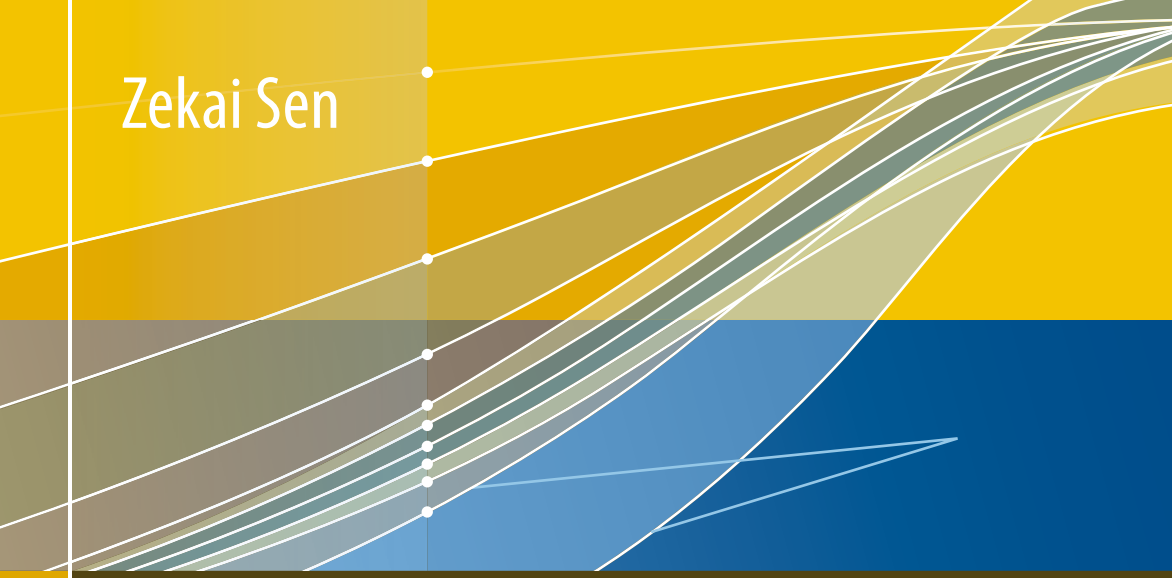


Zekai Sen

Spatial Modeling Principles in Earth Sciences

Second Edition

Springer

Spatial Modeling Principles in Earth Sciences

Zekai Sen

Spatial Modeling Principles in Earth Sciences

Second Edition

Springer

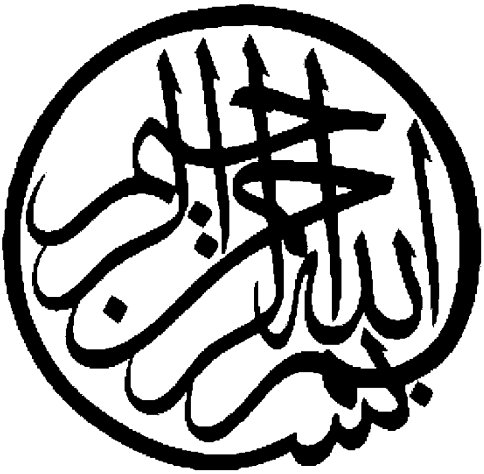
Zekai Sen
Faculty of Civil Engineering
Istanbul Technical University
Istanbul, Turkey

ISBN 978-3-319-41756-1 ISBN 978-3-319-41758-5 (eBook)
DOI 10.1007/978-3-319-41758-5

Library of Congress Control Number: 2016951726

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland



Our earth is the only place where we can live in harmony, peace, and cooperation for the betterment of humanity. It is the duty of each individual to try with utmost ambition to care for the earth environmental issues so that a sustainable future can be handed over to new generations. This is possible only through the scientific principles, where the earth systems and sciences are the major branches.
This book is dedicated to those who care for such a balance by logical, rational, scientific, and ethical considerations for the sake of other living creatures' rights.

Preface

Earth systems and sciences cover a vast amount of disciplines that are effective in the daily life of human beings. Among these are the atmospheric sciences, meteorology, hydrogeology, hydrology, petroleum geology, engineering geology, geophysics, and marine and planary systems. Their subtopics such as floods, groundwater and surface water resources, droughts, rock fractures, and earthquakes have basic measurements in the form of records of their past events and need to have suitable models for their spatio-temporal predictions. There are various software for the solution of problems related to earth systems and sciences concerning the environment, but unfortunately they are ready-made models, most often without any basic information. It is the main purpose of this book to present the basic knowledge and information about the evolution of earth sciences events on rational, logical, and at places scientific and philosophical features so that the reader can grasp the underlying principles in a simple and applicable manner. The book also directs the reader to proper basic references for further reading and enlarging the background information. I have gained almost all of the field, laboratory and theoretical as well actual applications, during my stay at the Faculty of Earth Sciences, Hydrogeology Department, King Abdulaziz University (KAU), and the Saudi Geological Survey (SGS), which are in Jeddah, Kingdom of Saudi Arabia. Additionally, especially on the atmospheric sciences, meteorology and surface water resources aspects are developed at Istanbul Technical University (İTÜ), Turkey.

It is well sought to adapt spatial modeling and simulation methodologies in actual earth sciences problem solutions for exploring the inherent variability such as in fracture frequencies, spacing, rock quality designation, grain size distribution, groundwater exploration and quality variations, and many similar random behaviors of the rock and porous medium. The book includes many innovative spatial modeling methodologies with actual application examples from real-life problems. Most of such innovative approaches appear for the first time in this book with the necessary interpretations and recommendations of their use in real life.

The second print of the book indicates the need for spatial modeling in earth sciences. The book has an additional chapter and also some recent methodological procedures in some chapters, which cannot be found in the first print. I wish that the content will be beneficial to anyone interested in spatial earth system modeling and simulation.

Throughout the first and this second edition preparation process, my wife Fatma Hanim has encouraged me to think that such a work will be beneficial to all humans in the world and those who are interested in the topics of this book. I appreciate the encouragement by Springer Publishing Company for the second printing of this book.

Istanbul, Turkey Zekai Sen
18 May 2016

Contents

Chapter 1
Introduction

Abstract Earth science events have spatial, temporal and spatio-temporal variabilities depending on the scale and the purpose of the assessments. Various earth sciences branches such as hydrogeology, engineering geology, petroleum geology, geophysics, and related topics have, in general, spatio-temporal variabilities that need for effective modeling techniques for proper estimation and planning the future events in each branch according to scientific principles. Determinism and especially uncertainty techniques are frequently used in the description and modeling these events conveniently through simple and rather complicated computer software. However, the basic principles in any software require simple and effective mathematical, probabilistic, statistical, stochastic and recently fuzzy methodologies or their combination for an objective solution of the problem based on field or laboratory data. This chapter provides comparetive and simple explanation of each one of these approaches.

Keywords Earth sciences • Model • Randomness • Probability • Random field • Statistics • Stochastic • Variability

1.1 General

There has been a good deal of discussion and curiosity about the natural event occurrences during the last century. These discussions have included comparisons between uncertainty in earth and atmospheric sciences and uncertainty in physics which has, inevitably it seems, led to the question of determinism and indeterminism in nature (Leopold and Langbein 1963; Krauskopf 1968; Mann 1970).

At the very core of scientific theories lies the notion of "cause" and "effect" relationship in an absolute certainty in scientific studies. One of the modern philosophers of science, Popper (1957), stated that "to give a causal explanation of a certain specific event means deducing a statement describing this event from two kinds of premises: from some universal laws, and from some singular or specific statements which we may call the specific initial conditions." According to him there must be a very special kind of connection between the premises and the conclusions of a causal explanation, and it must be deductive. In this manner, the

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_1

conclusion follows necessarily from the premises. Prior to any mathematical formulation, the premises and the conclusion consist of verbal (linguistic) statements. It is necessary to justify at every step of deductive argument by citing a logical rule that is concerned with the relationship among statements. On the other hand, the concept of "law" lies at the heart of deductive explanation and, therefore, at the heart of the certainty of our knowledge about specific events.

Recently, the scientific evolution of the methodologies has shown that the more the researchers try to clarify the boundaries of their domain of interest, the more they become blurred with other domains of research. For instance, as the hydrogeologist tries to model the groundwater pollution as one of the modern nuisances of humanity as far as the water resources are concerned, they need information about the geological environment of the aquifers, meteorological and atmospheric conditions for the groundwater recharge, and social and human settlement environmental issues for the pollution sources. Hence, many common philosophies, logical basic deductions, methodologies, and approaches become common to different disciplines, and the data processing is among the most important topics which include the same methodologies applicable to diversity of disciplines. The way that earth, environmental, and atmospheric scientists frame their questions varies enormously, but the solution algorithms may include the same or at least similar procedures. Some of the common questions that may be asked by various research groups are summarized as follows. Most of these questions have been explained by Johnston (1989).

Any natural phenomenon or its similitude occurs extensively over a region, and therefore, its recordings or observations at different locations pose some questions such as, for instance, are there relationships between phenomena in various locations? In such a question, the time is as if it is frozen and the phenomenon concerned is investigated over the area and its behavioral occurrence between the locations. An answer to this question may be provided descriptively in linguistic, subjective, and vague terms which may be understood even by nonspecialists in the discipline. However, their quantification necessitates objective methodologies which are one of the purposes of the context in this book. Another question that may be stated right at the beginning of the research in the earth, environmental, and atmospheric sciences is: are places different in terms of the phenomena present there? Such questions are the source of many people's interest in the subject.

Our minds are preconditioned on the Euclidian geometry, and consequently ordinary human beings are bound to think in 3D spaces as length, width, and depth in addition to the time as the fourth dimension. Hence, all the scientific formulations, differential equations, and others include space and time variabilities. All the earth, environmental, and atmospheric variables vary along these four dimensions. If their changes along the time are not considered, then it is said to be frozen in time, and therefore a steady situation remains along the time axis but variable in concern has variations along the space. A good example for such a change can be considered as geological events which do not change with human life time span. Iron content of a rock mass varies rather randomly from one point to another within the rock and hence spatial variation is considered. Another example

is the measurement of rainfall amounts at many irregularly located meteorology stations spread over an area, i.e., simultaneous measurement of rainfall amounts; again the time is frozen and the spatial variation is sought.

On the other hand, there are timewise variations which are referred to as the temporal variations in the natural phenomenon. For such a variation, it suffices to measure the event at a given location which is the case in any meteorology station. Depending on the time evolution of the event whether it is continuous or not, time series records can be obtained. A time series is the systematic measurement of any natural event along the time axis at regular time intervals. Depending on this time interval, time series is called as hourly, daily, weekly, monthly, or yearly time series. Contrary to time variations, it is not possible to consider space series where the records are kept at regular distances except in very specific cases. For example, if water samples along a river are taken at every 1 km, then the measurements provide a distance series in the regularity sense. In fact, distance series are very limited as if there are no such data sequences. On the other hand, depending on the interest of event, there are series which include time data, but they are not time series due to irregularity or randomness in the time intervals between successive occurrences of the same event. Flood and drought occurrences in hydrology correspond to such cases. One cannot know the duration of floods or droughts. Likewise, in meteorology the occurrence of precipitation or any dangerous levels of concentrations of air pollutants all do not have time series characteristics.

Any natural event evolves in the 4D human visualization domains, and consequently its records should involve the characteristics of both time and space variabilities. Any record that has this property is referred to as the spatiotemporal variation.

Mathematical, statistical, probabilistic, stochastic, and alike procedures are applicable only in the case of spatial or temporal variability in natural or artificial phenomena. It is not possible to consider any approach of earth sciences phenomena without the variability property, which is encountered everywhere almost explicitly but at times and places also implicitly. Explicit variability is the main source of reasoning, but implicit variability leads to thousands of imagination with different geometrical shapes on which one is then able to ponder and generate ideas and conclusions. It is possible to sum up that the variability is one of the fundamental ingredients of philosophic thinking, which can be separated into different rational components by mental restrictions based on the logic. Almost all social, physical, economical, and natural events in small scales and phenomena in large scales include variability at different levels (micro, meso, or macro) and types (geometry, kinematics, and dynamics). Hence, the very word "variability" deserves detailed qualitative understanding for the construction of our future quantitative models.

Proper understanding of earth sciences phenomenon is itself incomplete, rather vague, and cannot provide a unique or precise answer. However, in the case of data availability, the statistical methods support the phenomenon understanding and deducing meaningful and rational results. These methods suggest the way to weight the available data so as to compute best estimates and predictions with acceptable

error limits. Unfortunately statistics and use of its methods are taken as cookbook procedures without fundamental rational background in problem solving. There are many software programs available to use, but the results cannot be interpreted in a meaningful, plausible, and rational manner for the service of practical applications or further researches.

It is, therefore, the purpose of this book to provide fundamentals, logical basis, and insights into spatial statistical techniques that are frequently used in engineering applications and scientific research works. In this manner prior to cookbook procedure applications and software use, the professional can give his expert view about the phenomenon concerned and the treatment alternatives of the available data. The major problems in the spatial analysis are the point estimation from a set of data sampling points where the measurements are not found, areal average calculations and contour mapping of the regionalized variables. The main purpose of this chapter is to lay out the basic spatial variability ingredients, their simple conceptual grasp and models.

1.2 Earth Sciences Phenomena

The phenomenologic occurrences in earth sciences are natural events, and their prediction and control need scientific methodological approaches under the domain of uncertainty and risk conceptions in cases of design for mitigation against their dangerous consequences and inflictions on the society at large. The common features of these phenomena in general are their rather random behaviors in amounts, occurrence time and location, duration, intensity, and spatial coverages. Although future average behaviors are taken as a basis in any design, it is recommended in this book that risk concept at 5 or 10 % must be taken into consideration and accordingly the necessary design structures must be implemented for reduction of dangerous occurrences and impacts. The necessary scientific algorithms, models, procedures, programs, seminars, and methodologies and if possible a comprehensive software must be prepared for effective, speedy, and timely precautions. Earth sciences hazards must be assessed logically, conceptually, and numerically by an efficient monitoring system and following data treatments. In various chapters of this book, different methodological data-processing procedures are presented with factual data applications.

Earth sciences deal with spatial and temporal behaviors of natural phenomena at every scale for the purpose of predicting the future replicas of the similar phenomena which help to make significant decisions in planning, management, operation, and maintenance of natural occurrences related to social, environmental, and engineering activities. Since any of these phenomena cannot be accounted by measurements which involve uncertain behaviors, their analysis, control, and prediction need to use uncertainty techniques for significant achievements for future characteristic estimations. Many natural phenomena cannot be monitored at desired instances of time and locations in space, and such restrictive time and

location limitations bring additional irregularity in the measurements. Consequently, the analyst, in addition to uncertain temporal and spatial occurrences, has the problem of sampling the natural phenomenon at irregular sites and times. For instance, floods, earthquakes, car accidents, illnesses, and fracture occurrences are all among the irregularly distributed temporal and spatial events. Uncertainty and irregularity are the common properties of natural phenomenon measurements in earth and atmospheric researches, but the analytical solutions through numerical approximations all require regularly available initial and boundary conditions that cannot be obtained by lying regular measurement sites or time instances. In an uncertain environment, any cause will be associated with different effects each with different levels of possibilities. Herein, possibility means some preference index for the occurrence of each effect. The greater the possibility index, the more frequent the event occurrence.

Geology, hydrology, and meteorology are the sciences of lithosphere, hydrosphere, and atmosphere that consist of different rock, water, and air layers, their occurrences, distribution, physical and chemical genesis, mechanical properties, and structural constitutions and interrelationships. It is also one of the significant subjects of these phenomena to deal with historical evolution of rock, water, and atmosphere types and masses concerning their positions, movement, and internal contents. These features indicate the wide content of earth sciences studies. Unfortunately, these phenomena cannot be simulated under laboratory conditions, and therefore, field observations and measurements are the basic information sources of information.

In large scales, geological, hydrological, and meteorological compositions are very anisotropic and heterogeneous; in small scales, their homogeneity and isotropy increase but the practical representatively decreases. It is therefore necessary to study them in rather medium scales that can be assumable as homogeneous and isotropic. In any phenomenon study, the general procedure is to have its behavioral features and properties at small locations and by correlation to generalize to larger scales. Unfortunately, most often the description of earth sciences phenomena provides linguistic and verbal qualitative interpretations that are most often subjective and depend on the common consensus. The more the contribution to such consensus of experts, the better are the conclusions and interpretations, but even then it is not possible to estimate or derive general conclusions on an objective basis. It is, therefore, necessary to try and increase the effect of quantitative approaches and methodologies in earth sciences, especially by the use of personal computers, which help to minimize the calculation procedures and time requirements by software. However, software programs are attractive in appearance and usage, but without fundamental handling of procedures and methodologies, the results from these programs cannot be done properly with useful interpretations and satisfactory calculations that are acceptable by common expertise. This is one of the very sensitive issues that are mostly missing in software program training or ready software uses.

It is advised in this book that without knowing the fundamentals of earth sciences procedure, methodology, or data processing, one must avoid the use of

software programs. Otherwise, mechanical learning of software from colleagues and friends or during tutorials with missing fundamentals does not lead the user to write proper reports or even to discuss the results of software with somebody expert in the area who may not know software use.

Phenomena in earth sciences are of multitude types, and each needs at times special interpretation, but whatever the differences, there is one common basis, which is the data processing. Hence, any earth sciences equipped with proper data-processing technique with fundamentals can help others and make significant contributions to open literature, which is of great need for such interpretations. Besides, whatever the techniques and technological level reached, these phenomena show especially spatial (length, area, volume) differences at many scales, areas, locations, and depths, and consequently, development of any technique cannot cover the whole of these variations simultaneously, and there is always an empty space for further interpretations and researches. It is possible to consider the earth sciences all over the world in two categories, namely, conventionalists that are more descriptive working group with linguistic and verbal interpretations, and conclusions with very basic and simple calculations. The second group includes those who are well equipped with advanced quantitative techniques starting with simple probabilistic, statistical, stochastic, and deterministic mathematical models and calculation principles. The latter group might be very addicted to quantitative approaches with little linguistic, verbal, i.e., qualitative interpretations. The view taken in this book remains in between where the earth scientist should not ignore the verbal and linguistic descriptive information or conclusions, but he also looks for quantitative interpretation techniques to support the views and arrive at general conclusions. Unfortunately, most often earth scientists are within these two extremes and sometimes cannot understand each other although the same phenomena are considered. Both views are necessary but not exclusively. The best conclusions are possible with a good combination of two views. However, the priority is always with the linguistic and verbal descriptions, because in scientific training, these are the first stones of learning and their logical combinations rationalistically constitute quantitative descriptions. Even during the quantitative studies, the interpretations are based on descriptive, qualitative, linguistic, and verbal statements. It is well known in scientific arena that the basic information is in the forms of logical sentences, postulates, definitions, or theorems, and in earth sciences, even speculations might be proposed prior to quantitative studies, and the objectivity increases with shifting toward quantitative interpretations.

Any geological phenomenon can be viewed initially without detailed information to have the following general characteristics:

1. It does not change with time or at least within the lifetime of human, and therefore, geological variations have spatial characters, and these variations can be presented in the best possible way by convenient maps. For example, geological description leads to lithological variation of different rock types. In this simplest classification of rocks, the researcher is not able to look for different possibilities or ore reserves, water sources, oil field possibilities, etc.

Oil reserves cannot be in igneous or metamorphic rock units, and therefore, he/she has to restrict the attention on the areas of sedimentary rocks. It is possible to look for regions of groundwater, which is in connection with atmosphere, i.e., the rainfall leads to direct infiltration. This information implies shallow groundwater possibilities, and consequently quaternary sediments (wadi alluviums of present epoch geological activity) can be delimited from such a map.

2. Geological units are neither uniform nor isotropic nor heterogeneous in horizontal and vertical extends. A first glance on any geological map indicates obviously that in any direction (NS, EW, etc.) the geological variation is not constant. There is the succession of different rock types and subunits along the vertical direction than is referred to as stratigraphic along which neither the thickness of each unit nor the slope is constant. It is possible to conclude from these two points that the spatial geometry of geological phenomena is not definite, and furthermore, its description is not possible with Euclidean geometry which is based on lines, planes, and volumes of regular shapes. However, in practical calculations, the geometric dimensions are simplified so as to use Euclidean representations and make simple calculations. Otherwise, just the geometry can be intangible, hindering calculations. This indicates that in calculations besides other earth sciences quantities that will be mentioned later in this book, initially geometry causes approximations in the calculations, and in any calculation, geometry is involved. One can conclude from this statement that even if the other spatial variations are constant, approximation in geometry gives rise to approximate results. It is very convenient to mention here that the only definite Euclidean geometry exists at rock crystalline elements. However, their accumulation leads to irregular geometry that is not expressible by classical geometry.
3. Rock materials are not isotropic and heterogeneous even at small scales of few centimeters except at crystalline or atomic levels. Hence, the compositional variation within the geometrical boundaries differs from one point to another, which makes the quantitative appreciation almost impossible. In order to alleviate the situation mentally, researchers visualize an ideal world by simplifications leading to the features as isotropic and homogeneous.
4. Isotropy implies uniformity along any direction, i.e., directional property constancy. The homogeneity means constancy of any property at each point. These properties can be satisfied in artificial material produced by man, but natural material such as rocks and any natural phenomenon in hydrology and meteorology cannot have these properties in the absolute sense. However, provided that the directional or pointwise variations are not very appreciably different from each other, then the geological medium can be considered as homogeneous and isotropic on the average. In this last sentence, the word “average” is the most widely used parameter in quantitative descriptions, but there are many other averages that are used in the earth sciences evaluations. If there is not any specification with this world, then it will imply arithmetic average. Arithmetic average does not attach any weight or priority to any point or direction, i.e., it is an equal-weight average.

5. It can be concluded from the previous points that spatial variations cannot be deterministic in the sense of isotropy and homogeneity, and therefore, they can be considered as nondeterministic which implies uncertainty, and in turn it means that the spatial assessments and calculations cannot be adopted as the final crisp value. At this stage, rather than well-founded deterministic mathematical rules of calculation and evaluation, it is possible to deal with spatial assessments and evaluations by uncertainty calculations, which are probability, statistical, and stochastic processes.
6. Apart from the geometry and material features, the earth sciences event media also includes in a nondeterministic way tectonic effects such as fissures, fractures, faults, folds, or chemical solution cavities, which appear rather randomly. It is a scientific truth that the earth sciences phenomena cannot be studied with deterministic methodologies for meaningful and useful interpretations or applications. The nondeterministic, i.e., uncertainty, techniques such as probability, statistical, and stochastic methodologies are more suitable for the reflection of any spatial behavior.

1.3 Variability

Variability in earth sciences implies irregularities, randomness, and uncertainty, which cannot be predicted with certainty, and always there is a certain level of error involved such as practically acceptable $\pm 5\,\%$ or $\pm 10\,\%$ levels. These levels are also considered as risk amounts in any earth sciences design such as in engineering geology, hydrogeology, and geophysical event evaluation and atmospheric, hydrologic, and environmental scientific and engineering projects. The natural phenomena includes variability with uncertainty component.

Variability is a word that reflects different connotations that are commonly used in everyday life, but unfortunately without noticing its epistemological content. For instance, this word implies inequality, irregularity, heterogeneity, fluctuations, randomness, statistical variability, probability, stochasticity, and chaos. Since science is concerned with materialistic world, variability property can be in time or space (points, lines, areas, and volumes). Uncertainty is concerned with the haphazard variations in nature.

Earth, environmental, and atmospheric phenomena evolve with time and space, and their appreciation as to the occurrence, magnitude, and location is possible by observations and still better by measurements along the time or space reference systems. The basic information about the phenomenon is obtained from the measurements. Any measurement can be considered as the trace of the phenomenon at a given time and location. Hence, any measurement should be specified by time and location, but its magnitude is not at the hand of the researcher. Initial observance of the phenomenon leads to nonnumerical form of descriptive data that cannot be evaluated with uncertainty.

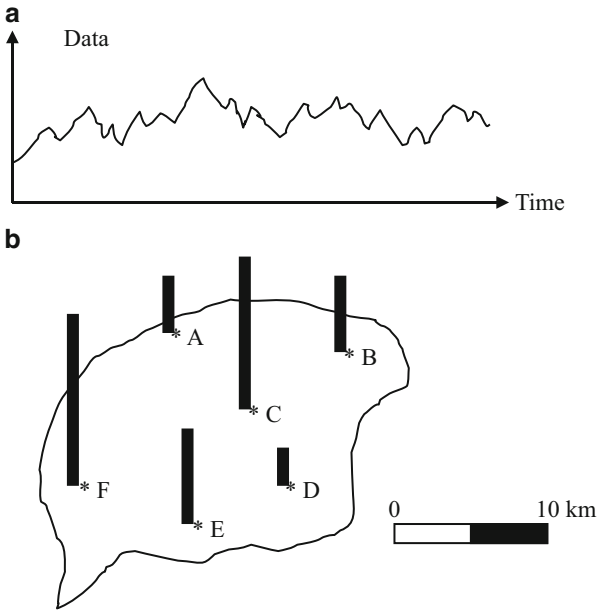


Fig. 1.1 Data variation by (**a**) time, (**b**) space

The worth of data in earth sciences and geology is very high since most of the interpretations and decisions are based on their qualitative and quantitative information contents. This information is hidden in representative field samples which are analyzed for extraction of numerical or descriptive characteristics. These characteristics are referred to as data. Data collection in earth sciences is difficult and expensive and requires special care for accurately representing the geological phenomenon. After all, various parameters necessary for the description and modeling of the geological event, such as bearing capacity, fracture frequency, aperture, orientation, effective strength, porosity, hydraulic conductivity, chemical contents, etc., are hidden within each sample, but they individually represent a specific point in space and time. Hence, it is possible to attach with data temporal and three spatial reference systems as shown in Fig. 1.1.

In geological sciences and applications, the concerned phenomenon can be examined and assessed through the collection of field data and accordingly meaningful solutions can be proposed. It is, therefore, necessary to make the best use of available data from different points. Geological data are collected either directly in the field or field samples are transferred to laboratories in order to make necessary analysis and measurements. For instance, in hydrogeology domain among field measurements are the groundwater table elevations, pH, and total dissolved solution (TDS) readings, whereas some of the laboratory measurements are chemical elements in parts per million (ppm), etc. There are also office calculations that yield also hydrogeological data such as hydraulic conductivity, transmissivity, and storage coefficients. In the meantime, many other data sources from soil surveys, topographic measurements, geological prospection, remote sensing evaluations,

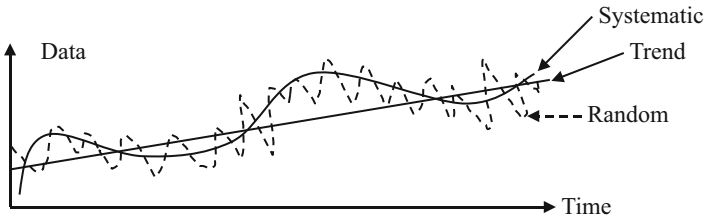


Fig. 1.2 Systematic (trend, seasonality) and unsystematic (random) components

and others may also support data for the investigation concerned. The common property of these measurements and calculations of laboratory analysis is that they include uncertainty attached at a particular time and sampling point only. Hence, the first question is how to deal with rather uncertain (randomly) varying data. At times, the data is random, sometimes, chaotic, and still in other cases irregular or very regular. These changes can be categorized into two broad classes as systematic and unsystematic. Systematic data yields mathematically depictable variations with time, space, or both. For instance, as the depth increases, so does the temperature and this variation is an example of systematic variation. Especially, if there is only one type of geological formation, this systematic variation becomes more pronounced. Otherwise, on the basis of rather systematic variation on the average, there are unsystematic deviations which might be irregular or random. Systematic and unsystematic data components are shown in Fig. 1.2.

