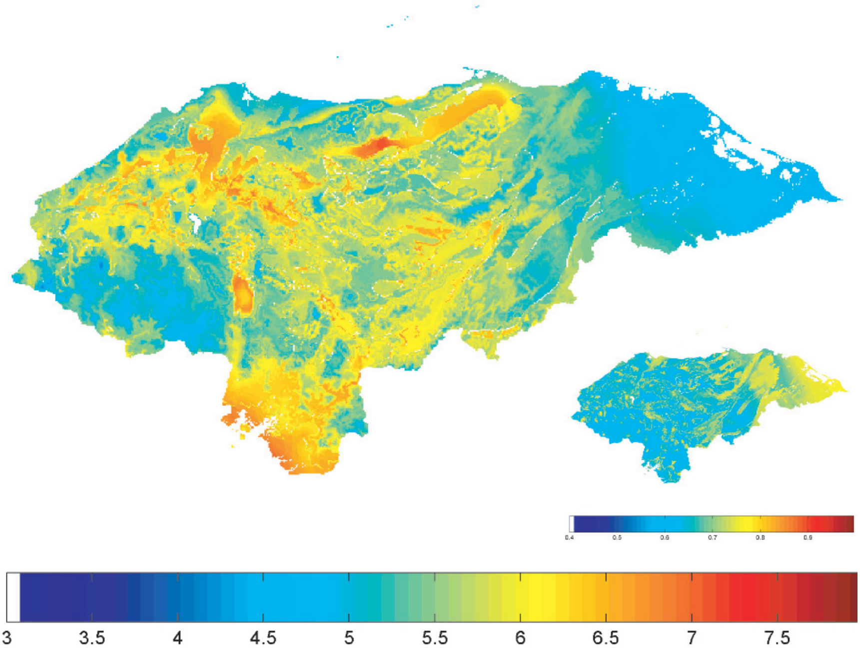


Plate 48 Predicted map of pH in topsoil and 67% confidence interval, without using spatial interpolation (See also Figure 33.5 on page 378)

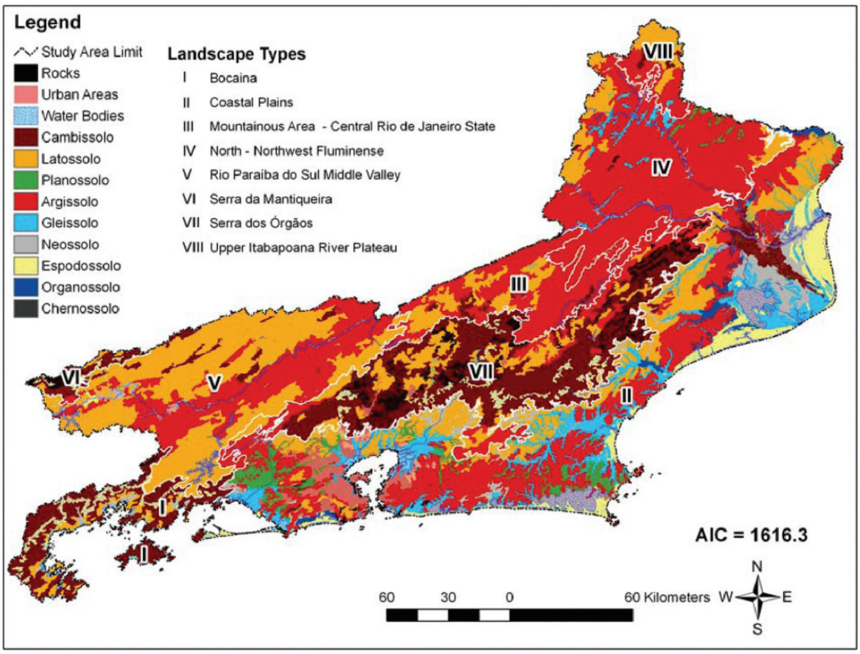


Plate 49 Digital soil class map (at order level of SiBCS) as predicted by the best model (M1) with landscape types (See also Figure 34.6 on page 394)

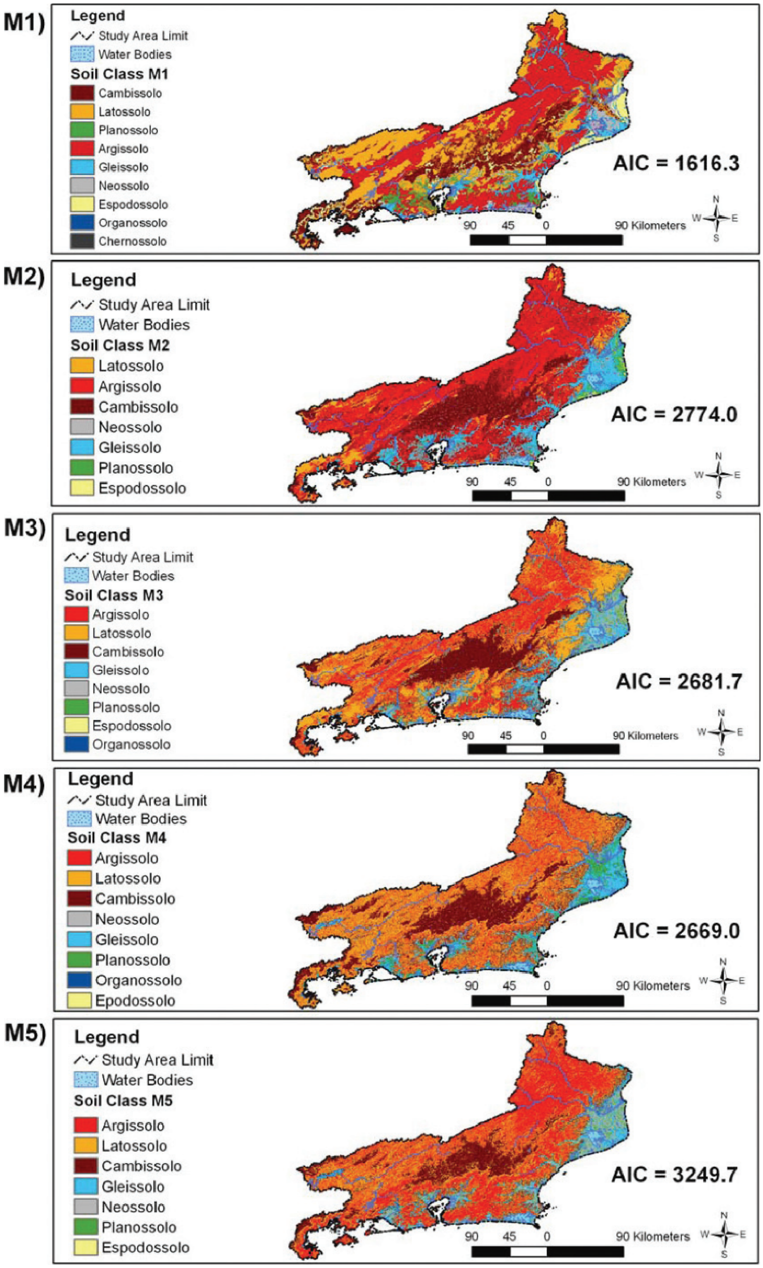


Plate 50 Resulting maps of predicted soil classes from the five models (See also Figure 34.5 on page 393)

Index