In many studies, the systematic changes (seasonality and trend) are referred to as the deterministic components, which are due to systematic natural (geography, astronomy, climatology) factors that are explainable to a certain extent. On the other hand, unsystematic variations are unexplainable or have random parts that need more probabilistic and statistical treatments.

There has been a good deal of discussion and curiosity about the natural event occurrences during the last century. These discussions have included comparisons between uncertainty in earth and atmospheric sciences and uncertainty in physics, which has, inevitably it seems, led to the question of determinism and indeterminism in nature (Leopold and Langbein 1963; Krauskopf 1968; Mann 1970).

At the very core of scientific theories lies the notion of "cause" and "effect" relationships in an absolute certainty in scientific studies. One of the modern philosophers of science, Popper (1957), stated that

> to give a causal explanation of a certain specific event means deducing a statement describing this event from two kinds of premises: from some universal laws and from some singular or specific statements which we may call the specific initial conditions

According to him there must be a very special kind of connections between the premises and the conclusions of a causal explanation, and it must be deductive. In this manner, the conclusion follows necessarily from the premises. Prior to any

mathematical formulation, the premises and the conclusion consist of verbal (linguistic) statements (Ross 1995; Şen 2010). It is necessary to justify at every step of deductive argument by citing a logical rule that is concerned with the relationship among causes and results. On the other hand, the concept of "law" lies at the heart of deductive explanation and, therefore, at the heart of the certainty of our knowledge about specific events.

Recently, the scientific evolution of the methodologies has shown that the more the researchers try to clarify the boundaries of their domain of interest, the more they become blurred with other domains of research. For instance, as the hydrogeologist tries to model the groundwater pollution as one of the modern nuisances of humanity, as far as the water resources are concerned, they need information about the geological environment of the aquifers, meteorological and atmospheric conditions for the groundwater recharge, and social and human settlement environmental issues for the pollution sources. Hence, many common philosophies, logical basic deductions, methodologies, and approaches become common to different disciplines, and uncertainty data processing is among the most important topics, which include the same methodologies applicable to diversity of disciplines. The way that earth, environmental, and atmospheric scientists frame their questions varies enormously, but the solution algorithms may include the same or at least similar procedures.

Any natural phenomenon or its similitude occurs extensively over a region, and therefore, its recordings (measurements) or observations at different locations pose some questions such as, for instance, are there relationships between phenomena in various locations? In such a question, the time is as if it is frozen, and the phenomenon concerned is investigated over the area and its behavioral occurrence between the locations. Frozen time considerations of any earth sciences events expose the spatial variability of the phenomenon concerned. An answer to this question may be provided descriptively in linguistic, subjective, and vague terms, which may be understood even by nonspecialists in the discipline. However, their quantification necessitates objective methodologies, which are one of the purposes of the context in this book. Another question that may be stated right at the beginning of the research in the earth, environmental, and atmospheric sciences is that are places different in terms of the phenomena present there? Such questions provide interest to researchers in the subject of spatial variability.

On the other hand, there are timewise variations, which are referred to as the temporal variations in natural phenomena. For such a variation, it suffices to measure the event at a given location, which is the case in any meteorology station or groundwater and petroleum well.

Any natural event evolves in the 4D human visualization domains, and consequently, its records should involve the characteristics of both time and space variabilities. Any record that has this property is referred to as to have spatiotemporal variation.

1.3.1 Temporal

Most of the natural phenomena and events take place with time, and hence, they leave time traces that are recordable by convenient instruments. Astronomic events have systematic time variations in regular diurnal, monthly, seasonal, and annual time steps, and these recorded traces appear in the form of time series or a set of time measurements. Depending on the time evolution of the event whether it is continuous or not, time series records can be obtained. A time series is the systematic measurement of any natural event along the time axis at regular time intervals. Depending on this time interval, time series is called as hourly, daily, weekly, monthly, or yearly time series. It is not possible to consider space series at regular distances and the records are kept at irregular locations except in very specific cases. For example, if water samples along a river are taken at every 1 km, then the measurements provide a distance series in the regularity sense. Such series are very limited as if there are no such data sequences in practical works.

On the other hand, depending on the interest of event, there are series along time axis but they are not time series due to irregularity or randomness in the time intervals between successive occurrences of the same event. Flood and drought occurrences in hydrology correspond to such cases. One cannot know the duration of floods or droughts. Likewise, in meteorology the occurrence instances of precipitation or any dangerous levels of concentrations such as air pollutants do not have regular time series characteristics.

1.3.2 Point

In earth sciences, records at a set of locations concerning any variable provide information of variability at the fixed point. For instance, at a single point, one can measure different variables, say, for instance, in the engineering geological studies, one can obtain the porosity, specific yield, failure resistance, friction angle, grain sizes, and alike measurements. In hydrogeology, taken water sample from a well may have different anions (Ca, Mg, K, Na) and cations (Cl, SO_4, CO_3, HCO_3), all of which provide specification of water quality collectively at a point. Point records are very important for regional and spatial assessments of earth sciences variable so as to obtain information about the behavior of the same variable at unmeasured sites. The soil specification as for the porosity, shear strength, Atterberg limits (shrinkage plastic and liquid), water content, compression strength, etc. is also among the point measurements that are useful in many geological, engineering, scientific, and agricultural activities.

1.3.3 Regional

Regional studies end up in the form of various 2D maps that provide pointwise information at any desired location. These can be produced from irregularly scattered point records in a region as contour lines that show equal value variables, polynomials that partition the region into a set of equal value regions or in the computer at a set of pixels. For the map construction over a region, there are different methodologies to convert the irregularly scattered records to a set of regular grid points, which are the basis prior to any application of the mapping procedures. Among the mapping procedures, there are different approaches such as the inverse distance, inverse distance square, regional geometric functions, and Kriging methodologies as explained in Chaps. 6 and 7.

1.3.4 Spatial

This is the main topic of the book and it is the three-dimensional, 3D, representation of the earth sciences phenomena. Such representations are valid for fence diagrams in geophysical prospection of the geological formations in any study area. This is the 3D representation of the contour maps or they can be obtained from digital elevation model (DEM) data for any region in the form of geomorphological (topographic) maps.

1.4 Determinism

This is not valid in natural earth sciences phenomena, because it denies the uncertainty involvement in natural earth sciences events where variability, irregularity, haphazardness, uncertainty, chaos, and any other type of uncertainty ingredient take place. Astronomic events are rather deterministic, but their effects on the earth sciences phenomena such as geology, earthquake, hydrology, meteorology, hydrogeology, and tsunami are all probabilistic or stochastic.

Deterministic phenomena are those in which outcomes of the individual events are predictable with complete certainty under any given set of circumstances, if the required initial conditions are known. In all the physical and astronomical sciences, traditionally deterministic nature of the phenomena is assumed. It is, therefore, necessary in the use of such approaches the validity of the assumption sets and initial conditions. In a way, with idealization concepts, assumptions, and simplifications, deterministic scientific researches yield conclusions in the forms of algorithms, procedures, or mathematical formulations which should be used with caution for restrictive circumstances. The very essence of determinism is the idealization and assumptions so that uncertain phenomenon becomes graspable

and conceivable to work with the available physical concepts and mathematical procedures. In a way, idealization and assumption sets render uncertain phenomenon into conceptually certain situation by trashing out the uncertainty components. A significant question that may be asked at this point is that, is there not any benefit from the deterministic approaches in the earth and atmospheric studies where the events are uncertain? The answer to this question is affirmative, because in spite of the simplifying assumptions and idealizations, the skeleton of the uncertain phenomenon is captured by the deterministic methods.

1.5 Uncertainty

Mann (1970) stated that the inability to predict specific events may stem from the fact that nature is intrinsically random. He defined random and randomness in a statistical sense to describe any phenomenon, which is unpredictable with some degree of uncertainty. On the other hand, deterministic phenomenon is predictable under a set of initial and boundary conditions. In general, the earth sciences events are considered as random. Consequently, randomness has been suggested as the ultimate and the most profound physical concept in earth sciences. In an intrinsically random phenomenon, exact predictions are rather impossible.

Uncertainty is introduced into any problem through the variation inherent in nature, through man's lack of understanding of all the causes and effects in physical systems and in practice through insufficient data. Even with a long history of data, one cannot predict the natural phenomenon except within an error band. As a result of uncertainties, the future can never be predicted completely by researchers. Consequently, the researcher must consider the uncertainty methods for the assessment of the occurrences and amounts of particular events and then try to determine their likelihood of occurrences.

The uncertainty in the geologic knowledge arises out of the conviction that earth generalizations are immensely complicated interactions of abstract, and often universal, physical laws. Earth sciences generalizations always contain the assumptions of boundary and initial conditions. In a way, the uncertainty in the predictions arises from the ignorance of the researcher to know the initial and boundary conditions in exactness. They cannot control these conditions with certainty. On the assumptions of physical theory, earth and atmospherically significant configurations are regarded as highly complex. This is true whether or not the "world" is deterministic. Physical laws, which are not formulated as universal statements, may impose uncertainty directly upon earth sciences events as in the case of inferences based on the principles of radioactive disintegration.

There has been a good deal of discussion and curiosity about the natural event occurrences during the last century. These discussions have included

comparisons between uncertainty in earth and atmospheric sciences and uncertainty in physics which has, inevitably it seems, led to the question of determinism and indeterminism in nature (Leopold and Langbein 1963; Krauskopf 1968; Mann 1970).

The uses of uncertainty techniques such as the probability, statistical, and stochastic methods in earth sciences have increased rapidly since the 1960s, and most of the researchers, as students and teachers, seek more training in these disciplines for dealing with uncertainty in a better quantitative way. In many professional journals, book and technical reports in the earth sciences studies include significant parts on the uncertainty techniques in dealing with uncertain natural phenomena. And yet relatively few scientists and engineers in these disciplines have a strong background in school mathematics, and the question is then how can they obtain sufficient knowledge of uncertainty methods including probability, statistical, and stochastic processes in describing natural phenomena and in appreciating the arguments which they must read and then digest for successful applications in making predictions and interpretations.

1.5.1 Probabilistic

The causality means that a result will proceed mechanistically from specific initial conditions and a triggering action or event (in the form of cause and effect through an environment). The causes of drilling a well are associated with two outcomes as wet and dry each with the same possibility index. The possibility of each drill to end up in "wet" phase is one-half and "dry" is also the same. Probability does not characterize a relationship among events, as does a statistical relationship, but rather a relationship among statements. In the probability approach, a probable inference will include as one of its premises a frequency hypothesis and an assertion that if conditions of a specified kind are realized, then conditions of another kind will be realized with a certain frequency in the long run. Herein, the conclusion cannot follow deductively from the premises, or the truth of the premises does not guarantee the truth of the conclusion. Hence, any probable inference is not simply an inference that yields an uncertain conclusion, but one that permits a certain degree of probability or rational credibility to be assigned to the conclusion. Since the earth sciences data have uncertainties, they appear according to probabilistic principles. In practical terms, probability is a percent value that is commonly used between people almost every day. Among the probabilistic questions are the oil hit possibilities by a drill, earthquake occurrence rate within a given duration, fracture percentage along a scanline on the rock outcrop, relative frequency of rainfall, etc. None of these questions can be answered with certainty, but through the probabilistic principles. It is, therefore, very significant that earth scientist should have some background about the probabilistic concepts.

One of the most frequently used concepts in the probability calculations is the event. It is defined as the collection of uncertain outcomes in the forms of sets, class,

or groups. Events might be elementary or compound depending on the decomposition. A compound event can be decomposed into at least two events, whereas an elementary event cannot be decomposed. For instance, if one things on the rainy and non-rainy days sequence then yesterday was non-rainy corresponds to an elementary event, but non-rainy days in the last 10 days is a compound event. In general, what is defined as elementary or compound depends on the problem under consideration and the purpose for which an analysis is being conducted. Occurrence of precipitation in any day may include different events such as the rainfall or snow or hail, and therefore, it is a compound event composed of these elementary events. If one says that the rainfall occurs or does not occur, these are complementary but elementary events. It is also possible to be interested in the amount of rainfall in addition to its occurrence; this is then a compound event with two parts, namely, the occurrence and the amount. For example, the flood disaster is a compound event that damage human life and/or property. In the probability calculations, the researcher prefers to decompose a compound event into underlying elementary events.

Although the probability is equivalent to daily usage of percentages, the problem in practice is how to define this percentage. For the probability definition, it is necessary to have different categories for the same data. In any probability study, the basic question is what is the percentage of data belonging to a given category? The answer to this question is explained in Chap. 2.

1.5.2 Statistical

Statistics is concerned with a set of parameters of a given data set and also any model in the form of mathematical expressions and their parameter values. Statistics is the branch of mathematics, where reliable and significant relationships are sought among different causative variables and the consequent variable.

Random and randomness are the two terms that are used in a statistical sense to describe any phenomenon, which is unpredictable with any degree of certainty for a specific event. An illuminating definition of random is provided by famous statistician Parzen (1960) as

> A random (or chance) phenomenon is an empirical phenomenon characterized by the property that its observation under a given set of circumstances does not always lead to the same observed outcome (so that there is no deterministic regularity) but rather to different outcomes in such a way that there is a statistical regularity.

The statistical regularity implies group and subgroup behaviors of a large number of observations so that the predictions can be made for each group more accurately than the individual predictions. For instance, provided that a long sequence of temperature observations is available at a location, it is then possible to say more confidently that the weather will be warm or cool or cold or hot

tomorrow than specifying exactly by predicting the degree of centigrade. As will be explained in later sections, the statistical regularities are a result of some astronomical, natural, environmental, and social effects. The global climate change discussions are based on the fossil fuel pollution of the lower atmospheric layers due to anthropogenic activities. The climate change effect is expressed by different researchers and even common men, but the intensity of such a change cannot be determined with certainty over the coming time epochs. Statistical regularity implies further complete unpredictability for single or individual events.

1.5.3 Stochastic

Stochastic processes are related to internal structure and behavior and natural or artificial events after probabilistic and statistical feature identifications. It is concerned with the serial dependence of a single time series or cross-dependence between any two time series. A detailed account of time series analysis and prediction is given in a classical book by Davis (1986) and Box and Jenkins (1970). Any natural, social, and economic records taken at a set of regular time or space intervals with uncertainty component are subject of stochastic processes for identification of the underlying generation mechanism and then for simulation and at times projections over future times or extrapolation on unmeasured spatial domains (Davis, J.C. 1986. Statistics and Data Analysis in Geology. John Wiley and Sons).

These are hybrid models in the sense that both statistical and probabilistic features of the historical data sequence are incorporated in the synthetic sequence generation. They have explicit mathematical expressions where the variable prediction at any time is considered as a function of some previous time measurements in a deterministic manner with a random component. Hence, it is possible to write all the stochastic processes mathematically as follows.

$$X_t = f_D(X_{t-1}, X_{t-2}, X_{t-3}, \ldots X_{t-\mathrm{k}}) + f_R(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \varepsilon_{t-m}) \qquad (1.1)$$

where the first term shows the deterministic component and the second term is for the random components. In the deterministic part the number of previous data points, m, is referred to as the deterministic lag. The second function on the right-hand side is the random component with random lag as m. In general, this second term is also referred to as the error term. It is possible to suggest different stochastic models from the implicit function in Eq. 1.1. In practice, almost all the classical stochastic processes have the explicit form of this last expression through linear term summations as will be explained in the following.

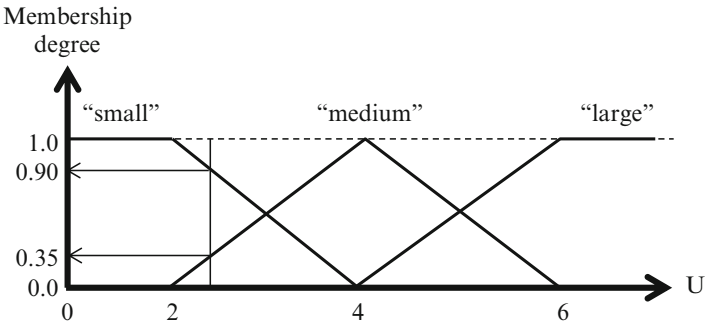


Fig. 1.3 Fuzzy set and membership degree

1.5.4 Fuzzy

The concept of "fuzzy sets" was introduced by Zadeh (1974) who pioneered the development of fuzzy logic instead of Aristotelian logic of two possibilities only. Unfortunately, this concept was not welcome into the literature since many uncertainty techniques such as the probability theory, statistical, and stochastic processes were commonly employed at that time but fuzzy logic has been developing since then and now being used especially in Japan for automatic control for commercially available products such as washing machines, cameras, and robotics. Many textbooks provide basic information on the concepts and operational fuzzy algorithms (Kaufmann and Gupta (1988, 1991), Klir and Folger (1988), Klir and Yuan (1995), Ross (1995), and Şen (2010)). The key idea in the fuzzy logic is the allowance of partial belonging of any object to different subsets of the universal set instead of belonging to a single set completely. Partial belonging to a set can be described numerically by a membership function which assumes numbers between 0 and 1.0 inclusive. For instance, Fig. 1.3 shows typical membership functions for small, medium, and large class sizes in a universe, U. Hence, these verbal assignments are the fuzzy subsets of the universal set.

Fuzzy membership functions may be in many forms but in practical applications simple straight-line functions are preferable like triangles and trapeziums. Especially, triangular functions with equal base widths are the simplest possible ones. For instance, Figure 1.3 shows the whole universe, U, space, which is subdivided into threesubsets with verbal attachments "small", "medium", and "large".

In this figure, set values with less than 2 are definitely "small"; those between 2 and 6 are "medium"; and values more than 6 are definitely large. However, intermediate values such as 2.2 and 3.5 are in between, that is, partially belong to subsets "small" and "medium." In fuzzy terminology, 2.2 has membership degree of 0.90 in "small" and 0.35 in "medium" but 0.0 in "large."

The main purpose of this approach is to replace "crisp" and "hard" objectives in many problem solving procedures by fuzzy ones. Unlike the usual constrain where, say, the variable in Fig. 1.3 must not exceed 2, a fuzzy constrain takes the form as

saying that the same variable should preferably be less than 2 and certainly should not exceed 4. This is tantamount in fuzzy sets term that the values less than 2 have membership of 1.0 but values greater than 4 have membership of 0.0 and values between 2 and 6 would have membership between 1.0 and 0.0. In order to make the subsequent calculations easy, usually the membership function is adopted as linear in practical applications. The objective then can be formulated as maximizing the minimum membership value, which has the effect of balancing the degree to which the objective is attained with degrees to which the constraints have to be relaxed from their optimal values.

1.5.5 Chaotic Uncertainty

The most important affair in any scientific study is the future prediction of the phenomenon concerned after the establishment of a reliable model. Although the prediction is concerned with the future unknown states, the establishment stage depends entirely on the past observations as numerical records of the system variables. In any study, the principal step is to identify the suitable model and modify it so as to represent the past observation sequences and general behaviors as closely as possible. Not all the developed models are successful in the practical applications. In many areas of scientific predictions and especially in the meteorology domain, the researchers are still far behind the identification of a suitable model. It is possible by physical principles in addition to various simplifying assumptions to describe many systems of empirical phenomenon by ordinary and partial differential equations, but their application, for instance, for prediction purposes requires the measurements and identification of initial and boundary conditions in space as well as time. Unfortunately, in most of the cases, the research is not furnished with reasonably sufficient data. In the cases of either the availability of unrealistic model or paucity of data, it is necessary to resort to some other simple but effective approaches. In many cases, the derivation of the partial differential equations for the representation of the concerned system is straightforward, but its implications are hindered due to either the incompatibility of the data with this model or the lack of sufficient data.

In many applications rather than the partial differential equations and basic physical principles, a simple but purpose-serving model is identified directly from the data. For instance, Yule (1927) has proposed such simple models by taking into account the sequential structural behavior of the available data. His purpose has been to treat the data on the basis of the stochastic process principles as a time series. Such a series was then regarded as one of the possible realizations among many other possibilities which are not known to the researcher. Recently, chaotic behaviors of dynamic systems also exhibited random-like behaviors which are rather different from the classical randomness in the stochastic processes. Hence, a question emerges as to how to distinguish between a chaotic and stochastic behavior. Although the chaotic behavior shows a fundamentally long-term pattern

in the form of strange attractors, it also suggests short-term prediction possibility. Any data in the form of time series might look like a random sequence, but it may include hidden short-term consistencies with few degrees of freedom. For instance, although fluid flows represent chaotic behaviors along the time axis, but when confined to time-independent state-space representation, it indicates an attractor of lower dimensions. In the case of such a strange attractor existence, its time evolution results in a time series, which hides the chaotic remnants. It is the main purpose of the chaos theory and methodology to try and identify such remnant chaotic behaviors in the time series by considering sequential correlation dimension values that are completely time independent. The classical serial correlation function assesses the time series in the time domain, but the serial correlation dimension investigates the given time series in the phase domain independent of the time.

A dynamic system is identified by a phase-space diagram whose trajectories define its evolution starting from an initial state. In order to enter this trajectory, it is necessary to know the initial state rather precisely. Different initial states even the ones that are of minute difference from each other enter the strange attractor and after the tremendous successive steps cover the whole strange attractor by chaotic transitions. Hence, another digression from the classical time series is that in the time series, the successive time steps are equal to each other but in the chaotic behaviors the steps are random but successive points remain on the same attractor. In the chaotic behaviors, it is not sufficient to identify the strange attractor completely, but for predictions what is necessary is the modeling of successive jumps that will remain on the attractor. The trajectories approach to a sub-space within the whole geometric pattern, and hence, the strange attractor can be captured independently from the initial conditions. Otherwise, the behaviors are not chaotic but stochastic with no attractor. In short strange attractors are made up of the points that cover a small percentage of the phase space but stochastic or completely random behaviors cover the whole phase space uniformly. Sub-coverage implies that the dimension of the chaotic time series is less than the embedding dimension of the phase space. The systems with strange attractors are dissipative in that the energy is not conserved.

If a system is deterministic, then its attractors have integer dimensions and they provide reliable long-term predictions. However, if the dynamic system is very sensitive to initial conditions, then its attractor has fractal dimension and therefore it is called as the strange attractor. For systems with strange attractors, long-term predictions are not reliable at all. Such chaotic dynamic systems should be exploited for short-term predictions with reliable mathematical tools by Grassberger and Procaccia (1983a, b). Detailed account of the chaotic dynamic systems is outside the scope of this book.

1.6 Random Field (RF)

Earth sciences events evolve by time but they have a regional nature and cover extensive irregular areas. Their monitoring, data collection, and assessment require special probabilistic, statistical, and stochastic approaches. One of the most well-known methodologies is the geostatistical procedures, where the regional variability is related to the distance along a preferred direction. Regional phenomenon can be identified by studying the spatial patterns at a set of sites (points). They are often based on single-site event definitions where the areal aspect is included by studying the spatial pattern of point values. The irregular measurement points are overlain by a mesh at which grid points, the assessment, and modeling fundamentals can be activated. Such a set of measurement points and grid is given in Fig. 1.4.

Earth sciences variables evolve discretely or continuously in space and time. Each of these variables can be sampled at a single site in space at a given time instant or interval. Such a space-time distribution is referred to as a spatiotemporal variability, and they cannot be predicted with certainty, and therefore, they are assumed random, and hence, it is necessary to study a new class of fields, namely, random fields (RFs), which are defined mathematically by a quadruple function $\xi(x, y, z, t)$, at the site with coordinates x, y, and z and time instant, t.

In general, RF is a generalization of a stochastic process, for which the random function of the coordinates (x, y, z, t) must be understood at each spatiotemporal point (x, y, z, t) as having a random value $\xi(x, y, z, t)$ that cannot be predicted exactly, but these values are subject to a certain probability distribution function (PDF). Hence, the complete description of a RF can be represented by finite-dimensional PDFs of the field at different locations. However, in practice, rather than the PDFs, statistical moments (parameters) are found useful in their assessments. In general, a RF has three types of moments (Şen 1980a):

(i) Space moments, which are the time products of the values of the field at different points at a fixed time
(ii) Time moments, which are the mean product of the values of the field at different times at a fixed point

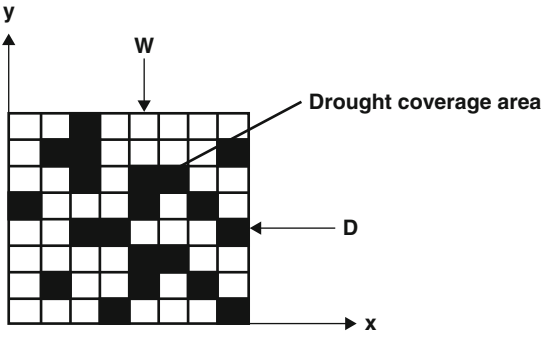


Fig. 1.4 Regional modeling set up

(iii) Space-time moments, which are the mean products from the values of the field at different points and times

A RF is homogeneous, if its PDF is invariant with respect to a shift in the system of points. Additionally, the RF is also statistically isotropic if the PDFs are invariant with respect to an arbitrary rotation of the system of points such as a solid body and to a mirror reflection of this system with respect to the arbitrary plane passing through the origin of the coordinate system. This implies that the statistical moments depend upon the configuration of the system of points for which they are formed, but not upon the position of the system in space. In case of homogeneous and isotropic RF, the correlation function depends only on the distance between the two points, which join them but not on the orientation of the line (Yevjevich and Karplus 1974; Şen 1980b, 2009).

For regional earth sciences event analysis, the space components of the field are assumed to remain in a certain area. Furthermore, in practice, information about the RF can be sampled at a finite number of sites within the drainage basin. With fixed coordinates of sites relative to a reference system, the sample hydrological phenomena constitute a multivariate stochastic process.

It is a generalization of a stochastic process such that the process parameters are not fixed values but can take a multitude of values. Their values are scattered in 2D or 3D space, and the values are spatially correlated in some way or another. In general, the closer the locations, the closer are the values.

References

Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden Day, San Francisco, p 475

Davis JC (1986) Statistic and data analysis in geology. John Wiley & Sons, Inc., New York, p 560

Kaufmann A, Gupta MM (1988) Fuzzy mathematical models in engineering and management science. North-Holland, Amsterdam

Grassberger P, Procaccia I (1983a) Physica, D 9, 198. Phys Rev Lett 50:346

Grassberger P, Procaccia I (1983b) Phys Rev A 28:2591

Johnston RJ (1989) Philosophy, ideology and geography. In: Gregory D, Walford R (eds) Horizons in human geography. MacMillan, Basingstoke, pp 48–66

Kaufmann A, Gupta MM (1991) Introduction to fuzzy arithmetic theory and application. Van Nostrand Reinhold, New York

Klir GJ, Folger TA (1988) Fuzzy sets uncertainty, and information. Prentice Hall, Englewood Cliffs

Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic theory and applications. Prentice Hall, Upper Saddle River

Krauskopf KB (1968) A tale of ten plutons. Geol Soc Am Bull 79:1–18

Leopold LB, Langbein WB (1963) Association and indeterminacy in geomorphology. In: Albritton CC Jr (ed) The fabric of geology. Addison Wesley, Reading, pp 184–192

Mann CJ (1970) Randomness in nature. Bull Geol Soc Am 81:95–104

Parzen E (1960) Probability theory, vol 2. Wan Nostrand, Princeton

Popper K (1957) Philosophy of science: a personal report. In: Mace CA (ed) British philosophy in the mid-century. Allen and Unwin, London

Ross TJ (1995) Fuzzy logic with engineering applications. Mc-Graw-Hill Book Co., New York: 600 p

Şen Z (1980a) Statistical analysis of hydrologic critical droughts. J Hydraul Eng Div, ASCE 106:99–115

Şen Z (1980b) Regional drought and flood frequency analysis. J Hydrol 46:258–263

Şen Z (2009) Spatial modeling principles in earth sciences, 1st edn. Springer, New York, p 351

Şen Z (2010) Fuzzy logic and hydrological modeling. CRC Press, Taylor and Francis Group, Boca Raton: 340 p

Yevjevich V, Karplus AK (1974) Area-time structure of the monthly precipitation process, Hydrology Paper No. 64. Colorado State University, Fort Collins

Yule G (1927) On a method of investigating periodicities in disturbed series, with special application to Wolfert's sun spot numbers. Phil Trans R Soc A 226:267–298

Zadeh LA (1974) A fuzzy-algorithmic approach to the definition of complex or imprecise concepts, ERL Memo M474. Univ. of Calif, Berkeley

Chapter 2
Sampling and Deterministic Modeling Methods

Abstract Prior to any earth sciences modeling, the basic data quality, reliability, and sufficiency are the basic questions that must be grasped by the specialists in the earth sciences domain so that she/he can continue for better description and modeling of the phenomenon concerned. For this purpose, field, laboratory, or satellite observations and measurements are basic foundation lines toward a successful problem solution. Sampling categorization of the data and the properties is important in any data treatment work, especially in earth sciences domain. The internal structure of data set as dependent of independent, the sample length, and their random distribution behavior by means of a theoretical probability distribution function are to be evaluated objectively in any study. In earth sciences, often irregularly located spatial data are available, and therefore, the representative area or (area of influence) should be defined for each measurement location. For this purpose, this chapter presents different spatial methodologies including regionalization, inverse distance methods, triangularization, polygonizations, areal coverage probability, regional extreme value probabilities, and spatio-temporal modeling.

Keywords Droughts • Observation • Measurement • Numerical data • Sampling • Small sample • Triangularization • Polygonization • Percentage polygon

2.1 General

Scientific and engineering solutions can be given about any earth sciences phenomena through relevant spatial modeling techniques provided that representative data are available. However, in many cases, it is difficult and expensive to collect the field data, and therefore, it is necessary to make the best use of available linguistic information, knowledge, and numerical data to estimate the spatial (regional) behavior of the event with relevant parameter estimations and suitable models. Available data provide numerical information at a set of finite points, but the professional must fill in the gaps using information, knowledge, and understanding about the phenomena with expert views. Data are the treasure of knowledge and information leading to meaningful interpretations.

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_2

Observations are also potential source of information, which provides linguistic rational and logical expressions in the form of premises. Data imply numerical measurements using different instruments either in the field or laboratory. Observations are not numerical but rather verbal data that assist to describe and identify the phenomenon concerned.

The development of data estimation methods can be traced back to Gauss (1809), who suggested the technique of deterministic least-squares approach and employed it in a relatively simple orbit measurement problem. The next significant contribution to the extensive subject of estimation theory occurred more than 100 years later when Fisher (1912), working with PDF, introduced the approach of maximum likelihood estimation. However, Wiener (1942, 1949) set forth a procedure for the frequency domain design of statistically optimal filters. The technique addressed the continuous-time problem in terms of correlation functions and the continuous filter impulse response. Moreover, the Wiener solution does not lend itself very well to the corresponding discrete data problem, nor it is easily extended to more complicated time-variable, multiple-input/output problems. It was limited to statistically stationary processes and provided optimal estimates only in the steady-state regime. In the same time period, Kolmogorov (1941) treated the discrete-time problem.

In this chapter, observation and data types are explained, and their preliminary simple logical treatments for useful spatial information deductions are presented and applied through examples.

2.2 Observations

They provide information on the phenomenon through sense organs, which cannot provide numerical measurements but their expressions are linguistic (verbal) descriptions. In any study, the collection of such information is unavoidable, and they are very precious in the construction of conceptions and models for the control of the phenomenon concerned. Observations may be expressed rather subjectively by different persons but experts may deduce the best set of verbal information. Depending on the personal experience and background, an observation may instigate different conceptualization and impression on each person. In a way observations provide subjective information about the behavior of the phenomenon concerned. In some branches of scientific applications, observational descriptions are the only source of data that help for future predictions. Even though observations may be achieved through some instruments as long as their description remains in verbal terms, they are not numerical data. Observations were very significant in the early developments of the scientific and technological developments especially before the seventeenth century, but they became more important in modern times including linguistic implications and logical deductions explaining the fundamentals of any natural or man-made event (Zadeh 1965). For instance, in general, geological description of rocks can be

made by field observations, and concise linguistic categorizations are then planned for others to understand again linguistically. It is important to stress at this point that linguistic expressions of observations help to categorize the event. Although such a categorization set forward crisp and mutually exclusive classes according to classical Aristotelian logic, recently, fuzzy logic classification including mutually inclusive classes is suggested by Zadeh (1973) and it is most frequently used in every discipline in an increasing rate. In many disciplines observations are extremely important such as in geological sciences, medicine, social studies, physiology, military movements, economics, social sciences, meteorology, engineering, etc. At times they are more valuable than numerical data, but unfortunately, their role is almost forgotten due to recent mechanical and software programs that work with numerical data.

Example 2.1 What type of observational information can be obtained when one takes hand specimen from a rock? Even a nonspecialist in geology tries to deduce the following basic linguistic information based on his/her observation and inspection of the specimen through a combined use of his/her sense organs.

1. Shape: Regular, irregular, round, spiky, elongated, flat, etc. This information can be supported for detailed knowledge with addition of adjectives, such as "rather," "quite," "extremely," "moderately," and so on. Note that these words imply fuzzy information.
2. Color: Any color can be attached to whole specimen or different colors for different parts. Detailed information can be provided again by fuzzy adjectives such as "open," "dark," "gray," etc.
3. Texture: The words for the expression of this feature are "porous," "fissured," "fractured," "sandy," "gravelly," "silty," etc.
4. Taste: The previous descriptions are through the eye but the tongue can also provide information as "saline," "sour," "sweet," "brackish," and so on.
5. Weight: It is possible to judge approximate weight of the specimen and have description feelings as "light," "heavy," "medium," "very heavy," and "floatable," and likewise other descriptions can also be specified.
6. Hardness: The relative hardness of two minerals is defined by scratching each with the other and seeing which one is gouged. It is defined by an arbitrary scale of ten standard minerals, arranged in Mohr's scale of hardness and subjectively numbered in scale based on degrees of increasing hardness from 1 to 10. The hardness scale provides guidance for the classification of the hand specimen according to Table 2.1, where the verbal information is converted to a scale through numbers.

Complicated events cannot be quantified through numbers and this leaves the only way of verbal description as a result of visual observations. It is possible to prepare such informational scales according to expert views.

Example 2.2 Earthquake effect on structures can be described according to Table 2.2 guidance, which is given by Mercalli (1912). The following is an abbreviated description of the 12 scales of Modified Mercalli Intensity.

Table 2.1 Mohr's scale of hardness

Hardness scale	Mineral	Example
1	Talc	Hydrated magnesium silicate
2	Gypsum	Hydrated calcium sulfate
3	Calcite	Calcium carbonate
4	Fluorspar	Calcium fluoride
5	Apatite	Calcium phosphate
6	Feldspar	Alkali silicate
7	Quartz	Silicate
8	Topaz	Aluminum silicate
9	Corundum	Alumina
10	Diamond	Carbon

Table 2.2 Mercalli earthquake intensity scales

Scales	Description
I	Not felt except by a very few under especially favorable conditions
II	Felt only by a few persons at rest, especially on upper floors of buildings. Delicately suspended objects may swing
III	Felt quite noticeably by persons indoors, especially on upper floors of buildings. Many people do not recognize it as an earthquake. Standing motor cars may rock slightly. Vibration similar to the passing of a truck. Duration estimated
IV	Felt indoors by many, outdoors by few during the day. At night, some awakened. Dishes, windows, doors disturbed; walls make cracking sound. Sensation like heavy truck striking building. Standing motor cars rocked noticeably
V	Felt by nearly everyone; many awakened. Some dishes, windows broken. Unstable objects overturned. Pendulum clocks may stop
VI	Felt by all, many frightened. Some heavy furniture moved; a few instances of fallen plaster. Damage slight
VII	Damage negligible in buildings of good design and construction; slight to moderate in well-built ordinary structures; considerable damage in poorly built or badly designed structures; some chimneys broken
VIII	Damage slight in specially designed structures; considerable damage in ordinary substantial buildings with partial collapse. Damage great in poorly built structures. Fall of chimneys, factory stacks, columns, monuments, walls. Heavy furniture overturned
IX	Damage considerable in specially designed structures; well-designed frame structures thrown out of plumb. Damage great in substantial buildings, with partial collapse. Buildings shifted off foundations
X	Some well-built wooden structures destroyed; most masonry and frame structures destroyed with foundations. Rail bent
XI	Few, if any (masonry) structures remain standing. Bridges destroyed. Rails bent greatly
XII	Damage total. Lines of sight and level are distorted. Objects thrown into the air

It is important to notice that the linguistic descriptions and scales are neither time nor space dependent but they have event basis. The reference to any system is not required apart from the logical rules.

2.3 Sampling

The very significant prospect of sampling is the distribution of measurement sites within the study area. Depending on the prevailing conditions, sometimes the scatter of sampling points is already set up due to previous human activities such as water well locations, oil drillings, settlement areas, roads, etc. However, in detailed studies at smaller scales, the researchers have to lay down the set of points so as to sample the concerned phenomenon in a representative manner. There are different techniques in deciding about the position of the sampling points. If nothing is known beforehand, then it may seem logical to select the sampling points at nodes or centers of a suitable mesh over the study area. This is the most uniform sampling procedure as shown in Fig. 2.1.

These sampling patterns can be divided into three different categories as the regular, random, and aggregated or clustered. Figure 2.1a, b is of regular sampling procedures. In Fig. 2.1c, d, the randomness is in small scales and the random pattern remains within the subareal regular grids. In Fig. 2.1b although each one of the subarea is sampled, in Fig. 2.1d, only randomly chosen subareas are sampled. In Fig. 2.1a, the maximum distance between the two neighboring points cannot be greater than the twice of the subareal diagonal length. In Fig. 2.1c, the distance is of several times the main diagonal length of the subarea. Large-scale random sampling patterns are given in Fig. 2.1e, where there is no restriction on the distance between the sampling points. In Fig. 2.1f, there are three clusters of the spatial sampling each with random sampling patterns. Such categorized samplings are possible depending on the areal occurrence of the phenomenon studied. For instance, if ore, water, or oil deposits are intact from each other at three neighboring areas, then the cluster sampling patterns arise.

Another feature of spatial sampling points is its uniformity concerning the frequency of occurrence per area. If the density defined in this manner is equal in each subarea, then the spatial sampling is uniform; otherwise, it is nonuniform. This definition implies that the regular sampling points in Fig. 2.1a, b in addition to small-scale random pattern in Fig. 2.1c are all uniform, because there is one point per subarea.

Uniformity gains significance if there are many sampling points within each subarea. In geological or meteorological studies, subareas are quadrangles between two successive longitudes and latitudes. For instance, such a situation is shown in Fig. 2.2 where each quadrangle has random number of random sampling points. Hence, the question is whether the sampling point distribution is uniform or not.

The pixels present a regular mesh over a region, which is conceptually similar to numerical solution of analytical models. The difference is that in the case of pixels, the measurements (brightness values) are known, but this is not the case in numerical analysis, where the values either at the center of each cell or the values at each node are necessary for calculation.

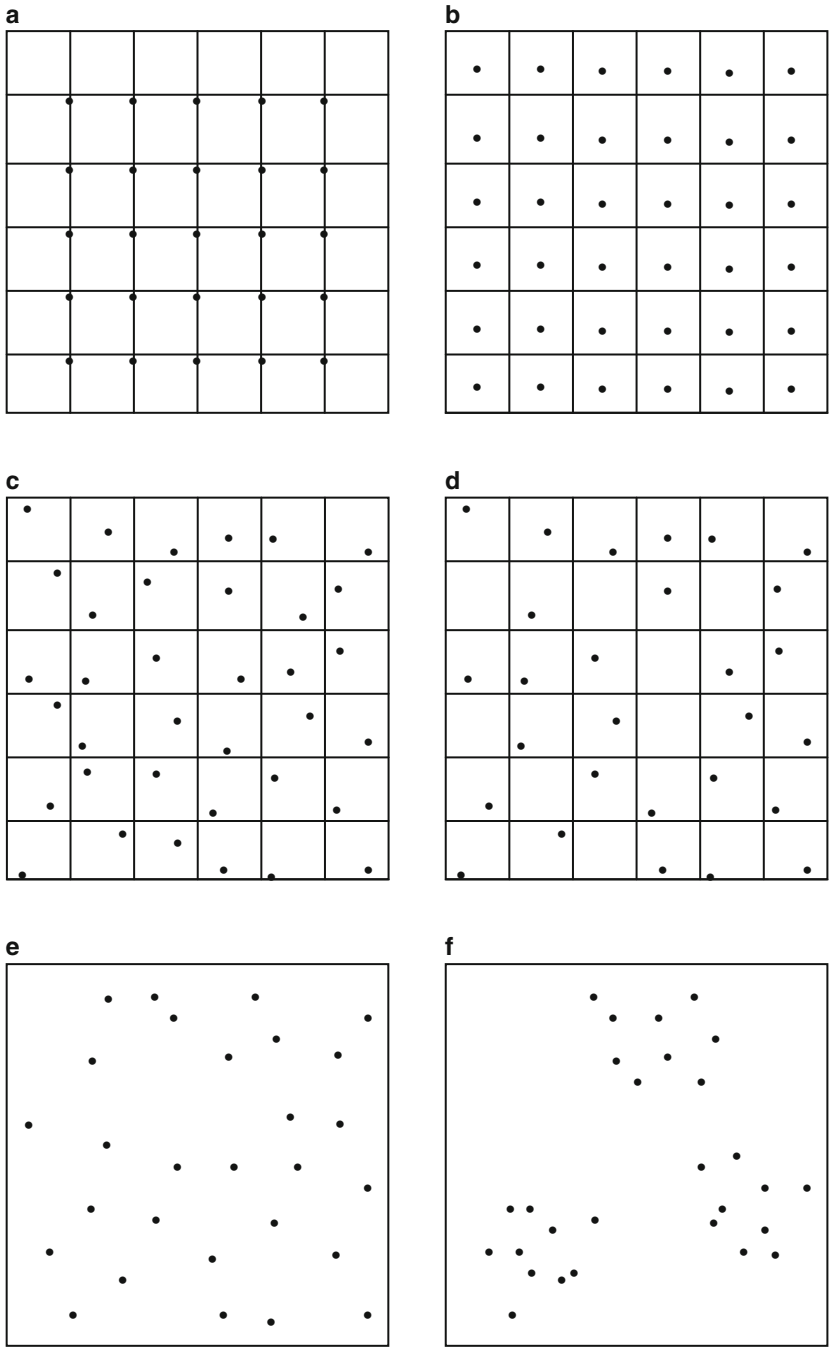


Fig. 2.1 Areal sampling patterns

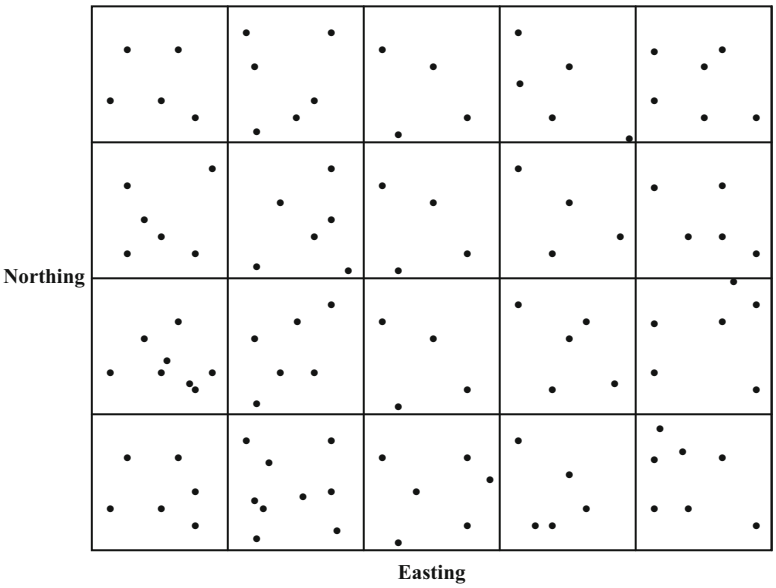


Fig. 2.2 Quadrangle sampling

A set of point data from a region on interested event such as the elevations at 25 different sites can be presented in three columns, two of which are location descriptions and the third column includes the regional variable, which is elevation in this case (see Table 2.3). This table is an example for triple values as mentioned before.

In this table, the scatter of sample points can be obtained from the second and third columns, which appear as in Fig. 2.3. The first appearance indicates that the sampling points are irregularly distributed in the study area. However, it is not possible to say whether their distribution is random or there is a systematic correlation. This question will be answered in Chap. 3.

This may not be a representative sampling set for the study area because there are no sampling points in the upper right and lower left parts of the area.

Although the spatial data is understood as the sampling of a variable with respect to longitudes and latitudes, it is also possible to consider another two variables instead of longitude and latitude. For instance, similar to Table 2.3 in Table 2.4, calcium, magnesium, and chloride variables are given in three columns. These values are taken at a single point in space at a set of time instances.

In this case, there are three alternatives for data treatment purposes. The following similar questions can be asked:

1. Does one want the change of chloride with respect to calcium and sodium?
2. Does one want the change of calcium with respect to chloride and sodium?
3. Does one want the change of sodium with respect to calcium and chloride?

Table 2.3 Elevation data

Points	Easting	Northing	Elevation (m)	Points	Easting	Northing	Elevation (m)
1	53,4371.3	2,449,324	60	14	569,732.8	2,353,515	750
2	589,853.8	2,362,837	116	15	620,966.6	2,362,115	570
3	539,568.8	2,432,734	570	16	612,983.4	2,391,577	110
4	605,430.4	2,359,240	228	17	596,556.5	2,397,926	700
5	626,244	2,350,163	710	18	619,828.7	2,399,006	60
6	619,177.6	2,370,404	60	19	527,607.4	2,385,670	710
7	570,437.2	2,454,979	520	20	658,855.3	2,376,279	1,240
8	560,157.4	2,447,557	280	21	520,716.1	2,377,359	540
9	563,857.7	2,453,414	11	22	655,595.2	2,355,951	280
10	615,788.4	2,361,155	660	23	646,632.6	2,390,925	125
11	543,309.8	2,427,824	720	24	650,305.8	2,366,972	90
12	572,452.4	2,389,806	910	25	641,826.4	2,348,446	350
13	604,572	2,358,312	11				

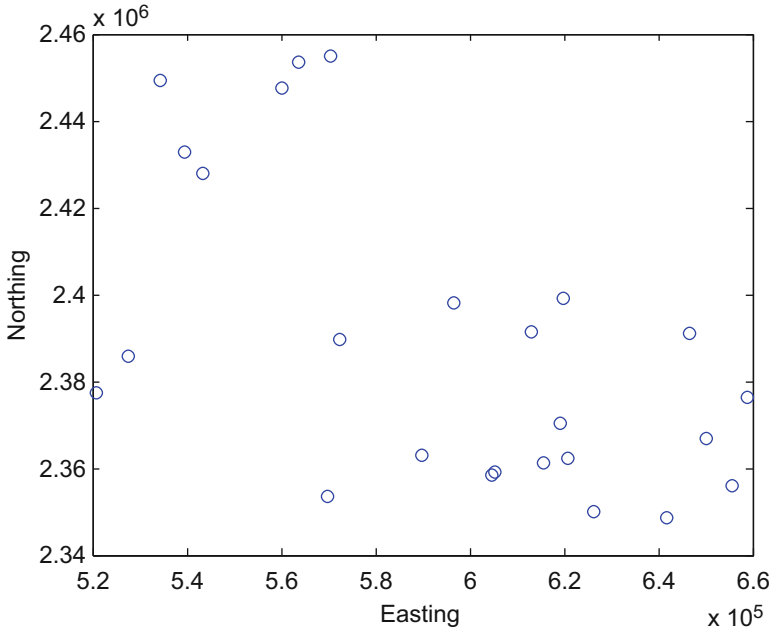


Fig. 2.3 Irregular sample points scatter

Of course, each one of these cases presents different scatter diagrams as presented in Fig. 2.4. In each case, the third variable plays the role of elevation as in the preparation of topographic maps.

In each case, the triplicate {e, n, z} has {e, n}, {e, z}, and (n, z} alternatives, which are referred to as scatter diagrams.

Table 2.4 Calcium, sodium, and chloride measurements

Point	Calcium (ppm)	Sodium (ppm)	Chloride (ppm)	Point	Calcium (ppm)	Sodium (ppm)	Chloride (ppm)
1	135.01	20.76	134.48	16	80.57	45.14	165.41
2	63.11	57.34	194.97	17	133.55	55.47	141.99
3	100.68	42.25	183.24	18	131.69	41.44	126.73
4	88.60	66.59	156.74	19	81.03	35.23	164.19
5	129.13	43.30	182.70	20	129.37	29.48	153.20
6	116.21	40.93	159.03	21	45.79	29.67	179.22
7	85.65	62.31	111.30	22	75.29	54.11	203.53
8	41.85	46.26	103.46	23	121.32	35.14	138.39
9	122.14	30.13	111.18	24	40.99	47.08	192.02
10	84.47	53.61	140.11	25	53.89	27.54	85.94
11	101.54	61.91	169.07	26	60.28	54.89	206.96
12	119.19	20.98	106.39	27	59.87	38.92	100.72
13	132.18	54.06	185.77	28	100.38	63.00	97.85
14	113.82	38.97	145.21	29	67.22	62.68	191.36
15	57.63	61.59	115.56	30	59.88	49.68	170.60

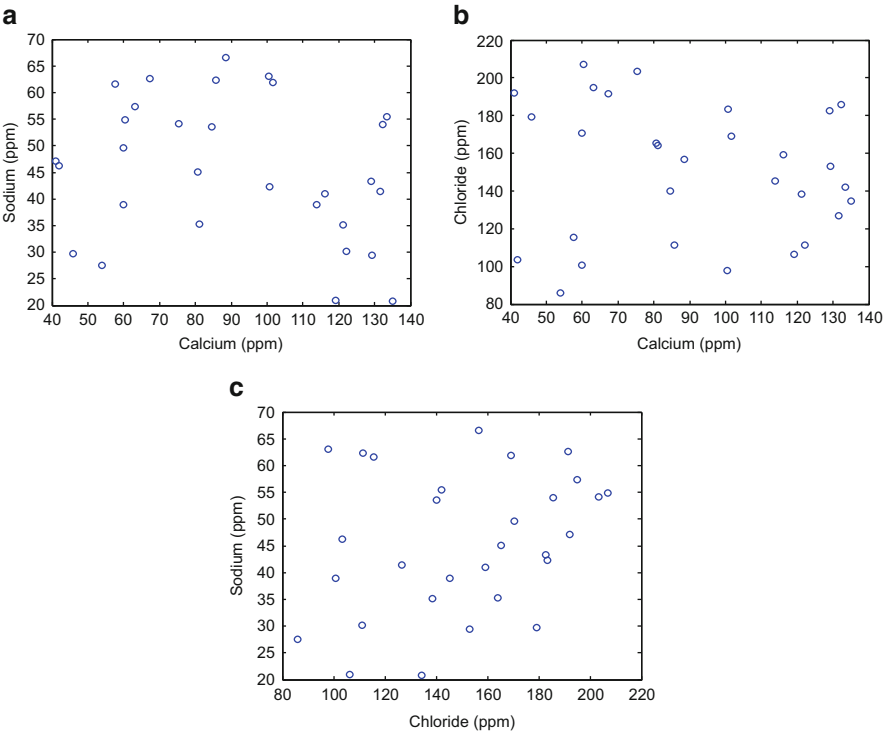


Fig. 2.4 Sample points scatter diagrams (**a**) Calcium-sodium, (**b**) Calcium-chloride, (**c**) Chloride-sodium

2.4 Numerical Data

Evolution of any event takes place both in time and space, but depending on the practical purposes, they can be viewed either temporally or spatially or spatio-temporally. Accordingly, instruments yield numerical data based on either time or space reference system. In this book, only spatial data interpretations and treatment processes are presented. It is assumed herein that the spatial phenomena cover continuously the whole study area. Since the most elemental part of space is a point in earth sciences, the basic sampling locations are points, which may be scattered in the study area either regularly or irregularly. Theoretically, there is infinite number of points but the sampling of all points is not conceivable practically. There are two ways of sampling for conceivable studies. These are:

1. Point sampling: The size of points is considered as dimensionless or infinitesimally small with no practically considerable influence area. This is referred to as sink or source in classical analytical solutions (see Fig. 2.5).
2. Pixel sampling: The size of each point is appreciable, and therefore the whole study area is covered with a mesh of very small squares (see Fig. 2.6).

In general, any point or pixel has a system of longitudes and latitudes. Hence, the whole earth surface is covered by quadrangles, which may be regarded as large-scale pixels. In practice, for Cartesian distance, area, and volume calculation purposes, longitudes and latitudes are converted to "northing" and "southing" values with respect to an initial reference point (see Figs. 2.5 and 2.66). This means that the elements of spatial point data include triple variables, namely, easting, e; northing, n; and the spatial variable measured at this location, say, z. In short we can show the point data as a triple {e, n, z}. Likewise, in addition to these triple values, any pixel includes its resolution size, r, which can be represented by a quadruple {e, n, z, r}. Even though the pixel size is small, one can practically calculate the number of pixels that is necessary to cover a given area.

Example 2.3 If a study region has an area of 45 km^2 and the pixel dimension is 100×100 m^2, what is the number of pixels for the representation of the whole

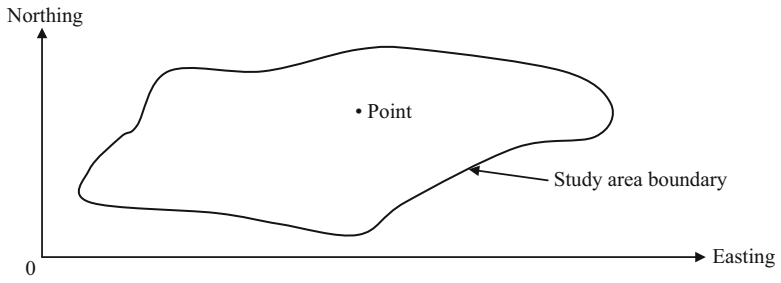


Fig. 2.5 Point sampling in an area

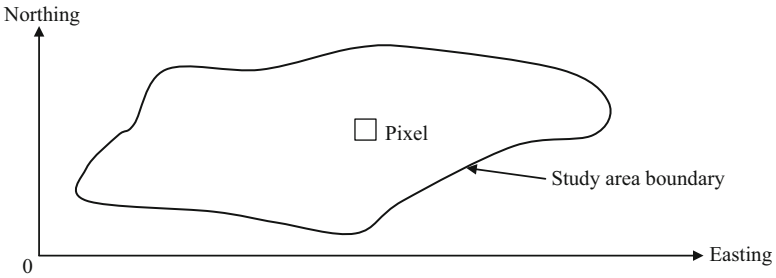


Fig. 2.6 Pixel sampling in an area

region? The number can be calculated as $45 \times 10^6/10^4 = 4500$ pixels. This simple calculation indicates that each pixel has an area of influence defined by its square area. However, it is not possible to make the same calculation for point data, since a point does not have an area of influence by itself. However, it is possible to define the area of influence for each point based on a set of neighboring points as will be explained later in this chapter.

Irregularity is not necessarily related to randomness and has its own connotation in earth and atmospheric sciences. A very brief distinction between these two terms is that although randomness is inherent in the behavior of natural events out of human control, irregularity implies human implications in the measurement and description or definition of natural events. For instance, the locations of meteorology stations or groundwater wells are irregularly scattered in the area or space and consequently they do not comply by any regular or random pattern. Once the irregularity is established, then it does not change with time and space easily until there are other additional inferences by human. Another good distinction in the earth and atmospheric sciences between the regularity and irregularity can be considered in the solution of differential equations by numerical techniques. Finite element method requires a definite and regular mesh to be laid over the solution domain of interest (study area) with regular mesh. Boundary and initial conditions must be defined at regular set of nodes. Various types of regularity are shown in Fig. 2.7. In practical studies measurements as initial conditions are available at a set of irregularly scattered station locations and hence there is not a desirable match between these locations and consequent regular nodes.

Many irregularities appear in practical applications and there is beauty even in the irregularity of the measurement points. Random sampling may be regarded as another type of irregular sampling procedure. All the meteorology station locations, water or oil well sites, soil-sampling points, etc. have irregular patterns in character. The transfer of knowledge from irregular points to regular nodes will be dealt with in detail later in this chapter and other chapters. Without such a knowledge and information transfer, none of the numerical solution models such as Kriging in earth sciences and especially general circulation models (GCM) in atmospheric sciences and others alike can be successfully applicable for practical solutions.

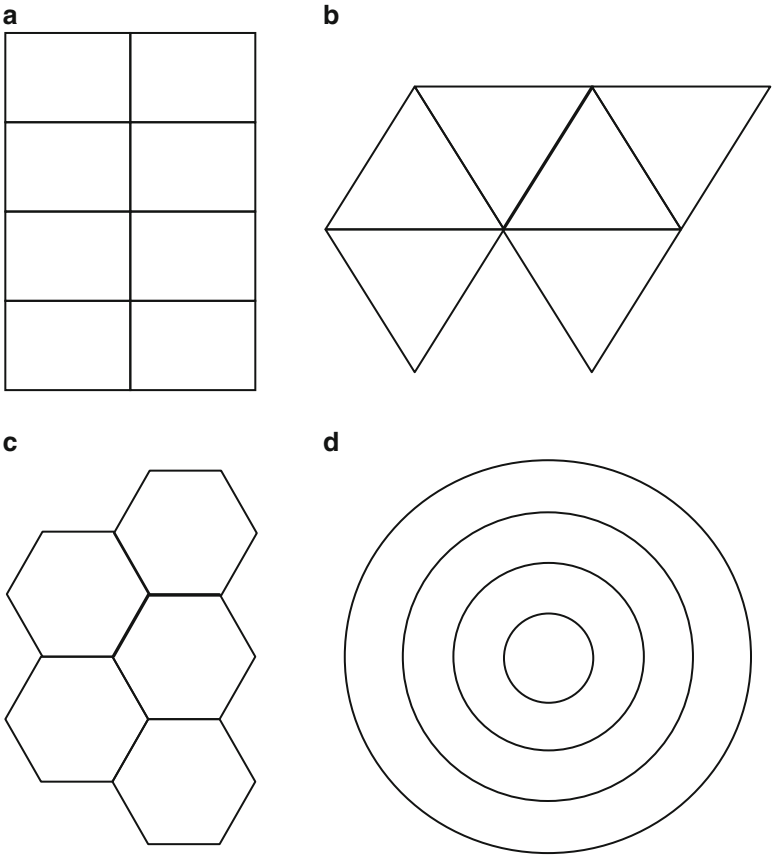


Fig. 2.7 Different regularities (**a**) rectangular; (**b**) triangular; (**c**) pentagonal, (**d**) radial

In theoretical and practical studies, randomness is accounted by the statistical and especially probabilistic methods, whereas irregularity does not need uncertainty methodology for their treatment but for the treatment of random events measured at these irregular points.

2.5 Number of Data

In this book spatial data will imply at the minimum the existence of three points within the study area, because they represent a spatial form as the simplest plain trend. However, it must not be understood that only triplets are necessary for spatial modeling. The practical question is how many data are necessary for spatial data modeling. There is not an easy answer to such a question. The ready answers as appear in open literature such as 12 or 30 data points are not logical but they may

provide vague and practical solutions depending on the circumstances. If the purpose is to check whether the data confirms with the normal probability distribution, then 12 data are enough. On the other hand, 30 is a number which is based on empirical observations that in the representation of average meteorological phenomenon, it is adopted by the World Meteorological Organization (WMO) as a normal period. Unfortunately, neither of these numbers can be adopted in practical applications without looking at the behavior of the concerned phenomenon features, properties, and environmental conditions. In order to have answer in a more realistic manner, the following points must be taken into consideration:

1. If the phenomenon is spatially homogeneous in a deterministic manner, then the number of sample will be as small as possible. In fact, in the case of perfectly homogeneous spatial distribution, only one sample point is enough. This is a very theoretical situation and difficult to come across in earth sciences studies. Logically, in such a case, there is no spatial variation, and therefore the variance of the spatial variation is equal to zero. This implies that the smaller the standard deviation, the smaller the representative sample size. Numerical expression of this situation can be achieved by standard deviation, similarity, or correlation coefficients (Şen 2002). The more the correlation coefficient (close to either + 1, i.e., completely positively proportional or −1 as the negatively inverse), the smaller will be the data number. Finally, the more the similarity coefficient (equal to 1), the smaller will be the data number.
2. If the regional phenomenon is randomly homogeneous (heterogeneous), it is then necessary to take as much data sample (measurement) as possible. However, there must be an upper limit for the data number, which may be decided not only on the basis of statistical criterion but at times more significantly on the cost of samples and on sampling occasions such as in rare events (floods, droughts, etc.).
3. Areal coverage of the samples must also be decided by taking into consideration the effective area of influence of each sample, which will be explained in Chap. 3 in detail.

2.5.1 Small Sample Length of Independent Models

In general, statistical parameters from finite length records are biased estimations of population (very long samples) counterparts. If population values (expectations) are not equal to the parameter estimations then there is a bias effect are biased effect. This is due to the lack of complete information, and consequently, the parameter will be under- or overestimated with respect to its population value. In practice, the true population parameters, which are independent of sample length, are not known, but it is possible to estimate them from available finite length records. As a rule of thumb, it is necessary to have at least 30 values in order to have normal parameter estimations. However, a 30-year period cannot be equally valid for all random or regionalized variables (ReV). It is very much a function of the correlation structure.

In short, the more the persistence (correlation), the smaller is the necessary data number. In practical applications, it is of great interest to know the number of data to have a stable value on the average. This is equivalent to saying that the variance of the average parameter estimation must be constant. From this point, Cramer (1946) showed theoretically for the normal independent ReV that the variance, $V_I(\overline{x})$, of the arithmetic averages for sample length n is

$$V_I(\overline{x}) = \frac{\sigma^2}{n} \tag{2.1}$$

where σ^2 is the unknown population variance and n is the number of data. This is also due to the central limit theorem that the average of random samples accords in normal PDF with mean equal to the average of the data, $\overline{x}$, and the variance of the averages is given as in Eq. 2.1. The square root of this expression is referred to as the standard error, e, of estimate of arithmetic mean. If the original data come from an independent ReV process with population mean, μ, and standard deviation, σ, then the finite sample averages will have the same arithmetic mean with variance as in Eq. 2.1. This means that the sample average is an unbiased estimate of the arithmetic mean with a standard error as

$$e = \frac{\sigma}{\sqrt{n}} \tag{2.2}$$

which decreases with the square root of the sample length. ReV are sampled over various space (or time) intervals. It is necessary to take 10 (90) % standard error (reliability, significance) level corresponding to standard deviate, x_{90}, value from a standard Gaussian distribution as shown in Fig. 2.8. This level separates the whole area under the standard normal PDF into two parts as reliability and risk regions (Şen 1998a).

It is possible to obtain a practical chart between three variables in Eq. 2.2, which gives the relationship between the data number depending on the reliability level as in Fig. 2.9.

Example 2.4 In an extensive area, there is an unconfined groundwater aquifer in rather homogenous sandstone. The field tests indicated that the radius of influence of each well is 650 m. So how many samples must be drilled such that there will not be interference between the adjacent wells in a total region of 5 km^2? For the numerical answer, it is necessary first to calculate the area of influence for each well as $3.14(650)^2 = 1326650$ m^2 $= 1.3266$ km^2. Now the number of wells can be found as $5/1.3266 = 3.769 - 4$ wells.

Example 2.5 In a region $n = 10$, well samples of electric conductivity (EC) in μmhos/cm are recorded as given in Table 2.5. If the purpose is to find the number of representative data for the average value, it is more than enough or more data is necessary.

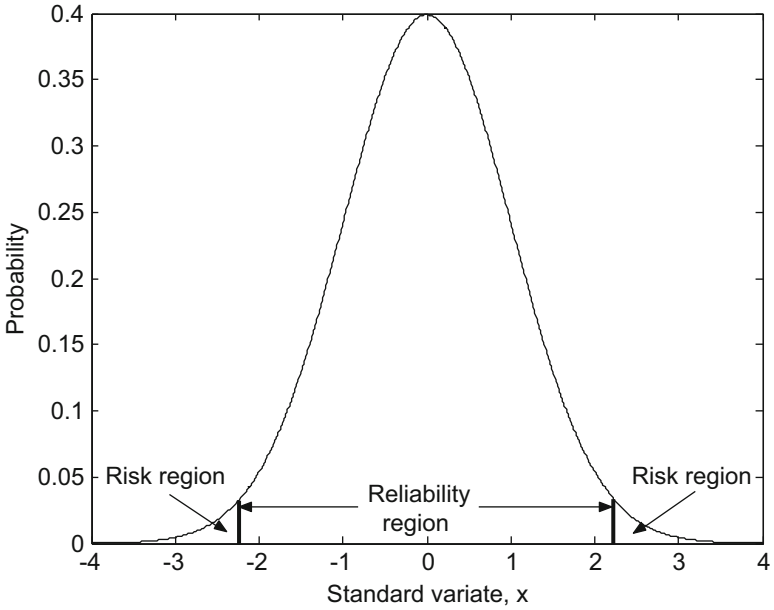


Fig. 2.8 Standard Gaussian PDF

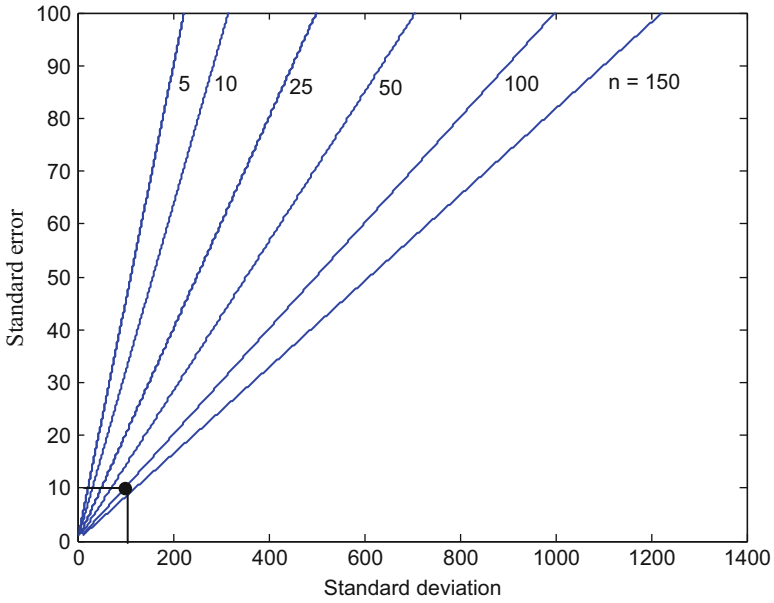


Fig. 2.9 Sample number with the standard deviation and reliability level

Table 2.5 EC samples

Data number	1	2	3	4	5	6	7	8	9	10
EC	770	1020	997	790	750	760	765	850	1029	900

The arithmetic average and the standard deviation of the data values are 863.1 μmhos/cm and 114.82 μmhos/cm, respectively. The samples are assumed to have spatial independence. Find the number of data for $e = 10\,\%$.

Since the standard deviation for the given example is 114.82 ppm, the number of representative data can be found as about 100 with 10 % standard error according to Fig. 2.9.

2.5.2 *Small Sample Length of Dependent Models*

The smaller the sampling interval (distance), the more is the correlation between nearby observations, and consequently the simple result in Eq. 2.2 cannot be applied directly to averages of dependent series. The physical phenomena within the earth sciences (as atmosphere) that give rise to such features are nonlinear dynamic systems with limited predictabilities. Therefore, it is not convenient to have a very persistent process in their systematic modeling, but rather lower-order processes seem more convenient without any mathematical complications. For instance, Schubert et al. (1992) proposed a first-order Markov process to provide a general picture of the persistence-based model behaviors compared to the general circulation model (GCM) in atmospheric sciences. First-order Markov processes have short memory of the correlation function, and therefore, they are not sufficient in the GCM. The same authors then attempted to improve the situation upon the statistical model by fitting low-order univariate autoregressive integrated moving average (ARIMA) models to the control run of the GCM. Detailed information concerning these models is available from Box and Jenkins (1976) as a natural extension of the first-order Markov model and they are useful in modeling atmospheric behaviors (Chu and Karz 1985). In addition to finite sample length, the autocorrelation structure of the process causes further source of bias (Kendall 1954; Quenouille 1956).

In order to model persistence within earth sciences ReV, herein, the ARIMA (1,0,1) model is considered. It is the mixture of separate stochastic processes including autoregressive and moving average models. The numbers in the argument as (1,0,1) imply that this type of ARIMA model is composed of first-order autoregressive (Markov) and first-order moving average processes with zero-order difference between successive values. Generally, the model is written mathematically as follows:

$$x_i = \phi x_{i-1} - \theta \varepsilon_{i-1} + \varepsilon_i \tag{2.3}$$

where ϕ and θ are the autoregressive and moving average parameters, respectively, and ε_i is a zero-mean independent (white noise) random variable. The

autocorrelation structure of the same process is presented in terms of the model parameters as (Box and Jenkins 1976)

$$\begin{aligned}\rho &= 0\\ \rho_1 &= \frac{(\phi-\theta)(1-\phi\theta)}{(1+\theta^2-2\phi\theta)}\\ \rho_i &= \phi\rho_{i-1} \quad (i \geq 2)\end{aligned} \tag{2.4}$$

These expressions reduce to the white noise case when ϕ and θ are equal to zero, to first-order Markov process case if $\theta = 0$ and $\phi = \rho$, and, finally, to the moving average process for $\phi = 0$. For this model, the variance of time averages can be calculated by using model parameters which are the fundamental quantities related to some statistical parameters that can be estimated from available data. In order to illustrate this point, let us consider data smoothened using a simple arithmetic average of length m less than record length n. Hence, the arithmetic average for such a subsample length n is

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.5}$$

By taking first the square and then the expectation operator on both sides leads after some algebra to

$$V(\overline{X}_n) = \frac{1}{n^2}\left[\sum_{i=1}^{n} E(X_i^2) + \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_iX_j)\right] - \mu^2 \tag{2.6}$$

Of course, it is assumed $E(X_i) = \mu$. It is well known from mathematical statistics that for identically distributed random variables, $E(X_i^2) = E(S^2) + \mu^2$ and $E(X_iX_j) = E(S^2)\rho^{|i-j|} + \mu^2$ and their substitution into the last expression yield

$$V_A(\overline{X}_n) = \frac{E(S^2)}{n^2}\left[n + 2\sum_{i=1}^{n}(n-i)\rho_i\right] \tag{2.7}$$

The substitution of ARIMA (1,0,1) autocorrelation structure from Eq. 2.4 leads to

$$V_A(\overline{X}_n) = \frac{E(S^2)}{n^2}\left\{n + \frac{2\rho_1}{(1-\phi)^2}[(1-\phi) - (1-\phi^n)]\right\} \tag{2.8}$$

This expression provides a common basis for the change calculations by giving the variance of the PDF of the ReV means for finite lengths, i.e., subsamples, from a complete data set. For large samples ($n > 30$), the distribution of means converges to a normal PDF due to the well-known central limit theorem in statistics. Square

root of Eq. 2.8 is equal to the standard deviation of such a normal PDF. Consequently, it can be used for determining whether the small sample mean value in a given REV is significantly different from its long-term mean value, μ, supposedly calculated from the whole record. For this purpose, the executions of the following steps are necessary:

1. Identify the underlying stochastic or ReV model for the given earth sciences phenomenon.
2. Find the theoretical variance of the averages by substituting the necessary model parameters into Eq. 2.8.
3. Consider the population PDF of the given data averages as a normal PDF with mean, μ, and standard deviation $\sigma_A = \sqrt{V_A(\overline{X}_n)}$.
4. Find the standard deviate, t, for means calculated from a given length, m, as

$$t = \frac{\overline{X}_n - \mu}{\sigma_A}. \tag{2.9}$$

5. Accept a certain level of significance, α_s, and find the corresponding confidence limit, $\pm t_c$, from the standard normal PDF table given in any statistics book (Davis 2002).
6. If $t \leq |t_c|$, then the conclusion is that only sampling errors cause random variability and there is no long-term persistence variability (change); otherwise, the earth sciences data are affected from systematic changes due to some atmospheric and/or environmental reasons.

For the first-order Markov process similar expressions to Eq. 2.8 can be obtained provided that ϕ is substituted by ρ_1, which leads to,

$$V_A(\overline{X}_n) = \frac{E(S^2)}{n^2}\left\{n + \frac{2\rho_1}{(1-\rho_1)^2}\left[(1-\rho_1) - \left(1-\rho_1^n\right)\right]\right\} \tag{2.10}$$

This expression reduces to Eq. 2.1 for $\rho_1 = 0$, which corresponds to the case of independent model as explained in the previous section.

Generally, even for very small samples ($n < 30$), one can use Chebyshev inequality which states that the probability of single value, say σ_A, selected at random to deviate from μ of the PDF more than $\pm\sqrt{V_A(\overline{X}_n)}$ is less than or equal to $1/K^2$, and it can be expressed in mathematical form as

$$P\left[|\sigma_A - \mu| \geq K\sqrt{V_A(\overline{X}_n)}\right] \leq 1/K^2 \tag{2.11}$$

This inequality yields upper and lower limits on the probability of a deviation of a given magnitude from the mean value. Hence, it is possible to find confidence intervals on either side of the average value.

Example 2.6 Let us calculate 95 % confidence (reliability) limits for an ARIMA (1, 0, 1) model with model parameters $\phi = 0.6$ and $\theta = 0.3$ for a finite length data, $n = 15$. First of all, from Eq. 2.4 one can find that $\rho_1 = 0.34$, the substitution of which with other relevant values into Eq. 2.8 gives $\sqrt{V_A(\overline{X}_n)} = 0.40\sigma$. Confidence level of 95 % implies that $1/K^2 = 0.95$ or $K = 1.026$ and accordingly $K\sqrt{V_A(\overline{X}_n)} = 0.41\sigma$. Consequently, the confidence interval limits are $\mu \pm 0.41\sigma$. Finally, if the calculated square root of variance of a finite data averages with $n = 15$ lies outside of these limits, then there is trend with the possibility of ReV change. Of course, for any desired small sample size, confidence interval limits can be calculated easily according to aforementioned procedure.

On the other hand, for the moving average case, $\phi = 0$, and hence from Eq. 2.4, $\rho_1 = -\theta/(1+\theta^2)$ and its substitution into Eq. 2.8 leads to

$$V_A(\overline{X}_n) = \frac{\sigma^2}{n^2}\left[n - \frac{2(n-1)\theta}{1+\theta^2}\right] \tag{2.12}$$

It is important to notice when this expression becomes identical to Eq. 2.1. On the other hand, the well-known statistical estimate of small sample variance, S^2, is given as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \tag{2.13}$$

The existence of n–1 rather than n in the denominator of Eq. 2.13 is because of obtaining unbiased variance estimate for independent processes. Indeed, taking the expectation of both sides leads to $E(S^2) = \sigma^2$. This indicates with no hesitation that it is possible to substitute σ^2 in Eq. 2.1 by its sample estimate for independent ReV. However, as will be shown in the following sequel, this is not true for dependent processes. It has already been shown by Şen (1974) that for dependent processes,

$$E(S^2) = \sigma^2\left[1 - \frac{2}{n(n-1)}\sum_{i=1}^{n}(n-i)\rho_i\right] \tag{2.14}$$

where ρ_i represents lag-i autocorrelation coefficient. This last expression shows that there is bias effect not only due to the finite sample length, n, but more significantly autocorrelation structure of the ReV. Substitution of Eq. 2.4 into Eq. 2.14 leads after some algebraic manipulations to

$$E(S^2) = \sigma^2\left\{1 - \frac{2\rho_1}{n(n-1)}\left[\frac{n(1-\phi) - (1-\phi^n)}{(1-\phi)^2}\right]\right\} \tag{2.15}$$

This expression proves that there is always an underestimation of the population variance from a given finite data set, the amount of bias which depends on the finite sample length, and the model parameters which govern the dependence structure. In order to obtain an unbiased estimate of the population variance, the sample estimate of the variance must be divided by the curly bracket term in Eq. 2.15 as explained in detail by Şen (1974).

A common point of the equations derived above is their asymptotic convergence to a common value (σ^2/n) as the finite size of data n increases. In order to define objectively a characteristic sample length for convergence, a certain level of relative error percentage α_r will be adopted and it is defined generally as

$$\alpha_r = 100\frac{\sigma^2 - E\left(S^2\right)}{\sigma^2}$$

Hence, substitution of Eq. 2.15 into this expression leads to the most general form of the relative error for the ARIMA (1,0,1) model as

$$\alpha_r = \frac{2\rho_1}{n(n-1)}\left[\frac{n(1-\phi) - (1-\phi^n)}{(1-\phi)^2}\right] \times 100 \tag{2.16}$$

Based on this expression, the change of relative error, α_r, with the sample length n for a set of ρ_1 values is presented as charts in terms of ϕ values in Fig. 2.10.

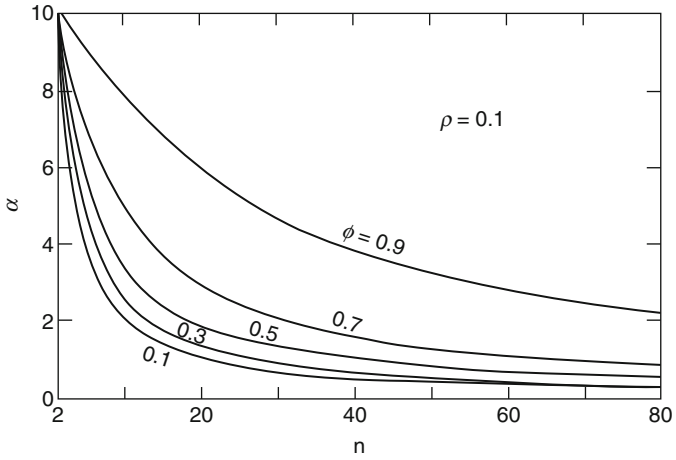


Fig. 2.10 Variance of average estimate relative error change with sample length ($\rho_1 = 0.1$).

A close inspection shows an exponential decrease in all the charts with sample length. Increase in ρ_1 implies increase in the relative error percentage for a given sample length. In these figures the Markov model case corresponds to the curves where $\rho_1 = \phi$. These charts are useful tools in finding equivalent independent model length for a given dependent model length provided that the relative error percentage is given. For instance, when 10 % error level is acceptable in the first-order Markov model small data mean variance estimation with $\rho_1 = 0.3$, then it is possible to find from chart in Fig. 2.10 that after at least eight samples, the ReV will not have dependence structure. However, for the same error level and dependence coefficient, an ARIMA (1,0,1) model requires at least 11, 16, and 48 samples for $\phi = 0.5$, 0.7, and 0.9, respectively.

These charts can also be used as indicators whether there is a trend in a given ReV. For example, if the underlying generating mechanism is identified through using procedures of Box and Jenkins (1976) as an ARIMA (1, 0, 1) model with $\rho_1 = 0.5$ and $\phi = 0.7$, then the error percentage is calculated from Eq. 2.16. If this error percentage is greater than the theoretically calculated counterpart, then there is a possibility of trend.

Another relative error for measuring any deviation from the independent model with no long-term systematic features can be defined by considering Eqs. 2.1 and 2.8 as

$$\beta = 100 \frac{V_A(\overline{X}_n) - V_I(\overline{X}_n)}{V_A(\overline{X}_n)}$$

This expression provides information about the necessary sample length concerning the variance of climatic subsample averages. It is possible to write this expression explicitly in terms of model parameters from Eqs. 2.1 and 2.8 as

$$\beta = 100 \left\{ 1 - \frac{1}{n + \frac{2\rho_1}{(1-\theta)^2}\left[(1-\theta) - (1-\theta^n)\right]} \right\} \tag{2.17}$$

Finally, if the variance of short data and the variance of subsample averages are calculated for the same sample length, then the relationship between these two types of relative errors can be found from Eqs. 2.16 and 2.17 as

$$\beta = 100 \left[1 - \frac{1}{1 + \frac{n-1}{100}\alpha_r} \right] \tag{2.18}$$

which implies that β is equal to zero only when α_r approaches to zero. Otherwise, for any given α_r, there is an implied β-type error. The graphical representation of Eq. 2.18 is given in Fig. 2.11. If any pair of α_r, β, and n values is known, then the third one can be determined from this chart easily.

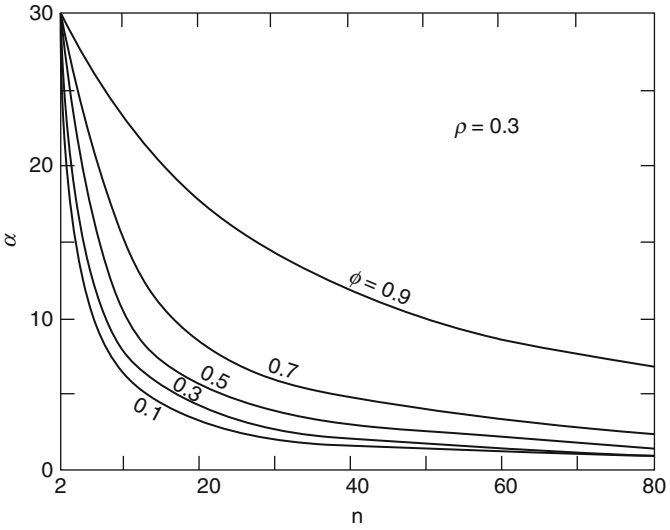


Fig. 2.11 Variance of climatic series average estimate relative error change with sample length ($\rho_1 = 0.3$)

2.6 Regional Representation

Decision-making about which model to use for spatial estimation should be based on some simple criteria as the number of data points, variability range, arithmetic average, standard deviation, geometrical configuration of sampling points, distances, and in general distance-regional variability. Classical parameters as the arithmetic average, standard deviation, skewness coefficient, etc. in terms of summary statistics can be readily known by the reader, and therefore, they will not be elaborated in this book (see Davis 2002).

2.6.1 Variability Range

In spatial analysis, the purpose is to estimate ReV at any desired point within the study area using the measurements at a set of sampling points. Such a task will depend not only on the internal variability at a fixed site of the variable but more significantly on the regional scatter of sampling points and number of points. Provided that the internal variability range does not differ more than 5 % between the maximum and the minimum data values, then one can depend on either the arithmetic average or more specifically on the mode (the most frequently appearing data value) as the sole representative for whole region. This method depends on any number of data without data number specification. If the maximum and the

minimum ReV values are Z_M and Z_m, respectively, then the range of data variability is defined as

$$R_Z = Z_M - Z_m \tag{2.19}$$

The maximum percentage error can be defined as follows:

$$e_m = 100\frac{R_Z}{Z_M} \tag{2.20}$$

If $e_m \leq 5$, then the regional representative value can be taken as the arithmetic average, $\overline{Z}$, at any point. This needs no detailed spatial modeling and it shows that the phenomenon is not complex within the area and its behavior is more or less homogeneous.

In order to be on a better side instead of arithmetic value, the mode, M_Z, value can be adopted. The mode is the most frequently occurring data value within the whole record.

Whatever the case is, the regionalized variable is represented by a constant value, and therefore, there will be deviations from this constant level, which are the errors that could not be accounted by the constant value. Mathematically, it is possible to suggest a procedure where the sum of square deviations (SSD) is the minimum. Hence, if the constant level of n data values $Z_i(i = 1,\ 2, \ldots, n)$ is indicated by Z_C, then the SSD becomes

$$\text{SSD} = \sum_{i=1}^{n} (Z_i - Z_C)^2 \tag{2.21}$$

which on expansion becomes

$$\text{SSD} = \sum_{i=1}^{n} Z_i^2 - 2Z_C \sum_{i=1}^{n} Z_i + nZ_C^2 \tag{2.22}$$

Since the minimum value is sought, the derivative of this last expression with respect to Z_C leads to

$$\frac{\partial(\text{SSD})}{\partial Z_C} = -2\sum_{i=1}^{n} Z_i + 2nZ_C \tag{2.23}$$

If this is set equal to zero, then the result becomes

$$Z_C = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{2.24}$$

which is equal to the arithmetic average, $\overline{Z}$. This proves that irrespective of any practical requirement, the use of arithmetic average as a representative value for the regionalized variable is a mathematical necessity. Further interpretation of Eq. 2.24 can be given after its explicit form as

$$Z_C = \frac{1}{n}Z_1 + \frac{1}{n}Z_2 + \frac{1}{n}Z_3 + \ldots + \frac{1}{n}Z_{n1} \tag{2.25}$$

which means that the arithmetic average is the summation of point data value multiplied by a factor ($1/n$ as weighting factor), which may have different interpretations:

1. The factors may be considered as weights for each data point, and therefore, they may represent some specific feature of the data point apart from the data point value. For instance, weight may be the influence area of the point or some important coefficient such as the relationship of this data value with available points. In this latter case, the factors may be the function of correlation between data point pairs in addition to distances. In Eq. 2.25, the factors are the same, which implies the isotropic behavior of the regionalized phenomenon that is not the case in natural events, and therefore, each factor is expected to be different than others.
2. In Eq. 2.25 equivalence of factors as $1/n$ reminds one that if there are n data points provided that they have equally likely future occurrence chances, the probability of occurrence for each data value is $p = 1/n$. This corresponds to random field case where the occurrences are completely independent from each other, i.e., there is no regional correlation within the regionalized variable. However, in actual situations these probabilities of occurrences are not equal.
3. Another significant point from Eq. 2.25 is that the summation of the factors or probabilities is always equal to 1. Furthermore, the factors are in percentages.
4. Finally, each factor is distance independent, whereas in any natural phenomenon, there are regional dependences, which may be simply expressed as inverse distance or inverse distance square weightings without natural consideration (see Sect. 2.6.2).

If the word "weight" is used instead of aforementioned factor or probability, then the arithmetic average expression in Eq. 2.25 can be written for more general uses as the weighted average as

$$Z_C = \sum_{i=1}^{n} \alpha_i Z_i \tag{2.26}$$

where α_i's are the weightings. Considering the percentage property of the weightings, this last expression can be written in its most explicit form as follows.

$$Z_C = \sum_{i=1}^{n} \left(\frac{W_i}{W_T}\right) Z_i \tag{2.27}$$

where W_i is the weight attached to ith data point and W_T represents the total weight as the summation of all weights:

$$W_T = W_1 + W_2 + \ldots + W_n \tag{2.28}$$

with n number of data points. The weights assigned to the estimation points are adjusted to sum to 1.0. A weighting process provides a function, which shows the change of influence with distance in an inversely proportional functional form. The closer are the estimation and measurement points the more is the weighting factor. This inverse distance-squared weighting function is then scaled so that it extends from one to zero over this distance. Equation 2.26 is in the form of linear multiple regression, and it furnishes the basis of all the linear estimation models in spatial analysis as will be explained in this book.

Logically, the use of mode value is preferable because it is the most frequently occurring data value within the given set of data points. The basis of mode is probabilistic rather than mathematical and it has the most likely probability of occurrence. A decision is to be made between the arithmetic average and the mode value in practical applications. If the PDF of the regionalized variable is symmetric, then the two concepts fall on each other. Otherwise, the use of mode must be preferable, but mathematically its calculation is not easy, and therefore, in any statistical or mathematical modeling, invariably arithmetic average is used. The reader should keep in mind that it yields the minimum error but does not abide with the most frequently occurring data value. The arithmetic average value renders the mathematical procedures into a tractable form.

2.6.2 *Inverse Distance Models*

As mentioned in the previous section, the general regional estimation procedure is given in Eqs. 2.26 and 2.27. In these expressions, the most important part is the attachment of weights. Once the weights are known, the remaining is simple calculation of the desired value at the point of estimation. Rational and logical procedure is straightforward, and one can conclude that as the distance, d, between two sites increases, their effect on each other becomes meager, and hence, if only the distance is considered without the regional variability feature of the phenomenon concerned, then the first rational approach suggests that the weights can be taken as inverse distances, 1/d. However, in such an approach, very small distances yield big weights and at zero distance it is not defined. Another path of thought is

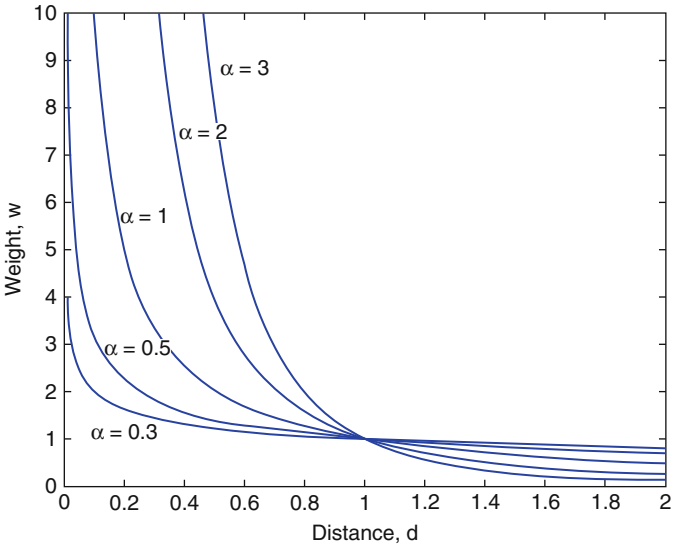


Fig. 2.12 Various inverse distance models

the gravitational force between two heavenly bodies where the attraction force is inversely proportional with the square of distance, $1/d^2$. It is even possible to generalize the inverse distance methodology as $1/d^{\alpha}$, where α is a power parameter assuming values greater than 1. The explicit expression used for the estimation, $\hat{Z}_j$ at point j according to IDP, is given as

$$\hat{Z}_j = \frac{\sum_{i=1}^{N} \frac{1}{d_{ij}^{\alpha}} Z_i}{\sum_{i=1}^{N} \frac{1}{d_{ij}^{\alpha}}} \tag{2.29}$$

where d_{ij} is the effective separation distance between grid node j and the neighboring points i, Z_i's are the neighboring point ReV values, and α is the weighting (smoothing) power. It is possible to derive special features from this general expression which are well known in the literature. The last equation becomes equal to inverse distance and inverse square distance for $\alpha = 1$ and $\alpha = 2$, respectively. The slopes at the points used in the estimation procedure are weighted according to the distances between the estimation node and other points. Various inverse distance functions are presented in Fig. 2.12.

Weighting is assigned to data through the use of a weighting power that controls how the weighting factors drop off as distance from a grid node increases. The greater the weighting power, the less effect points far from the grid node have during interpolation. As the power increases, the grid node value approaches the value of the nearest point. For a smaller power, the weights are more evenly

distributed among the neighboring data points. Normally, inverse distance function behaves as an exact interpolator but still does not take into consideration the inherent ReV variability in the phenomenon concerned. In calculating a grid node, the weights assigned to the data points are fractions and the sum of all the weights is equal to 1.0 (Eq. 2.28). When a particular observation is coincident with a grid node, the distance between that observation and the grid node is 0.0, and that observation is given a weight of 1.0, while all other observations are given weights of 0.0. Thus, the grid node is assigned the value of the coincident observation. Here α is a mechanism for buffering this behavior. When one assigns a nonzero smoothing parameter, no point is given an overwhelming weight so that no point is given a weighting factor equal to 1.0.

One of the characteristics of inverse distance function is the generation of "bull's-eyes" surrounding the position of observations within the gridded area. One can assign a smoothing parameter during inverse distance function to reduce the "bull's-eye" effect by smoothing the interpolated grid. Inverse distance function is a very fast method for gridding. With less than 500 points, one can use all data search types and gridding proceeds rapidly.

2.7 Subareal Partition

The whole study area may be partitioned into regular or irregular subareas each with the same feature; hence the regional changes are considered as partially homogeneous within each subarea. For instance, in Figs. 2.3 and 2.4 although the subareas are in the form of regular quadrangles, sample points are located within these subareas in a systematic or random manner. In this section, instead of points subareas are considered in the modeling of ReV variability.

2.7.1 Triangularization

The number of data points is also important in deciding which spatial prediction model must be adopted. No need to say that spatial variability is describable by at least three data points, which give the shape of a triangle, and therefore, in the early approaches before the advent of computers, triangularization method was preferred due to its simplicity. If there are n data points, it is possible to obtain $n - 2$ adjacent triangles. For instance, five sampling points (A, B, C, D, and E) in Fig. 2.13 yield to three triangles. Each corner is a measurement station. The ReV within each triangle may be represented as the arithmetic average of the data values at three apices. In this manner rather than the use of arithmetic average over all the study area, it is partitioned into triangular subareas, where the arithmetic averages are used. In this manner, the amount of the error in the global arithmetic average usage is reduced significantly.

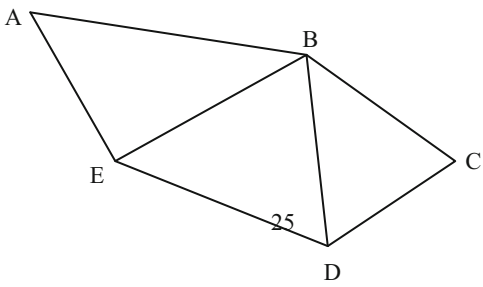


Fig. 2.13 Triangularization

The problem in triangularization is that one point record enters more than one subarea. The question is, is it not better to define an influence area for each subarea separately? So that the whole study region is divided into a set of subareas with influence of the data point.

In practice, given the scarcity of gauges and the spatial variability of ReV, for instance, in the case of precipitation, many storms completely miss several gauges within a drainage area. Therefore, two basic tasks must be performed, namely, the assessment of the representativeness of point rainfall and picture derivations of spatial patterns which reflect reality. Summer (1988) states, "In the ideal world, it should be possible to track and then model, mathematically or statistically the passage of a storm across an area with a great degree of accuracy and precision. In reality, this is very difficult, and one must be content with generalized spatial models of storm structure relating intensity or depth to storm area." Kriging and stochastic methods for the areal average estimation (AAE) based on the spatial correlation coefficient are summarized by Bras and Rodriguez-Iturbe (1985). However, the use of these methods needs recordings at many stations for the results to be reliable. Tabios and Salas (1985) compared several AAE methods with rainfall variability and concluded that a geostatistical method (ordinary and universal) (Kriging, Chap. 4) with spatial correlation structure is superior to Thiessen polygons, polynomial interpretation, and inverse distance weighting. Hevesi et al. (1992) suggested the use of multivariate geostatistical techniques for areal precipitation estimation in mountainous terrain. Reliable estimates by these techniques are particularly difficult when the areal coverage of stations is sparse or when precipitation characteristics vary greatly with locations. Such situations frequently occur in arid regions due to sporadic and haphazard meteorological occurrences. On the other hand, Kedem et al. (1990) have shown by considering satellite images and simple probability models that the higher the rainfall, the smaller the affected area over large regions. All these methods require high-speed computers, and they are not as practical as conventional procedures such as the arithmetic average, Thiessen polygons, or isohyetal map techniques which do not require much data (Chow 1964).

An alternative AAE calculation method is presented by Akin (1971). It is assumed that the precipitation over the subarea varies linearly between the

three-corner (triangular) gauge points. Thus at any point (x, y) interior to the nth subarea, the precipitation height $H_n(x, y)$ is expressed as

$$H_n(x, y) = \alpha_n + \beta_n x + \gamma_n y \tag{2.30}$$

where α_n, β_n, and γ_n are constants related to the gauge measurements at the corners. It is possible to write three simultaneous equations one for each apex (i, j, k) as

$$\begin{aligned} H_i(x, y) &= \alpha_n + \beta_n x_i + \gamma_n y_i \\ H_j(x, y) &= \alpha_n + \beta_n x_j + \gamma_n y_j \\ H_k(x, y) &= \alpha_n + \beta_n x_k + \gamma_n y_k \end{aligned} \tag{2.31}$$

The solution of constants from these equations in terms of known quantities leads to

$$\begin{aligned} \alpha_n &= \left[a_i H_i + a_j H_j + a_k H_k\right]/2A_n \\ \beta_n &= \left[b_i H_i + b_j H_j + b_k H_k\right]/2A_n \\ \gamma_n &= \left[c_i H_i + c_j H_j + c_k H_k\right]/2A_n \end{aligned} \tag{2.32}$$

where

$$\begin{aligned} a_i &= x_j y_k - x_k y_j \\ b_i &= y_j - y_k \\ c_i &= x_k - x_j \end{aligned} \tag{2.33}$$

Following a cyclic permutation of (i, j, k), the subarea A_n can be calculated as follows:

$$A_n = \left[a_i + a_j + a_k\right]/2 \tag{2.34}$$

The differential volume of rainfall at any point within the subarea is defined as

$$dQ = H(x, y)dA \tag{2.35}$$

so that the total volume of rainfall associated with the subarea becomes theoretically as

$$Q_n = \int\int \left[\alpha_n + \beta_n x_n + \gamma_n y_n\right] dxdy \tag{2.36}$$

where the substitution of the above relevant expressions leads after some algebra, finally, to the volume of rainfall for the nth subareas as

$$Q_n = A_n\left[\alpha_n + \beta_n\left(x_i + x_j + x_k\right)/3 + \gamma_n\left(y_i + y_j + y_k\right)/3\right] \tag{2.37}$$

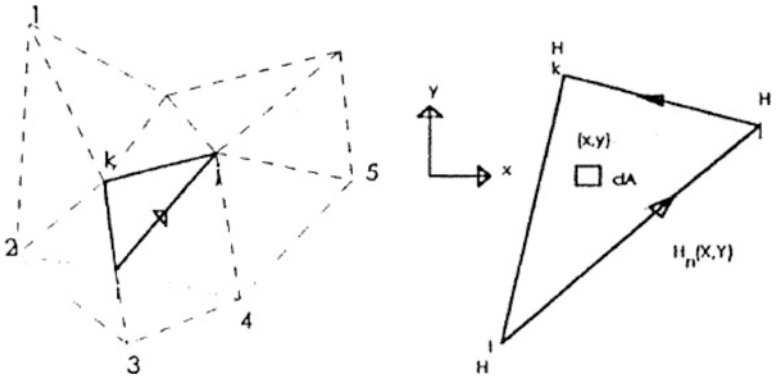


Fig. 2.14 Notations used for a triangular subarea

In the case of m triangular subareas, the total rainfall volume becomes

$$Q = \sum_{n=1}^{m} Q_n \tag{2.38}$$

and the corresponding total area is

$$A = \sum_{n=1}^{m} A_n \tag{2.39}$$

Finally, the ratio of Eqs. 2.38 and 2.39 gives the AAE height over m subareas as

$$\overline{H} = \frac{Q}{A}$$

By means of this procedure, the AAE area and volume are easily calculated if the gauge locations and rainfall amounts are known.

The development is analogous to some elementary concepts used in finite element analysis techniques. Consider a region of interest with station locations and amounts that are plotted on a map, then a series of straight lines are drawn arbitrarily to connect every gauge points with the adjacent gauges. Straight lines are drawn in anticlockwise direction as shown in Fig. 2.14. These straight lines should produce a series of triangles "not necessary to have the same shape."

Each triangle area is known as subarea and the corners of the triangles are shown by $[i,\ j,\ k]$. Precipitation values at the corners are denoted as $H_i, H_j, H_k (i, j, k = 1,\ 2, \ldots, n)$, where n is the number of subareas. Triangularization of all the stations for the northern part of Libya is shown in Fig. 2.15. On the other hand, Table 2.6 shows subareas for this study.

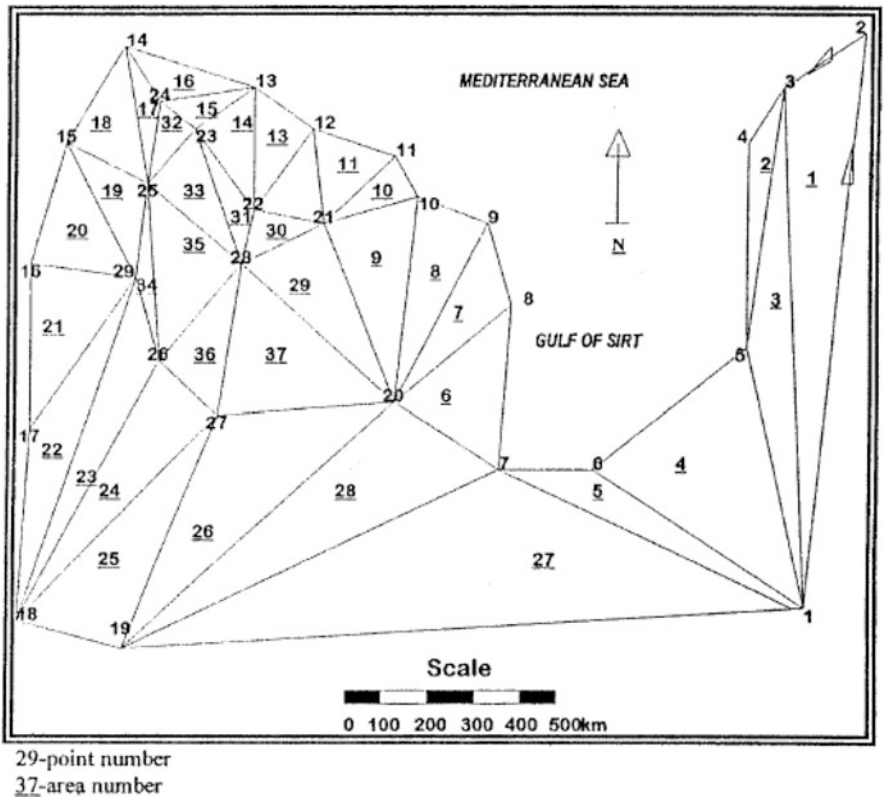


Fig. 2.15 Triangular meshes for northern Libya

Convenient software is developed to calculate the subareas, associated AAR depths, and volumes falling on each subarea as well as the total area of interest. The results are shown for monthly AARs in Table 2.7, while the seasonal AAR depths and volumes are presented in Table 2.8.

These values are computed by developed software over the study area as a summation of 37 subareas already shown in Fig. 2.15 and Table 2.8.

2.8 Polygonizations

The idea is to surround each data point with a polygonal area of influence so that the number of points will be equal to the number of subareas. In the case of n data points, there will be n area of influence.

Table 2.6 Values of i, j, and k for each subarea

Subareal number	i	j	k	Subareal number	i	j	k
1	1	2	3	20	15	16	29
2	3	4	5	21	16	17	29
3	3	5	1	22	17	18	29
4	5	6	1	23	18	26	29
5	6	7	1	24	18	27	26
6	7	8	20	25	18	18	27
7	7	8	20	26	19	20	27
8	9	10	20	27	19	1	7
9	10	21	20	28	19	7	20
10	20	11	21	29	20	21	25
11	11	12	21	30	21	22	28
12	12	22	21	31	22	23	28
13	12	13	22	32	15	25	24
14	13	23	22	33	28	23	25
15	13	24	23	34	26	25	29
16	13	14	24	35	26	28	25
17	14	25	24	36	27	28	26
18	14	15	25	37	20	28	27
19	15	29	25				

Table 2.7 Monthly averages of rainfall depths and volumes

Months	Average depth (mm)	Average volume ($\times 10^6$ m^3)
Jan	28.2	97.400
Feb	12.7	64.889
Mar	16.2	47.684
Apr	4.4	16.564
May	5.5	19.980
Jun	0.41	1.600
Jul	0.03	0.160
Aug	0.17	0.800
Sep	2.9	9.060
Oct	7.13	23.230
Nov	15.7	39.600
Dec	28.9	112.10

Table 2.8 Seasonal averages of rainfall depths and volumes

Season	Average depth (mm)	Average volume ($\times 10^6$ m^3)
Winter	72.800	274.389
Spring	26.100	84.228
Summer	0.613	2.560
Autumn	25.700	71.890

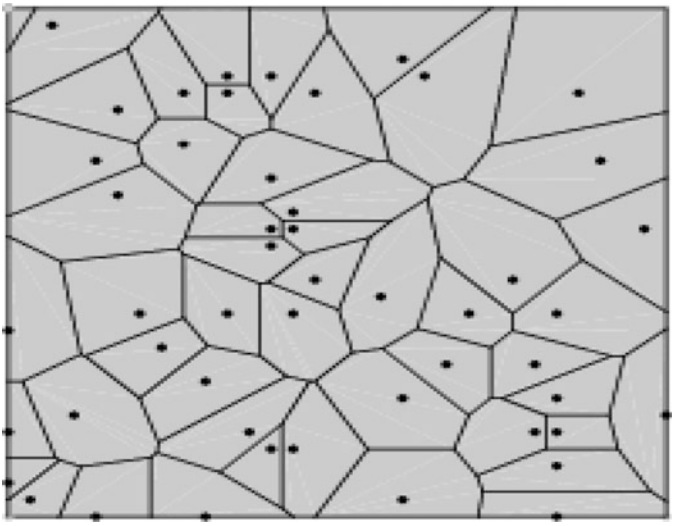


Fig. 2.16 Polygonizations

2.8.1 Delaney, Varoni, and Thiessen Polygons

These polygonal methods are all related to each other with simple differences but the basic logic is the same. Different names are used in different disciplines for the same method. According to these polygon methods, the study area can be partitioned into a set of convex polygons each containing only one measurement point such that every point within a given polygon is closer to the measurement point than any other measurement points. Each polygon defines the area of influence around the measurement point (Fig. 2.16). Each one of these polygons is also called as Thiessen polygon whose boundaries define the area that is closer to each point relative to all other points.

They are geometrically defined by the perpendicular bisectors of the lines between all points. A Voronoi diagram is sometimes also known as a Dirichlet (1850) tessellation with cells that are Dirichlet regions, Thiessen or Voronoi polygons (Dirichlet 1850; Voronoi 1907; Thiessen 1912). On the other hand, the Delaunay triangulation and Voronoi diagram in 2D space are dual to each other in the graph theoretical sense. Voronoi diagrams are named after a Russian mathematician who defined and studied the general n-dimensional case in 1908. Voronoi diagrams that are used in geophysics and meteorology to analyze spatially distributed data (such as rainfall measurements) are called Thiessen polygons. In climatology, Voronoi diagrams are used to calculate the rainfall of an area, based on a series of point measurements. In this usage, they are generally referred to as Thiessen polygons.

The Thiessen method is quick to apply because once the sub-polygons are fixed with a set of observation points, they remain the same all the time. The only change occurs artificially when additional observation points are added to the available set of measurement points. It is based on the hypothesis that for each point in the area, the best estimate of ReV is the measurement physically closest to that point. This concept is implemented by drawing perpendicular bisectors to straight lines connecting each two measurement stations, which yields with the consideration of the watershed boundary a set of closed areas known as Thiessen polygons. Based on the given measurement stations, the sub-polygons are obtained according to the following steps:

1. Connect each station to each nearby station with a straight line. These lines cannot cross and should connect only the nearest stations. The end product is several triangles.
2. Each side of the triangles is then bisected with a perpendicular line, thus forming polygons around each station.
3. Using an appropriate method, calculate the total area, A, and subareas represented by each polygon ($A_1, A_2, \ldots, A_n$). The number of subareas is equal to the number of measurement locations (Eq. 2.28).
4. Calculate the areal average estimation (AAE) of ReV as Z_C according to the weighted average formulation in Eq. 2.27. In this equation, W_i's correspond to subarea at measurement location i with measurement Z_i and W_T is the total area.

Example 2.7 Consider the measurement locations A, B, C, and D as shown in Fig. 2.17. Let the precipitation values be as 3.5 mm, 2.9 mm, 5.8 mm, and 4.2 mm, respectively. Draw the Thiessen sub-polygons and calculate the areal average ReV value. The total area is given as $A = 50\ \text{km}^2$, and after the polygon partitioning each sub-polygon area is calculated as $A_A = 12\ \text{km}^2$, $B_A = 17\ \text{km}^2$, $C_A = 13\ \text{km}^2$, and $D_A = 8\ \text{km}^2$.

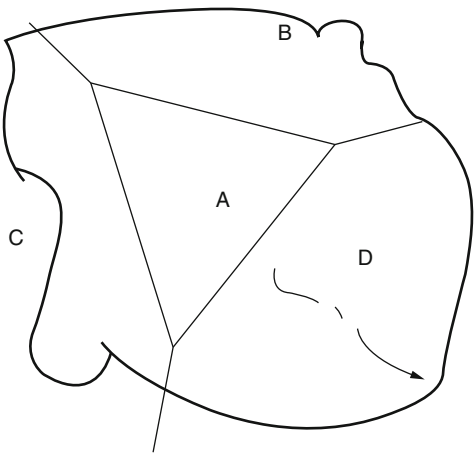


Fig. 2.17 Thiessen polygons

Using Eq. 2.27, the Thiessen areal average of ReV estimation for the entire area becomes

$$Z_C = \frac{12 \times 3.5 + 17 \times 2.9 + 13 \times 5.8 + 8 \times 4.2}{50} = 4.174 \text{ mm}$$

In contrast, the simple arithmetic average of precipitations is 4.1 inches. Differences between arithmetic and Thiessen averages increase for nonuniform ReV when the Thiessen areas differ widely. In the above given example, since the ReV measurements are close to each other, the AAE of ReV is close to the arithmetic average. However, if the ReV values at the same measurement locations are 2.5 mm, 15.3 mm, 8.5 mm, and 22.9 mm, then the AAE ReV value becomes as

$$Z_C = \frac{12 \times 2.5 + 17 \times 15.3 + 13 \times 8.5 + 8 \times 22.9}{50} = 11.675 \text{ mm}$$

which indicates significant difference from the arithmetic average.

Unfortunately, the most commonly used Thiessen (1912) polygon method for AAE calculations does not consider areal precipitation amounts recorded at individual stations in the partition of the whole catchment area into smaller polygonal subareas. Therefore, once the polygons are obtained on the basis station location configuration, they remain the same as long as the measurement locations do not change or there are no additional stations. However, it is logically plausible to expect that the subareas should change in response to the spatial variation of the phenomenon concerned. In other words, the partition should be based not only on the measurement location network configuration but also on the ReV measurements at each location.

2.8.2 *Percentage-Weighted Polygon (PWP) Method*

It is a new, simple, practical, and objective AAE method for determining the areal average of the spatial event based on a sparse and/or irregular network of measurement locations (Şen 1998b). This method takes into account ReV measurement percentage weightings for each station and also has geometrical advantages, i.e., a better representation of the ReV on the study area compared to the conventional Thiessen polygon method. For instance, ReV data such as precipitation show a considerable spatial variation over any region as explained by Huff and Neill (1957), Stout (1960), Jackson (1972), and Summer (1988). Wilson and Atwater (1972) suggested that this variation is due to differences in the type and scale of precipitation-producing models, which are strongly influenced by local or regional factors such as topography and wind direction. The AAE of the ReV over an area is most conveniently determined from a well-sited network of measurement locations, which show the local variations of ReV. For most areas, each measurement station is assumed to represent a very considerable area around it. However, this is a

restrictive and frequently invalid assumption because of the spatial variation of ReV, especially for short distances (durations), such as during severe storms (Summer 1988). There is no guarantee that point measurements provide reliable estimation for immediate surrounding areas. Relief often leads to large variations in ReV over relatively short distances. In the case of say precipitation, if upward motion of air occurs uniformly over thousands of square kilometers, the associated precipitation has usually light or moderate intensity and may continue for a long time. On the other hand, convective storms accompanied by compensating downdrafts (as best illustrated by the thunderstorm) may be extremely intense but their areal extent and local duration are comparatively limited. In practice, given the scarcity of gauges and the spatial variability of ReV, many stations may completely miss measurement. Therefore, two basic tasks must be performed:

1. Assess the representativeness of point measurement.
2. Derive pictures of spatial patterns, which reflect reality.

After deciding on the triangles, the following procedure is necessary for dividing the study area into polygons leading to the percentage weighting method. If the precipitation values at three apices of a triangle are A, B, and C, then their respective percentages are

$$p_A = 100A/(A + B + C) \tag{2.40}$$

$$p_B = 100B/(A + B + C) \tag{2.41}$$

and

$$p_C = 100C/(A + B + C), \tag{2.42}$$

respectively. Hence, it is possible to find the three-variable percentage data of constant sums for each triangle. A 2D plot of three variables can be shown on a triangular graph paper as one point (see Fig. 2.18). Such papers are very common tools in earth sciences (Koch and Link 1971).

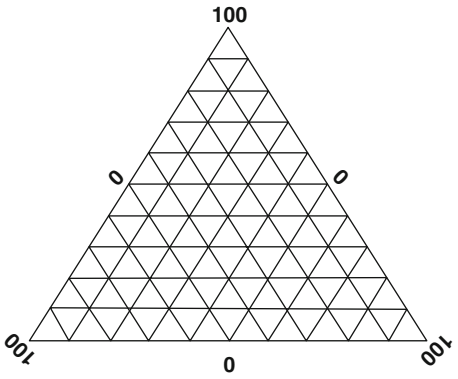


Fig. 2.18 The triangular coordinate paper

In order to demonstrate the method more explicitly, the following step-by-step algorithm is provided which can be computerized:

1. Draw lines between each adjacent pair of precipitation stations. Hence, a set of triangles is obtained that cover the study area.
2. For each triangle calculate the precipitation percentage at its apices according to Eqs. 2.40, 2.41 and 2.42. Consider in each station that each apex has the value of 100 percentages with zero percentage on the opposite side.
3. Consider bisector, which connects an apex to the midpoint of the opposite side, and graduate it into 100 equal pieces.
4. By making use of one precipitation percentage calculated in step 2, mark it along the convenient bisector starting from the opposite side toward the apex.
5. Draw a parallel line from this marked point in step 4 to the side opposite to the apex considered with its precipitation percentage.
6. Repeat steps 4 and 5 for the next precipitation percentage and find similar parallel line this time to another opposite side.
7. The intersection of these two lines defines the key point for the triangle considered.
8. In order to check the correctness of this key point, repeat steps 4 and 5 for the remaining third precipitation percentage value. If the parallel line to the side crosses through the aforementioned key point, then the procedure of finding the key point for the triangle is complete. Otherwise, there is a mistake either in precipitation percentage calculations or in the location of marked points along the bisectors in steps 3 through 6 inclusive.
9. Return to step 2 and repeat the process for the triangles constructed in step 1. In this manner each triangle will have its key point. The location of this point within the triangle depends on the percentages of recorded precipitation amounts at the three adjacent apices. The greater the precipitation percentage for an apex, the closer the point will lie to this apex. It is not necessary that the triangles resulting from a given set of stations in a network should be exactly equilateral. However, in the Thiessen method for an obtuse-angle triangle, the intersection of the three perpendicular bisectors occurs outside the triangular area.
10. Key points at adjacent triangles are connected with each other to form polygons each including a single precipitation station.
11. The boundaries of polygons around the basin perimeter are defined by drawing a perpendicular to the sides of triangles from the key points. Now, the division of the whole basin area into subareas is complete.

The triangular or polygon methods are preferred in practice when there are several data points about 8–10.

Example 2.8 A simple example of plotting on a triangular coordinate paper is presented in Fig. 2.19 for an obtuse-angle triangle.

If the precipitation amounts at three apices are $A = 12.5$ cm, $B = 20.1$ cm, and $C = 7.4$ cm, then the corresponding percentages from Eqs. 2.40, 2.41 and 2.42 are

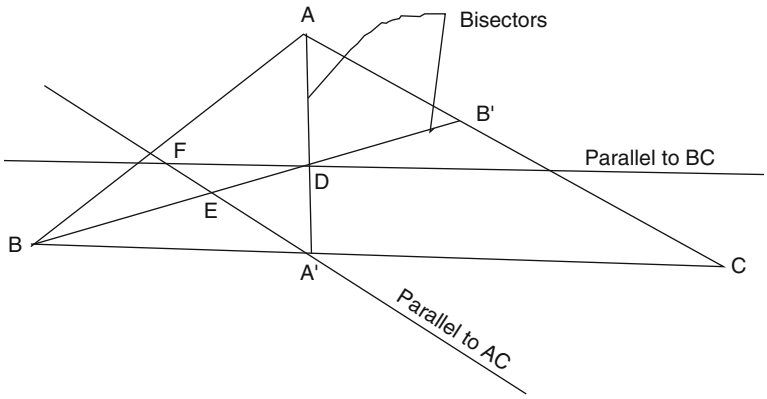


Fig. 2.19 Subdivision of triangular area

$pA = 31$, $pB = 50$, and $pC = 19$. In Fig. 2.19, point D on the A-A′ bisector corresponds to 31 % of the A-A′ length starting from A′ which lies on the B-C side of the triangle representing zero percentage. A parallel line to side B–C is drawn from point D. Likewise, the next percentage is considered for the precipitation amount at apex B. The bisector B-B′ is drawn starting from point B′ on the side A–C toward B. On the bisector, point E corresponding to 50 % is depicted. A parallel line from this point to the side AC is drawn. Finally, the intersection of these two parallels at point F defines the "key point" for triangle ABC. The following steps are necessary for the implementation of the PWP:

1. Three adjacent stations are considered such as in Fig. 2.19 where each apex is coded by its longitude (X_A, X_B, and X_C), latitude (Y_A, Y_B, and Y_C), and precipitation value (Z_A, Z_B, and Z_C).
2. The slopes (m_{AB}, m_{BC}, and m_{CA}) of triangle sides are calculated by making use of the apices' coordinates.
3. Determine the straight-line equations perpendicular to each of the sides but crossing from the opposite apex by analytical formulations. First of all, the coordinates of the projection point such as A', and B' of each apex on the opposite side must be calculated. For instance, the coordinates $X_A^{'}$ and $Y_A^{'}$ of point A' can be expressed in terms of known quantities as

$$X'_A = \frac{1}{m_{BC}^2 + 1}\left[m_{BC}^2(X_B) + m_{BC}(Y_A - Y_B) + X_A\right] \tag{2.43}$$

$$Y'_A = \frac{1}{m_{BC}}\left(X_A - X'_A\right) + Y_A \tag{2.44}$$

where

$$m_{BC} = \frac{Y_B - Y_C}{X_B - X_C} \tag{2.45}$$

Similarly, the coordinates $X_B{}'$ and $Y_B{}'$ of point B' on AC side are

$$X'_B = \frac{1}{m_{CA}^2 + 1}\left[m_{CA}^2(X_C) + m_{CA}(Y_B - Y_C) + X_B\right] \tag{2.46}$$

$$Y'_B = \frac{1}{m_{CA}}\left(X_B - X'_B\right) + Y_B \tag{2.47}$$

where

$$m_{CA} = \frac{Y_C - Y_A}{X_C - X_A} \tag{2.48}$$

4. The precipitation amounts (Z_A, Z_B, and Z_C) are used in order to find points along the perpendiculars starting from the side toward the apex which divide each one in proportions λ_1, λ_2, and λ_3. By use of these ratios and the previous known quantities, the coordinates of points A'', B'', and C'' are determined. For instance, the coordinates $X_A{}''$ and $Y_A{}''$ are defined as

$$X''_A = X'_A + \lambda_1\left(X_A - X'_A\right) \tag{2.49}$$

$$Y''_A = Y'_A + \lambda_2\left(Y_A - Y'_A\right) \tag{2.50}$$

5. On the other hand, from the precipitation measurements at the three apices of any triangle, the total precipitation T value becomes

$$T = Z_A + Z_B + Z_C$$

which can be written in terms of percentages as

$$Z_A/T + Z_B/T + Z_C/T = 1.0$$

or

$$\lambda_1 + \lambda_2 + \lambda_3 = 1.0$$

Hence, lambda values are defined as the percentage of precipitation amounts at the three apices of a triangle. Similarly, the coordinates of point B'' can be written with the relevant notations as

$$X''_B = X'_B + \lambda_2\left(X_B - X'_B\right) \tag{2.51}$$

$$Y''_B = Y'_B + \lambda_2\left(Y_B - Y'_B\right) \tag{2.52}$$

where

$$\lambda_2 = \frac{Z_B}{Z_A + Z_B + Z_C} \tag{2.53}$$

6. The straight lines are drawn that pass through point A'' and parallel to the BC side of the triangle and similarly parallel to AC crossing through B''. Let these parallel lines be D_1 and D_2. The equations of these lines are given as follows:

$$D_1 = m_{BC}X + \left(Y''_{A''} - m_{BC}X''_{A''}\right) \tag{2.54}$$

$$D_2 = m_{CA}X + \left(Y''_{B''} - m_{CA}X''_{B''}\right) \tag{2.55}$$

7. The intersection point of lines D_1 and D_2 is referred to as the "key point," M_K, for triangle ABC. The coordinates of this key point (X_K, Y_K) can be found by simultaneous solution of Eqs. 2.54 and 2.55.
8. Previous steps are repeated for each triangle representing the subarea within the whole catchment.
9. Adjacent key points are connected with each other hence constituting polygons each enclosing a single meteorology station. The key point coordinates become

$$X_k = \frac{1}{m_{CA} - m_{BC}}\left[\left(m_{CA}X''_B - m_{BC}X''_A\right) - \left(Y''_B - Y''_A\right)\right] \tag{2.56}$$

$$Y_k = m_{BC}\left(X_k - X''_A\right) + Y''_A \tag{2.57}$$

10. The area of each polygon is calculated by one of the available convenient mathematical methods.
11. Multiplication of each polygon area by its representative station precipitation value gives the volume of water that is gathered over this polygonal subarea.
12. Summation of these volumes for all the relevant polygons that cover the whole catchment area gives the total water volume that falls over the catchment area.
13. Division of this total water volume by the total catchment area yields the AAE value for the catchment.

In general, the comparison of the Thiessen and PWP methods indicates the following significant points:

1. If triangulation of stations produces obtuse triangles, then the Thiessen method gives bisector intersections that fall outside the triangle area. However, in the PWP method, all the key points lie within the triangle itself. In this way the common effect of precipitation amounts at triangle apices falls within the triangle domain.
2. In the Thiessen method, the station lies within the subareal polygon at almost equal distances from the sides. This implies that almost equal weights are given to precipitation amounts in the separation of polygons. On the contrary, the PWP method produces polygon sides closer to the station depending on the relative

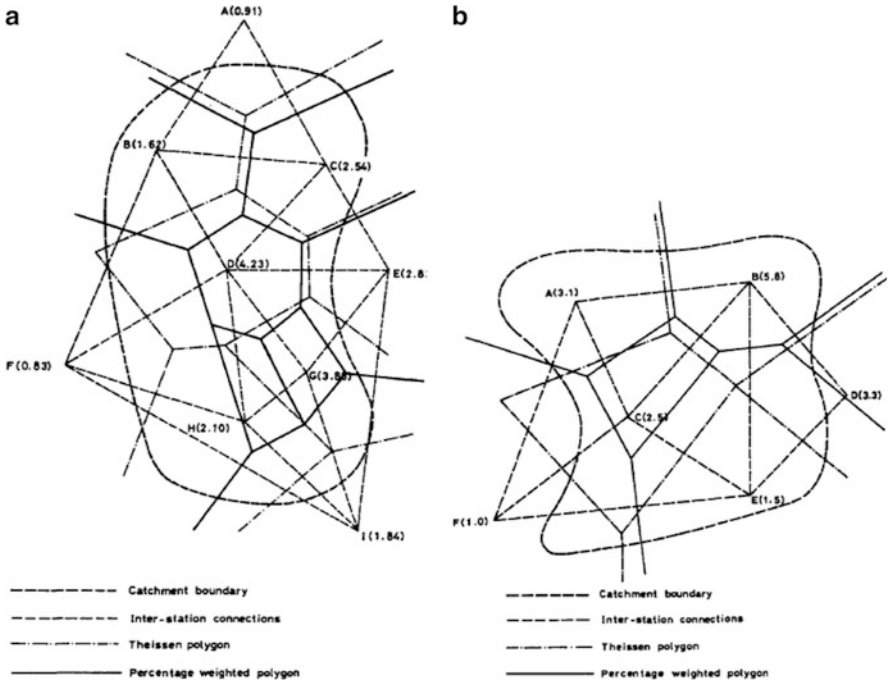


Fig. 2.20 Precipitation data. **a** Wiesner, **b** Bruce and Clark

magnitude of precipitation amounts between adjacent stations. This implies that in the PWP method, the center of recorded precipitation amounts at three neighboring stations (not the geometrical center of station locations) plays the most significant role in the areal partition calculations.

3. The size of any closed polygon for a station in the PWP method is always smaller than the corresponding Thiessen method. Hence, more refined partition of the catchment area into subareas is obtained. This implies that more refined AAE values result by use of the PWP method. In fact, in the PWP method, the biggest precipitation record station results in the smallest subarea.
4. Variations in the precipitation amounts, due to some different storms, change the polygonal partition in the PWP method, whereas Thiessen polygons remain the same.

Example 2.9 The implementation of PWP method was applied first to spatial precipitation data given by Wiesner (1970) and Bruce and Clark (1966) as in Fig. 2.20, where the geometric configuration of precipitation stations is given for the same data according to the Thiessen and PWP methods. In order to assess the reliability of these methods, the average sum of square deviations (ASSD) from the AAE value is calculated for each method as

$$\text{ASSD} = \frac{1}{n}\sum_{i=1}^{n}(P_i - \text{AAE})^2 \tag{2.58}$$

Table 2.9 AAE of precipitation

	Wiesner data		Bruce and Clark data	
Method	AAE (inches)	ASSD (inches^2)	AAE (inches)	ASSD (inches^2)
Thiessen	2.67	1.34	3.10	2.54
PWP	2.45	1.15	2.50	2.37

where n is the number of stations. The results for various methods are shown in Table 2.9.

The PWP method yields smaller AAE and ASSD values than the Thiessen method.

As mentioned earlier, with the Thiessen method, the obtuse triangle BDF, a perpendicular bisector intersection appears to the left of BF line (see Fig. 2.20a). By considering precipitation amounts at stations B and F, it is not possible to expect a value of about 4 inches. On the other hand, in the PWP method, there is no such confusion. Wiesner (1970) has shown after completion of an isohyetal map for the same precipitation data that the storm peak occurs within a small area around station D in Fig. 2.20a. This is also confirmed by the PWP method as shown in Fig. 2.20a. Again, in the same figure, the Thiessen method yields only one closed polygon, whereas PWP provides three such polygons. Extensions of Thiessen boundary polygons between the pairs of stations F and I and E and I suffer from the same drawbacks in that it is not possible to have precipitation values greater than 1.8 inches between F and I and more than 2.83 inches between E and I. Similar general interpretations are valid also for Fig. 2.20b where there are two odd bisectors' intersection points which lie outside the triangles resulting from the application of the Thiessen method. These lie to the left of line AF and below line FE. Comparatively, even in this figure, the closed polygon from the PWP method is smaller in size than the Thiessen case. The PWP method can be applied to local or regional scales due to the following reasons:

1. The assumption that convective, intense precipitation is comparatively limited in areal extent and duration is valid for small or regional scales. Thinking in larger space and time scales (e.g., global and monthly), there are also broader areas with intense precipitation, such as the large-scale convergence zones (ITCZ and SPCZ) and the monsoon regions.
2. The limited areal extent and duration of intense convective precipitation compared to stratiform, light precipitation are generally compensated by the fact that many storms might be missed by the rain gauge network (see "Introduction" section). So if the rain gauge network is nearly equally distributed, there is not any reason to compensate for this a second time by reducing the areal weight of heavier precipitation, unless there is not really an indication that intense precipitation is overrepresented due to the spatial distribution of the rain gauges in the network. This might be the case in arid regions, because there are almost no stations in desert regions However, in mountainous areas precipitation is often

Table 2.10 Total monthly rainfall values of stations (mm)

Stations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
Silvan	104	116	116	84	61	6	0	1	1	19	69	116
Siirt	115	107	111	106	66	9	1	0	5	48	86	102
Ahlat	64	67	79	84	62	23	9	4	15	50	68	53
Bitlis	180	156	141	107	65	19	5	3	14	58	97	130
Van	42	35	46	58	41	17	6	3	12	44	49	83
Şırnak	136	121	143	141	64	5	2	1	6	33	83	123
Başkale	47	43	69	102	100	42	20	9	16	39	43	38
Nusaybin	92	62	63	63	40	1	1	0	1	16	43	79
Şemdinli	106	155	151	195	56	9	3	4	5	77	119	97
Hakkari	102	105	125	146	57	15	3	1	10	26	76	91

underestimated, because most rain gauge stations are located in valleys and do not catch very well the heavy precipitation on the slopes or mountain peaks.

3. The proposed PWP method is very convenient in areas where the precipitation stations are irregularly located.

Example 2.10 In order to determine the AAE of rainfall from ten different meteorology stations, the PWP method is applied together with the other conventional methods (arithmetic mean, Thiessen polygon, isohyetal map technique) to the Southeastern Anatolia Region of Turkey (Bayraktar et al. 2005). Total monthly rainfall values of these stations in 1993 are used and presented in Table 2.10.

1/500 000 scaled maps are used to draw the subareas, which are measured with a planimetry. For each method, AAR values are calculated with the help of Eq. 2.27.

Application map of Thiessen method and areal values of Thiessen polygons are merely given in Fig. 2.21, because the key points remain the same as long as the meteorology stations do not change. Monthly isohyetal maps are prepared for January, April, July, and October as in Fig. 2.22.

For the same months, rainfall values and subareas are given in Fig. 2.23a–d. In the PWP method, which is the main purpose of this study, values of rainfall and percentage weightings are calculated for each of the three adjacent stations constituting sub-triangles.

During the application, PWP rainfall values are calculated by considering Eqs. 2.40, 2.41 and 2.42 are used for determination of subareas. PWP method calculations and subareal values are given in Fig. 2.23a–d. For the comparison of different methods, all results are presented collectively in Table 2.11.

As can be seen from this application, the PWP method yields more reliable results and smaller AAE values depending on the regional variability of the rainfall amounts over the catchment area. For instance, in July, rainfall values have many regional variations. This is due to the semiarid characteristics of the study area and frequent occurrence of convective rainfalls, which are not expected to cover large areas. It is noted that in July, Başkale station surrounding

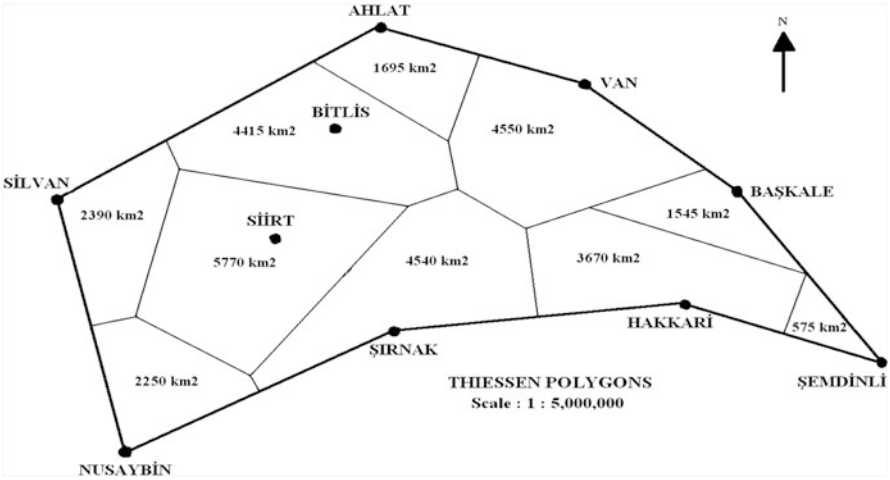


Fig. 2.21 Application map of Thiessen method (January, April, July, October 1993)

appears as the most rainfall reception area, which is represented by a subarea of 1545 km^2 in the Thiessen method and by a subarea of 3140 km^2 and 519 km^2 in the arithmetic mean and the isohyetal map technique, respectively. On the other hand, it is represented by a subarea of 153 km^2 when PWP method is considered. Hence, the representation of the most rainfall intercepting meteorology station with the smallest subarea gives rise to a smaller value of AAE for the catchment area during convective storms. The monthly AAEs of the catchment area in this month according to isohyetal map, Thiessen polygon, and arithmetic mean techniques are estimated as 44 %, 41 %, and 54 %, respectively. They are smaller than the PWP method AAE value. The areal rainfall variability in October is comparatively less than July because of the north polar maritime air mass penetration into the study area which gives rise to frontal rainfall occurrences. It is well known that the frontal rainfalls have more extensive areal coverage than the convective rainfall types and consequently the percentages are smaller in the PWP method calculations. This is tantamount to saying that percentage calculations from the three adjacent stations are not very different as they are in the convective rainfall events. It is timely to state that similar effects can be expected from the orographic rainfall occurrences to the convective rainfall types in mountainous areas. Such distinctions cannot be considered in the Thiessen approach, where the subareas are determined according to the geometric configuration of the station locations without the consideration of actually recorded rainfall amounts at these stations. For example, in October, the AAE calculations by using the PWP method yielded 13 %, 15 %, and 18 % smaller values than other three conventional methods. This is due to the lesser areal variability of rainfall in this month than in July. Furthermore, the AAE rainfall amounts by use of the isohyet, Thiessen and PWP methods are 12 %, 14 % and 14 %, respectively.

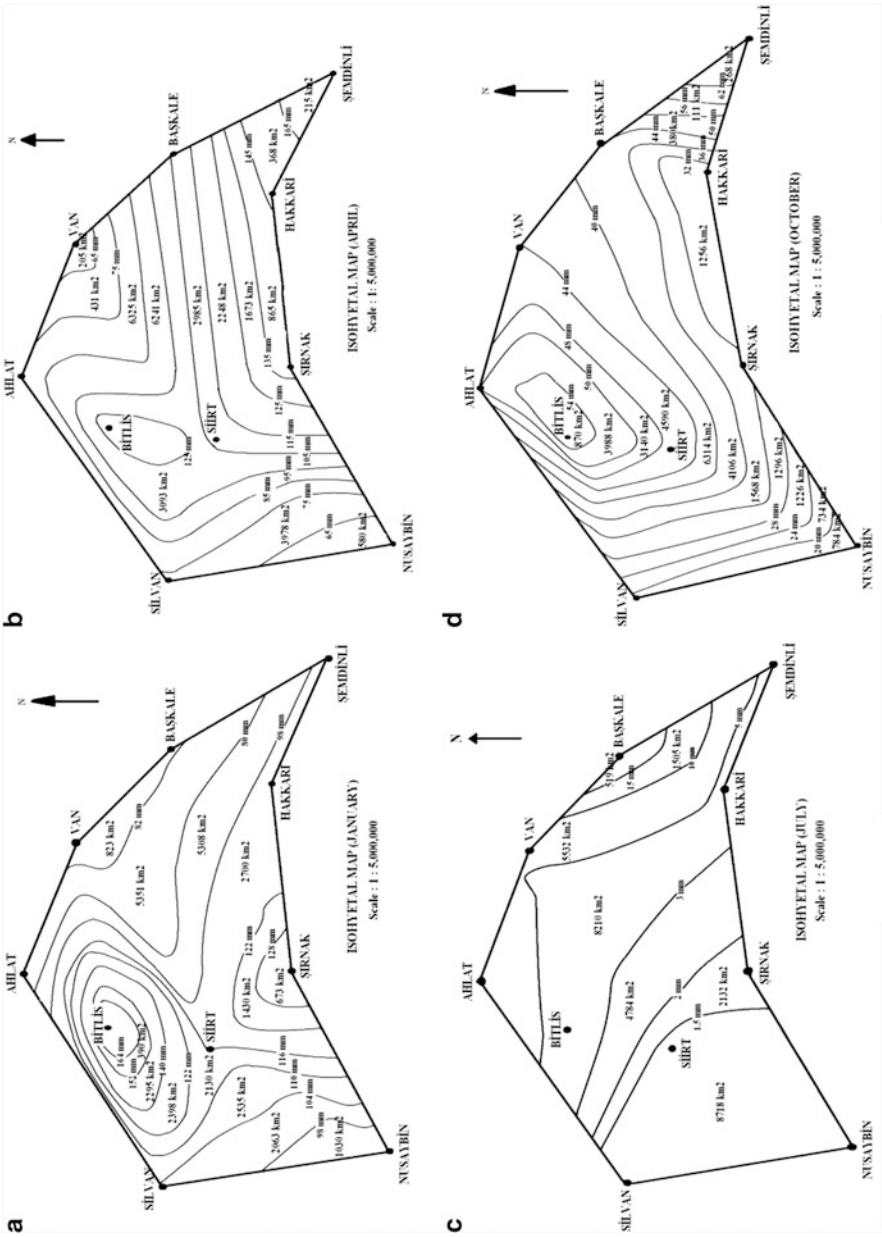


Fig. 2.22 Application map of isohyetal technique. (**a**) January, (**b**) April, (**c**) July, (**d**) October

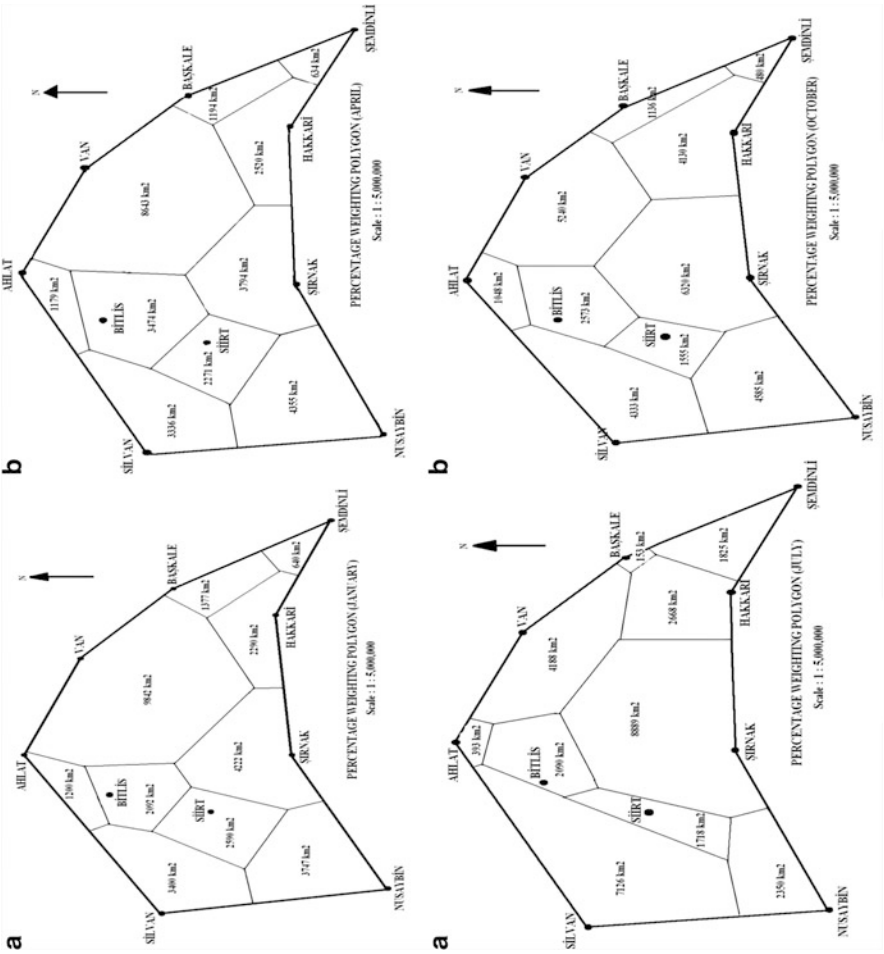


Fig. 2.23 Application map of PWP technique. (**a**) January, (**b**) April, (**c**) July, (**d**) October

Table 2.11 AAE value for each method (mm)

Method	January	April	July	October	Annual
Percentage weighting polygon	89.22	92.70	2.32	33.75	54.49
Isohyetal map	102.72	102.72	4.12	38.96	62.13
Thiessen polygon	106.30	104.30	3.94	39.74	63.57
The arithmetic average	98.76	108.49	5.00	40.97	63.31

As the frontal and more homogeneous type of rainfalls are recorded during a specific period, then it is expected that the results of the PWP and Thiessen polygon methodologies will approach to each other, provided that there are approximate rainfall records at each meteorology station. However, in practice the use of PWP method is recommended. In fact, whatever the circumstances are, it is preferable to adopt PWP method since it yields almost the same result as the Thiessen method if the conditions are satisfied as the homogeneous type of rainfalls.

2.9 Areal Coverage Probability

A rigorous probabilistic methodology is developed on the basis of random field concept for determining various probability of precipitation (PoP) areal coverage probability (ACP) over a region. The necessary analytical formulations are derived with the restrictive assumption that the precipitation occurrences at different sites and times are independent from each other. However, the regional heterogeneity in precipitation occurrence is exemplified by either nonidentical probability of precipitation values or nonuniform threshold levels. Finally, extreme value probabilities of precipitation coverage area are formulated in their most general forms. The probability statements derived are helpful in predicting the regional probable potential future precipitation or drought coverage areas. The drought basic time unit can be adopted as hour, day, week, month, or year in addition to unit area for spatial analysis. If the basic unit is day, then each day can be labeled as "rainy" and "non-rainy" and accordingly the calculations are performed (Şen 2008). Similarly, in the case of regular division of an area into a set of subareas as in Fig. 2.24b, each subunit may have either a "dry" or "wet" spell. For instance, in Fig. 2.24a, b, there are six and nine time and subarea units, respectively.

If "dry" units are shown by black and "wet" units by white colors, then in Fig. 2.25, there are four and five time and subareal dry units, respectively.

As the number of basic unit increases, then the "dry," D, and "wet," W, time and spatial features start to become clearer as in Fig. 2.26.

In this section precipitation phenomenon is modeled as a random field where timewise probabilities at a fixed site are referred to as the PoP and spatial

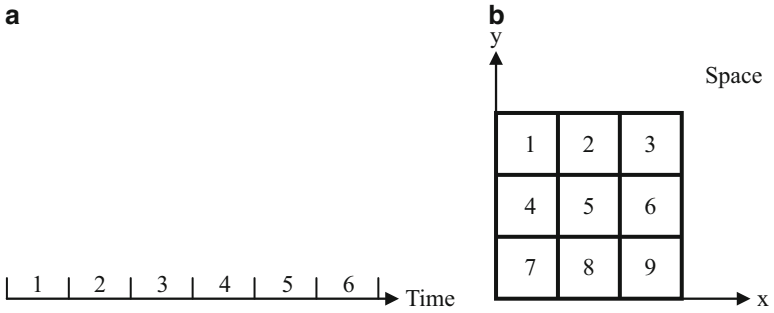


Fig. 2.24 Time and spatial units

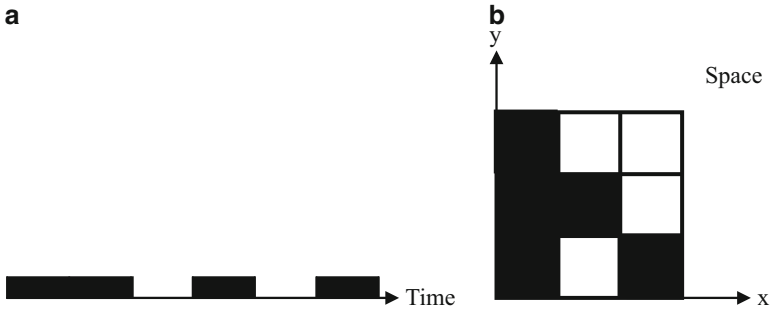


Fig. 2.25 Dry time and subareal units

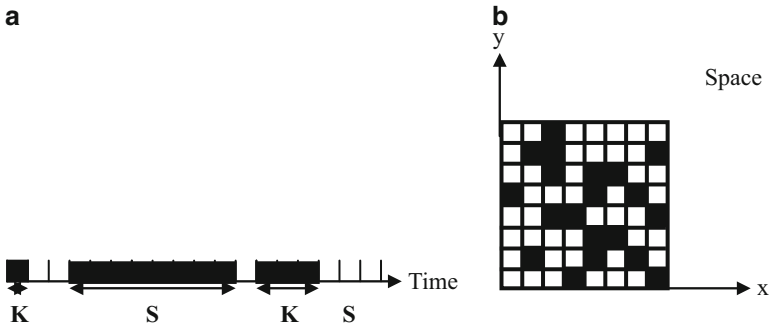


Fig. 2.26 Dry durations and areas

probabilities for a fixed time instant are ACP. The areal probability is, in fact, the fraction of the area hit by rainfall. It does not provide a means for depicting which of the subareas is going to be hit with precipitation event. However, it simply represents the estimate of what the fractional coverage will

be. Furthermore, it is clearly a conditional probability being conditioned by whether at least one precipitation event actually occurs in the subarea during a certain time interval.

The PoP at any desired threshold value, x_0, such as standard 0.01 inches is equivalent to the exceedance probability of this value. If the PDF of precipitation at site i is denoted by $f_i(X)$, then the PoP, p_i, at this site is

$$p_i = \int_{x_0}^{\infty} f_i(X)dX$$

If the forecast area is thought in terms of n sample sites, then the average areal probability p can be defined formally as

$$p = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{2.59}$$

In the special case where the point probability is uniform over the forecast area, p is equal to the common value of PoP. Note that since p is a lumped value, it does not, in general, contain detailed information provided by the set of individual point probabilities. However, if for some reason some of the subareas vary in size, then each probability value must be percentage area weighted. It should be noted that the areas associated with subareas need to be small enough that a single probability value can be applied to each one.

2.9.1 Theoretical Treatment

In the most general case, none of the sites have equal PoPs which implies that the random field is heterogeneous. It is quite likely that probabilities might vary from place to place even within a single area. In practice, in addition to the spatial correlation variations, the following three situations give rise to heterogeneous and anisotropic random fields. These are:

1. Identical PDFs of precipitation at different sites but nonuniform threshold levels
2. Nonidentical PDF of precipitation at sites but uniform threshold level
3. Nonidentical PDFs at sites and nonuniform threshold levels

Let the PoP and its complementary probability at jth site within a region of n sites be given by p_j and $q_j, (j = 1, 2, \ldots, n)$, respectively. They are pointwise mutually exclusive and complementary, i.e., $p_j + q_j = 1$. The ACP, P($A = i$)

including i sites, can be evaluated through enumeration technique after the application of summation and multiplication rules of probability theory. Hence,

$$P(A=i)=\frac{1}{i!}\sum_{\substack{j_1=1\\ j_2\neq j_1}}^{n}p_{j1}\sum_{\substack{j_2=j_1\\ j_i\neq j_{i-1}}}^{n}p_{j2}\ \cdot\ \cdot\ \sum_{\substack{j_i\\ k\neq j_i}}^{n}p_{j_i}\ \prod_{k=1}^{n}q_k \tag{2.60}$$

or succinctly

$$P(A=i)=\frac{1}{i!}\sum_{\substack{j_1=1\\ j_2\neq j_1}}^{n}p_{j1}\sum_{\substack{j_2=j_1\\ j_i\neq j_{i-1}k\neq j_i}}^{n}p_{j2}\cdot\cdot\sum_{j_i}^{n}p_{j_i}\ \prod_{k=1}^{n}q_k \tag{2.61}$$

$$P(A=i)=\frac{1}{i!}\left(\prod_{\substack{k_1=1\\ j_{k_1}\neq j_m}}^{n}\sum_{\substack{j_{k_1}=1\\ k_2\neq j_m\\ m=1,2,\ \cdot\ \cdot\ j_{k_{m-1}}}}^{n}p_{j_{k_1}}\right)\ \prod_{k_2=1}^{n}q_{k_2} \tag{2.62}$$

where the multiplication of i summations in the brackets includes all the possible combinations of i precipitation occurrences at n sites, whereas the last multiplication term corresponds to possible no-precipitation combinations. For identical PoPs, the term in brackets simplifies to $n(n-1)..(n-i+1)p^i$ and the last multiplication to q^{n-i}; hence, Eq. 2.62 reduces to

$$P(A=i)=\binom{n}{i}p^i q^{n-i} \tag{2.63}$$

which is actually the binomial distribution with two-stage Bernoulli trials (Feller 1957).

This is the well-known binomial distribution with mean np and variance npq. The probability, p_A, that all the sites, hence area, are covered by precipitation at an instant can be found from Eq. 2.62 as

$$p_A=\prod_{i=1}^{n}p_i \tag{2.64}$$

Comparison of Eqs. 2.59 and 2.64 yields

$$p>p_A \tag{2.65}$$

and accordingly

$$\min(p_1,p_2,\ldots..,p_n)<p<\max(p_1,p_2,\ldots..,p_n) \tag{2.66}$$

The conditional probability of partial area, given a group of randomly or systematically selected k precipitation occurrences or nonoccurrences, is of great interest in the practical applications. The joint probability that any k_1 of n sites have significant precipitation occurrences can be defined by virtue of independence and hence multiplication rule of the probability theory as

$$P(X_1 > x_0, X_2 > x_0, \ldots, X_{k_1} > x_0) = p^{k_1} \tag{2.67}$$

Assuming that any one of the sites is equally likely to experience a precipitation, the conditional probability that only a certain group of i sites has precipitation can be written as

$$P(X_1 > x_0, X_2 > x_0, \ldots, X_{k_1} > x_0 / A = i) = \left(\frac{i}{n}\right)^{k_1}$$

By definition the joint probability becomes

$$P(X_1 > x_0, X_2 > x_0, \ldots, X_{k_1} > x_0, A = i) = \left(\frac{i}{n}\right)^{k_1} P(A = i)$$

The conditional ACP of precipitation can then be obtained after dividing this last expression by Eq. 2.67 which leads to

$$P(A = i / X_1 > x_0, X_2 > x_0, \ldots, X_{k_1} > x_0) = \left(\frac{i}{np}\right)^{k_1} P(A = i) \tag{2.68}$$

Hence, in order to obtain the most general case of heterogeneous PoP's conditional areal coverage, it is necessary to substitute Eq. 2.62 into Eq. 2.68. On the other hand, for homogeneously distributed PoPs, Eq. 2.68 takes its explicit form by substitution of Eq. 2.63:

$$P(A = i / X_1 > x_0, X_2 > x_0, \ldots, X_{k_1} > x_0) = \left(\frac{i}{np}\right)^{k_1} P(A = i) \tag{2.69}$$

Since some initial information is given with certainty, the conditional probability in Eq. 2.69 has more information content than the original unconditional distribution in Eq. 2.63. This fact can be objectively shown by considering the variances of PDFs in Eqs. (2.63) and (2.69) which are npq and (n–1)pq, respectively. Hence, the conditional variance is smaller; therefore, conditional ACP is more certain. However, the unconditional areal precipitation coverage expectation from Eq. 2.63 is equal to np, whereas the conditional coverage area expectation is greater and can be obtained from Eq. 2.68 as np + q.

Similarly, the conditional ACP of precipitation given that a group of k_2 sites have no precipitation can be found for homogeneous PoPs as

$$P(A = i/X_1 < x_0, X_2 < x_0, \ldots, X_{k_2} < x_0,) = \left(\frac{n-i}{\mathrm{nq}}\right)\binom{n}{i} q^i p^{n-1} \tag{2.70}$$

Finally, the conditional ACP of precipitation given that a group of k_1 sites has precipitation and another group of k_2 sites has no precipitation is obtained as

$$P\big(A = i/X_1 > x_0, X_2 > x_0, \ldots, X_{k_1} > x_0, X_{k_{1+1}} > x_0, X_{k_1+2} < x_0, \ldots X_{k_2} < x_0,\big) = \left(\frac{i}{\mathrm{nq}}\right)\left(\frac{n-i}{\mathrm{np}}\right)\binom{n}{i} q^i p^{n-1} \tag{2.71}$$

The probability expressions in Eqs. 2.68, 2.69, 2.70 and 2.71 can be effectively employed to study regional precipitation occurrence patterns.

2.9.2 Extreme Value Probabilities

In practice, it is also important to know the probability of maximum and minimum areal coverage of precipitation. The probability of maximum areal coverage, A_M, of precipitation can be found provided that the precipitation coverage area evolution along the time axis is assumed to be independent from each other. In general, for m time instances, the probability of maximum areal coverage of precipitation to be less than or equal to an integer value, i, can be written as

$$P(A_M \leq i) = [P(A \leq i)]^m \tag{2.72}$$

where $P(A \leq i)$ can be evaluated, in general, from Eq. 2.62 for heterogeneous PoPs by employing

$$P(A \leq i) = \sum_{j=0}^{i} P(A = j)$$

The substitution of which into Eq. 2.72 leads to

$$P(A_M \leq i) = \left[\left(\sum_{j=1}^{i} P(A = j)\right)\right]^m$$

However, it is a well-known fact that $P(A_M = i) = P(A_M \leq i) - P(A_M \leq i-1)$, and therefore, one can obtain consequently that

$$P(A_M = i) = \left[\sum_{j=0}^{i}\binom{n}{j} p^j q^{n-j}\right]^m - \left[\sum_{j=0}^{i-1}\binom{n}{j} p^j q^{n-j}\right]^m \tag{2.73}$$

It is then possible to rewrite Eq. 2.73 for homogeneous PoPs from Eq. 2.63 explicitly as

$$P(A_M = i) = \left[\sum_{j=0}^{i} \binom{n}{j} p^j q^{n-j}\right]^m - \left[\sum_{j=0}^{i-1} \binom{n}{j} p^j q^{n-j}\right]^m$$

Hence, the probability $P(A_M = n)$ that the whole area is covered by precipitation can be obtained from Eq. 2.70 as

$$P(A_M = n) = 1 - (1 - p^n)^m \tag{2.74}$$

One can interpret from this expression that for small regions, the number of sites is also small, in which case the probability in Eq. 2.74 is not zero, and there is a chance for the whole area to be covered by precipitation. Similarly, the probability of minimum areal coverage, A_m, of precipitation can be written for homogeneous PoPs as

$$P(A_m = i) = \left[\sum_{j=i}^{n} \binom{n}{j} p^j q^{n-j}\right]^m - \left[\sum_{j=i+1}^{n} \binom{n}{j} p^j q^{n-j}\right]^m \tag{2.75}$$

2.10 Spatio-Temporal Drought Theory and Analysis

Let an agricultural land be divided into m mutually exclusive subareas each with the same spatial and temporal drought chance. The Bernoulli distribution theory can be employed to find the extent of drought area, A_d, during a time interval, Δt. The probability of n_1 subareas stricken by drought can be written according to Bernoulli distribution as (Feller 1957)

$$P_{\Delta t}(A_d = n_1) = \binom{m}{n_1} p_r^{n_1} q_r^{m-n_1} \qquad p_r + q_r = 1.0 \tag{2.76}$$

This implies that out of m possible drought-prone subareas, n_1 have deficit and hence the areal coverage of drought is equal to n_1 or in percentages n_1/m. For the subsequent time Δt interval, there are $(m–n_1)$ drought-prone subareas. Assuming that the evolution of possible deficit and surplus spells along the time axis is independent over mutually exclusive subareas, similar to the concept in Eq. 2.76, it is possible to write for the second time interval that

$$P_{2\Delta t}(A_d = n_2) = \sum_{n_1=0}^{n_2} \binom{m}{n_1} \binom{m-n_1}{n_2-n_1} p_r^{n_1} p_r^{n_2-n_1} q_r^{m-n_1} q_r^{m-n_2} \tag{2.77}$$

where n_2 is the total number of drought-affected subareas during the second time interval. By considering Eq. 2.76, this expression can be rewritten succinctly in the form of recurrence relationship as

$$P_{2\Delta t}(A_d = n_2) = \sum_{n_1=0}^{n_2} P_{\Delta t}(A_d = n_1) P(A_d = n_2 - n_1) \tag{2.78}$$

where $P(A_d = n_2 - n_1)$ is the probability of additional $(n_2 - n_1)$ subareas to be affected by deficit during the second time interval out of remaining (m–n_1) potential subareas from the previous time interval. With the same logic, extension of Eq. 2.78 for any successive time interval, i, furnishes all the required drought area probabilities as

$$P_{i\Delta t}(A_d = n_i) = \sum_{n_{i-1}=0}^{n_i} P_{(i-1)\Delta t}(A_d = n_{i-1}) P(A_d = n_i - n_{i-1}) \tag{2.79}$$

For $i = 1$, this expression reduces to its simplest case which does not consider time variability of drought occurrences as presented by Şen (1980) and experimentally on digital computers by Tase (1976). Furthermore, the probability of agricultural drought area to be equal to or less than a specific number of subareas j can be evaluated from Eq. 2.79 according to

$$P_{i\Delta t}(A_d \le j) = \sum_{k=0}^{j} P_{i\Delta t}(A_d = k) \tag{2.80}$$

The probability of having, $n_1^{'}$, deficit subareas given that there are n_1 deficit subareas at the beginning of the same time interval within the whole region can be expressed as

$$P_{\Delta t}\left(A_d = n_1' | A_d = n_1\right) = \binom{m}{n_1} \binom{n_1}{n_1 - n_1'} p_r^{n_1} p_t^{n_1 - n_1'} q_r^{m-n} q_t^{n_1'}$$

or, shortly,

$$P_{\Delta t}\left(A_d = n_1' | A_d = n_1\right) = P_{\Delta t}(A_d = n_1) \binom{n_1}{n_1 - n_1'} p_t^{n_1 - n_1'} q_t^{n_1'} \tag{2.81}$$

It should be noted that always $n_1 \ge n_1'$ and the difference, $j = n_1 - n_1'$ gives the number of transitions. On the basis of Eq. 2.81, a general equation for the marginal probability of observing $n_1^{'}$ deficit spells at the end of the same time interval, after simple algebra, becomes

$$P_{\Delta t}\left(A_d = n_1'\right) = \sum_{k=0}^{m=n_i'} P_{i\Delta t}\left(A_d = k + n_i'\right) \binom{k = n_i'}{k} p_t^k q_t^{n_i'} \tag{2.82}$$

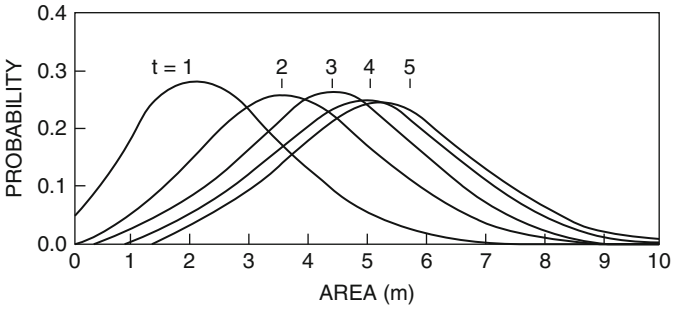


Fig. 2.27 Probability of drought area for multi-seasonal model ($m = 10$; $p_r = 0.3$; $p_t = 0.2$)

Hence, the regional agricultural drought occurrences during the second time interval follow similarly to this last expression, and generally, for the ith step, it takes the following form:

$$P_{i\Delta t}(A_d = n_i') = \sum_{k=0}^{m=n_i'} P_{i\Delta t}(A_d = k + n_i') \binom{k + n_i'}{k} p_t^k q_t^{n_i'} \tag{2.83}$$

Its validity has been verified using digital computers. The PDFs of areal agricultural droughts for this model are shown in Fig. 2.27 with parameters $m = 10, p_r = 0.3$, $p_t = 0.2$ and $i = 1,2,3,4$ and 5.

The probability functions exhibit almost symmetrical forms irrespective of time intervals although they have very small positive skewness.

Another version of the multi-seasonal model is interesting when the number of continuously deficit subareas appears along the whole observation period. In such a case, the probability of drought area in the first time interval can be calculated from Eq. 2.76. At the end of the second time interval, the probability of j subareas with two successive deficits given that already n_1 subareas had SMC deficit in the previous interval can be expressed as

$$P_{2\Delta t}(A_d = j | A_d = n_1) = P\Delta t(A_d = n_1) \binom{n_1}{j} p_t^j q_t^{n_1 - j} \tag{2.84}$$

This expression yields the probability of having n_1 subareas to have deficit out of which j subareas are hit by two deficits, i.e., there are $(n_1 - j)$ subareas with one deficit. Hence, the marginal probability of continuous deficit subarea numbers is

$$P_{2\Delta t}(A_d = j) = \sum_{k=0}^{m-j} P_{\Delta t}(A_d = k + j) \binom{k + j}{j} p_t^j q_t^k$$

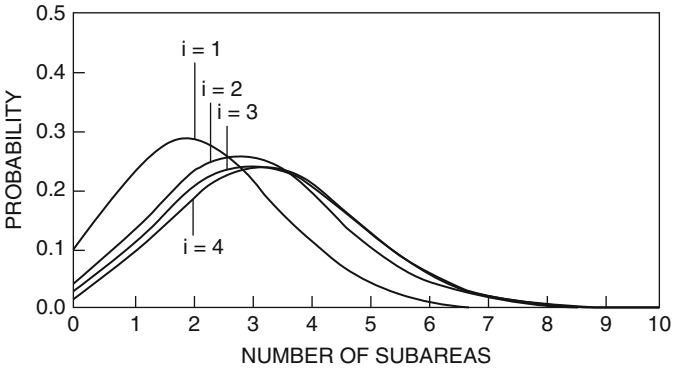


Fig. 2.28 Drought area probability for multi-seasonal model ($m = 10;\ p_r = 0.3;\ p_t = 0.5$)

In general, for the ith time interval, it is possible to write

$$P_{i\Delta t}(A_d = j) = \sum_{k=0}^{m-j} P_{(i-1)\Delta t}(A_d = k + j)\binom{k+j}{j} p_t^j q_t^k \tag{2.85}$$

The numerical solutions of this expression are presented in Fig. 2.28 for $m = 10, p_r = 0.3$ and $p_t = 0.5$.. The probability distribution function is positively skewed.

2.10.1 Drought Parameters

The global assessment of model performances can be achieved on the basis of drought parameters such as averages, i.e., expectations and variances, but for drought predictions, the PDF expressions as derived above are significant. The expected, i.e., average number of deficits, $E_i(A_d)$, over a region of m subareas during time interval, $i\Delta t$, is defined as

$$E_i(A_d) = \sum_{k=0}^{m} kP_{i\Delta t}(A_d = k) \tag{2.86}$$

Similarly, the variance, $V_i(A_d)$, of drought-affected area is given by definition as

$$V_i(A_d) = \sum_{k=0}^{m} k^2 P_{i\Delta t}(A_d = k) - E_i^2(A_d) \tag{2.87}$$

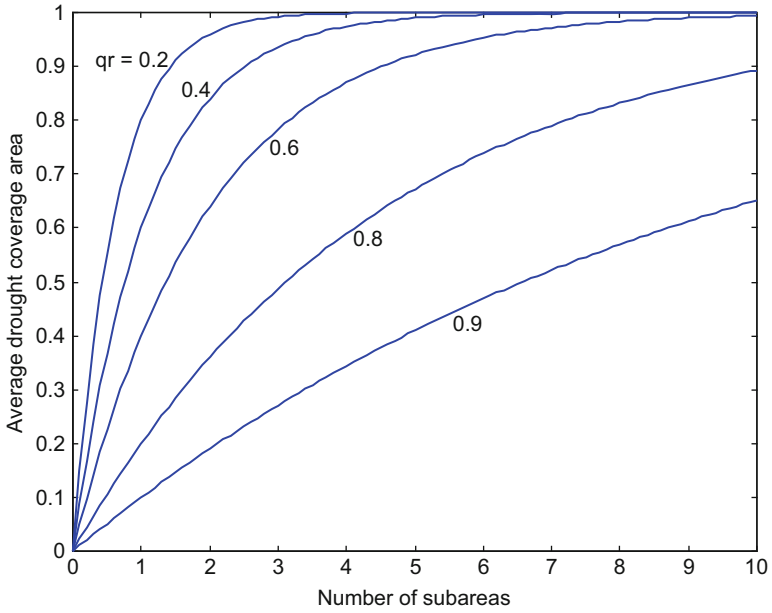


Fig. 2.29 Drought percentage areal coverage

Substitution of Eq. 2.79 into Eq. 2.86 leads to drought-stricken average area within the whole region as

$$E_i(A_d) = mp_r \sum_{k=0}^{i-1} q_r^k \tag{2.88}$$

or, succinctly,

$$E_i(A_d) = m\left(1 - q_r^i\right) \tag{2.89}$$

Furthermore, the percentage of agricultural drought-stricken area, P_A^i, can be calculated by dividing both sides by the total number of subareas, m, leading to

$$P_A^i = \left(1 - q_r^i\right) \tag{2.90}$$

Figure 2.29 shows the change of drought-stricken area percentage with the number of deficit subareas, i, for given deficit probability, q_r.

For regional drought variations in the first time interval ($i = 1$), from Eq. 2.90, $P_A^1 = p_r$. On the other hand, for the whole area to be covered by drought theoretically, $i \rightarrow$c ∞, and therefore, $P_A^\infty = 1$. It is obvious that the temporal agricultural drought percentage for a region of m subareas at any time, i, is $p_r \leq P_A^i \leq 1$.

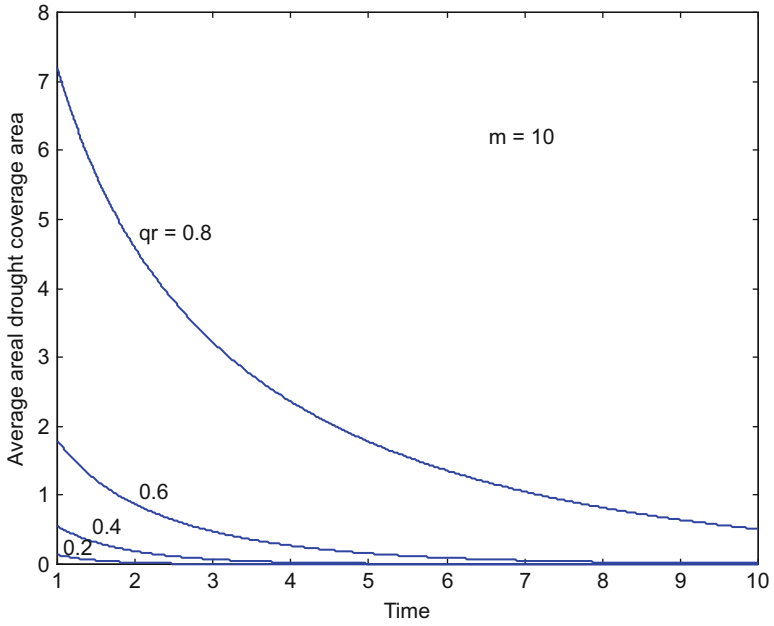


Fig. 2.30 Drought time change by p_r

As the probability of deficit spell in a region increases, the average drought area attains to its maximum value in relatively shorter time as can be written from Eq. 2.90:

$$i = \mathrm{Ln}\left(1 - P_A^i\right)/\mathrm{Ln}(1 - p_r) \tag{2.91}$$

Hence, this expression provides the opportunity to estimate average time period that is required for a certain percentage of the region to be covered by drought. Figure 2.30 indicates the change of i with p_r, which is the surplus probability.

Furthermore, in practical applications the probability of deficit can be approximated empirically as $1/m$ or preferably as $1/(m+1)$. The substitution of these conditions into Eq. 2.91 gives

$$i = \mathrm{Ln}\left(1 - P_A^i\right)/[\mathrm{Ln}(m/(m+1)] \tag{2.92}$$

This expression confirms that regional drought durations are affected mainly by its size rather than its shape as was claimed by Tase and Yevjevich (1978).

The next significant regional drought parameter is the variance which is a measure of drought variability. In general, the smaller the variance, the smaller is the areal drought coverage percentage. The variance of the regional persistence model can be found from Eq. 2.76 and 2.79 after some algebra as

$$V_i(A_d) = m\left(1 - q_r^t\right)q_r^t \tag{2.93}$$

Similarly, it is better in practice to standardize this variance by dividing both sides by the total area, m, which gives percentage variance, $P_V^{i,T}$, in the case of i subarea coverage after T time duration:

$$\mathrm{P}_V^{i,T} = \left(1 - q_r^i\right) q_r^T \tag{2.94}$$

Furthermore, consideration of Eq. 2.90 together with this last expression yields the relationship between percentages of average and variance drought coverage as

$$P_V^{i,T} = P_A^i q_r^T \tag{2.95}$$

Figure 2.31 shows the change of regional drought variance percentage with i deficit affected number of subareas at different times for a given deficit probability, $q_r = 0.7$.

Another application of spatio-temporal agricultural drought occurrence is possible by considering spatial deficit probability, q_r, and temporal surplus probability, p_t. It is rather cumbersome to find a concise expression for the expectation of this case at all times. However, for the first time interval,

$$E_1(A_d) = m(1 - q_r)(1 - p_t) \tag{2.96}$$

which exposes explicitly the contribution of the regional and temporal dry spell effects on the average areal drought. Figure 2.32 shows drought spatial and temporal average variations for given probabilities.

Finally, from the variance of the drought area coverage by simple spatio-temporal model considerations, it is possible to derive for the first time interval that

$$V_1(A_d) = E_1(A_d)(q_r + p_t p_r) \tag{2.97}$$

The numerical solution of this expression is given for various combinations of p_r and p_t in Fig. 2.33.

As a result the PDFs of regional and temporal agricultural droughts are derived for independent dry and wet spell occurrences. Two basically different probabilistic models are proposed for regional drought modeling. Regional drought parameter variations are assessed graphically. The following conclusions are valid for regional and temporal drought occurrences:

1. Drought occurrences are dependent on the regional and temporal dry and wet spell probabilities as well as size of the region considered.
2. Drought area distribution within a region without considering temporal probabilities becomes negatively skewed as its size increases. Initially, it can be approximated by a normal distribution. For multi-seasonal model the same distribution has an approximate normal PDF provided that continuous drought duration is not considered. Otherwise, it is positively skewed.
3. Drought probabilities over a region are more affected by its size.

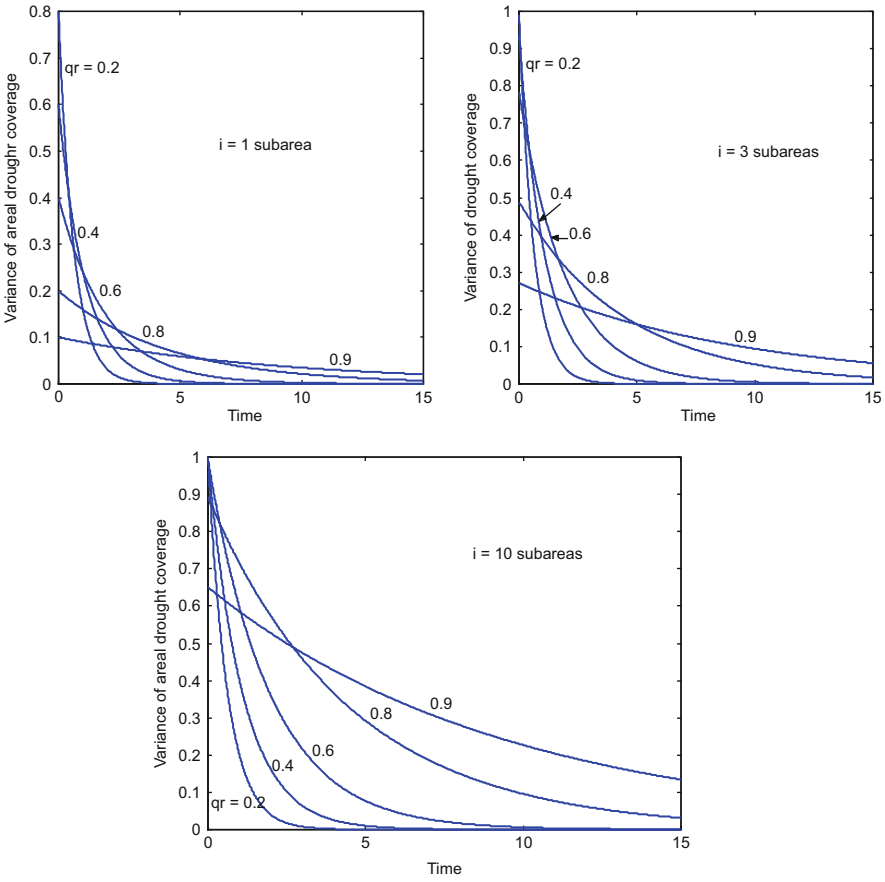


Fig. 2.31 Areal drought variance percentage variation

2.11 Spatio-Temporal Modeling

Conceptual definition of dry and wet spells implies probabilistic chance combinations of recurrent events. For instance, dry (wet) spell can be expressed in terms of deficit (surplus) period of basic time intervals. In probability, statistical, and stochastic approaches, the stationary is taken as a fundamental assumption. Therefore, these formulations cannot be reliable in cases of nonstationary.

In regional behavior of the spells, homogeneity or heterogeneity of the areal extent comes into view. If each subarea of the drought (flood)-stricken region has the same chance of being dry or wet, then such an event has homogeneous features; otherwise, heterogeneity must be considered in the modeling. From a climatic viewpoint, dry and wet spells have frequent quasi-cyclical events, and such periodicities are usually difficult to preserve in models, lack suitable explanations, and

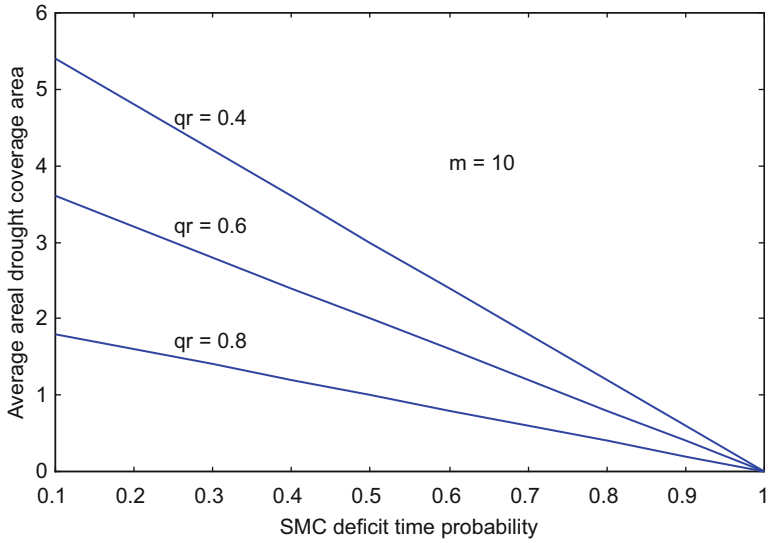


Fig. 2.32 Average areal drought coverage by considering spatio-temporal variations during the first time interval

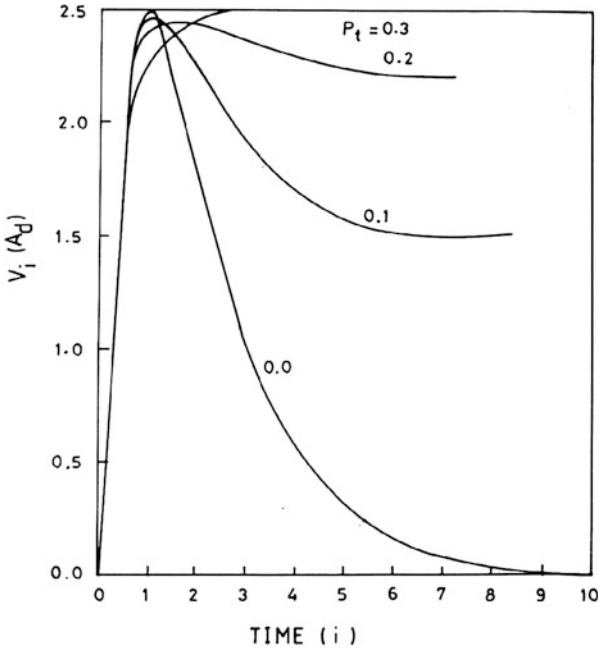


Fig. 2.33 Average variance of drought area for various basic probability values

have rather low predictive utility. For reliable modeling heterogeneity has been considered in this paper.

Floods and especially droughts are extensive regional covers in continental scales. They strike not only one country but several countries or subareas (sites) within the same country in different proportions. In literature, almost all the theoretical studies are confined to temporal assessment with few studies concerning areal coverage.

In order to model the coverage of occurrence patterns, the study area can be divided into characteristically distinctive and mutual exclusive subareas on the basis of prevailing rainfall considerations. It is not necessary that each subarea should be equal in size. It is accepted generally that severe dry and wet spells arise as a result of apparent chance variations of the atmospheric circulation. For instance, dry spells are initiated always by a shortage of precipitation and many of the traditional approaches to drought definition concentration on rainfall analysis only (Smith 2001). Simple assessment of drought severity depends largely on the magnitude and regional extent of precipitation deficiencies from mean climatic conditions. Precipitation is the most easily available data than any other effective variables. For the sake of simplicity, the precipitation will be considered as the main variable, which gives rise to regional and temporal variations of wet and dry spells. Figure 2.34 indicates subareas along the time axis. At initial time instant, there are n_1 dry spells, but after one time step, Δt, there are $n_1^{'}$ dry spells (Şen 2014).

In regional studies, clustering of dry spells in a region is referred to as drought area; otherwise, wet area is valid. In this section, two different regional models are explained. The first model relies on regional dry and wet areal spell probabilities, p_r and q_r, respectively. Mutually exclusiveness implies that $p_r + q_r = 1.0$. This model assumes that once a subarea of usable land is hit by a dry spell, it remains under this state in the subsequent time instances. Hence, as time passes, the number of dry spell hit subareas steadily increases until the whole region is covered by drought. Such a regional model has been referred to as regional persistence model by Şen (1980). The application of this model is convenient for arid and semiarid regions where long drought periods exist.

The second model takes into account the regional as well as temporal occurrence probabilities of wet and dry spells. The probabilities of temporal precipitation deficit (dry) p_t and surplus (wet) q_t occurrences are mutually exclusive, $p_t + q_t = 1.0$. In this model, in an already drought-stricken area, subareas are subject to temporal drought effects (dry or wet spells) in the next time interval. This model is referred to as multi-seasonal model, because it can be applied for a duration which may include several dry and wet periods.

Let the area be divided into m mutually exclusive subareas each with the same spatial (homogeneous) and temporal (stationary) drought chances. The Bernoulli distribution theory yields extend of dry area, A_d, during any time interval, Δt. The probability of n_1 subareas stricken by drought out of m subareas is given as (Feller 1957)

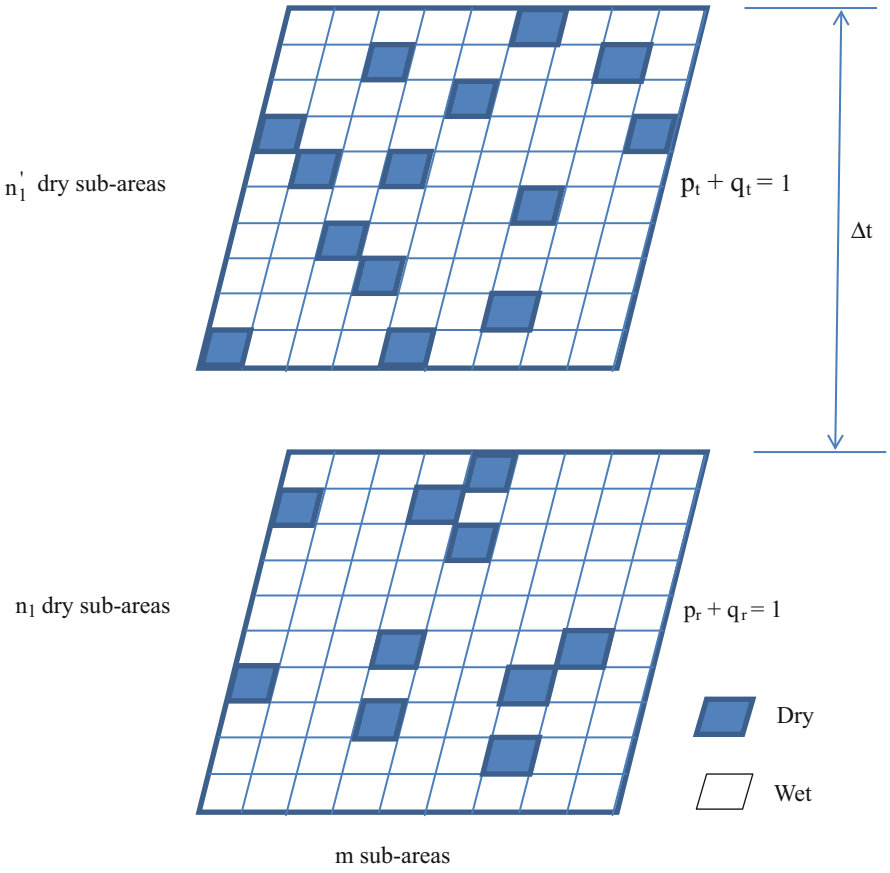


Fig. 2.34 Spatio-temporal wet and dry patter grid

$$P_{\Delta t}(A_d = n_1|m) = \binom{m}{n_1} p^{n_1} q^{m-n_1} \qquad p + q = 1.0 \tag{2.98}$$

By considering explanations given for Fig. 2.5 in case of heterogeneous subareal probabilities, the general form of Eq. 2.98 is derived in this paper as

$$P_{\Delta t}(A_d = n_1|m) = \frac{1}{n_1!} \underbrace{\sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \cdots\cdots \sum_{j_{n1}=1}^{m}}_{n_1-\text{fold}} p_{rj_1} p_{rj_2} \cdots\cdots p_{rj_k} \underbrace{\prod_{j_k=1}^{m}}_{(m-n_1)-\text{fold}} q_{rj_k}$$
$$j_2 \neq j_1 \quad j_{n_1} \neq j_k \qquad j_k \neq j_l \tag{2.99}$$
$$k = 1, 2, \ldots\ldots n_1 - 1 \qquad l = 1, 2, \ldots\ldots n_1$$

where p_{rji} is the regional probability of dry spell hit of subarea, j_i. Herein, indices i take its value according to arrangements beneath the summation and multiplication signs.

The n_1-fold summation terms imply the exhaustive alternatives of dry periods and (m-n)-fold multiplication term is for collective wet spells at the remaining subareas. In case of homogeneous areal probability at each site, Eq. 2.99 reduces to Eq. 2.98. In order to verify the validity of Eq. 2.99, simple calculation procedure is shown in the following two examples.

Example 2.11 The composition of Eq. 2.99 for $m=4$ and $n_1=2$, for which Eq. 2.99 reduces to

$$P_{\Delta t}(A_d=2|4)=\frac{1}{2!}\sum_{\substack{j_1=1\\ j_2\neq j_1}}^{4}\sum_{\substack{j_2=1\\ j_4\neq j_l,j_2}}^{4}p_{rj_1}p_{rj_2}\prod_{j_3=1}^{4}q_{rj_3}$$

Table 2.12 presents in detail all the necessary steps for the calculation of the final probability from this last expression with similar calculations from Eq. 2.98 in a comparative manner.

The summation of the terms in column V gives the right-hand side of the previous expression except 1/2! factor. Hence, the formulation can be written explicitly as

$$\begin{aligned}&P_{\Delta t}(A_d=2|4)\\&=\frac{1}{2!}\begin{pmatrix}P_{\mathrm{r}1}P_{\mathrm{r}2}q_{\mathrm{r}3}q_{\mathrm{r}4}+p_{\mathrm{r}1}p_{\mathrm{r}3}q_{\mathrm{r}2}q_{\mathrm{r}4}+p_{\mathrm{r}1}p_{\mathrm{r}4}q_{\mathrm{r}2}q_{\mathrm{r}3}+p_{\mathrm{r}2}p_{\mathrm{r}1}q_{\mathrm{r}3}q_{\mathrm{r}4}+p_{\mathrm{r}2}p_{\mathrm{r}3}q_{\mathrm{r}1}q_{\mathrm{r}4}\\+p_{\mathrm{r}2}p_{\mathrm{r}4}q_{\mathrm{r}1}q_{\mathrm{r}3}+p_{\mathrm{r}3}p_{\mathrm{r}1}q_{\mathrm{r}2}q_{\mathrm{r}4}+p_{\mathrm{r}3}p_{\mathrm{r}2}q_{\mathrm{r}1}q_{\mathrm{r}4}+p_{\mathrm{r}3}p_{\mathrm{r}4}q_{\mathrm{r}1}q_{3}+p_{\mathrm{r}4}p_{\mathrm{r}1}q_{\mathrm{r}2}q_{\mathrm{r}3}\\+p_{\mathrm{r}4}p_{\mathrm{r}2}q_{\mathrm{r}1}q_{\mathrm{r}3}+p_{\mathrm{r}4}p_{\mathrm{r}3}q_{\mathrm{r}1}q_{\mathrm{r}2}\end{pmatrix}\end{aligned}$$

In case of homogeneous point probabilities, this detailed expression reduces to the following simple form:

$$P_{\Delta t}(A_d=2|4)=\frac{1}{1.2}\left(4\times3\times p^2q^2\right)=\frac{12}{2}p^2q^2$$

On the other hand, Eq. 2.98 yields

$$P_{\Delta t}(A_d=2|4)=\binom{4}{2}p^2q^{4-2}=6p^2q^2$$

Hence, Eqs. 2.98 and 2.99 yield the same result in both cases, homogeneous and heterogeneous subareal probabilities.

Example 2.12 Now let us consider the case of $m=6$ and $n_1=4$. According to Eq. 2.99, the corresponding formulation becomes

Table 2.12 Detailed probability calculations

	Heterogeneous probabilities ($p_{r1} \neq p_{r2} \neq p_{r3} \neq p_{r4}$)				Homogeneous ($p = p_{r1} = p_{r2} = p_{r3} = p_{r4}$)
j_1	j_2	$\sum_{j_1=1}^{4} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{4} P_{rj_1} p_{rj_2}$	$\prod_{\substack{j_3=1 \\ j_3 \neq j_2}}^{4} q_{rj_3}$	$P_{\Delta t}(A_d = 2\|4)$	$P_{\Delta t}(A_d = 2\|4)$
I	**II**	**III**	**IV**	**V**	**VI**
1	2	$p_{r1}p_{r2}$	$q_{r3}q_{r4}$	$p_{r1}p_{r2}q_{r3}q_{r4}$	p^2q^2
	3	$p_{r1}p_{r3}$	$q_{r2}q_{r4}$	$p_{r1}p_{r3}q_{r2}q_{r4}$	p^2q^2
	4	$p_{r1}p_{r4}$	$q_{r2}q_{r3}$	$p_{r1}p_{r4}q_{r2}q_{r3}$	p^2q^2
2	1	$p_{r2}p_{r1}$	$q_{r3}q_{r4}$	$p_{r2}p_{r1}q_{r3}q_{r4}$	p^2q^2
	3	$p_{r2}p_{r3}$	$q_{r1}q_{r4}$	$p_{r2}p_{r3}q_{r1}q_{r4}$	p^2q^2
	4	$p_{r2}p_{r4}$	$q_{r1}q_{r3}$	$p_{r2}p_{r4}q_{r1}q_{r3}$	p^2q^2
3	1	$p_{r3}p_{r1}$	$q_{r2}q_{r4}$	$p_{r3}p_{r1}q_{r2}q_{r4}$	p^2q^2
	2	$p_{r3}p_{r2}$	$q_{r1}q_{r4}$	$p_{r3}p_{r2}q_{r1}q_{r4}$	p^2q^2
	4	$p_{r3}p_{r4}$	$q_{r1}q_{r3}$	$p_{r3}p_{r4}q_{r1}q_{r3}$	p^2q^2
4	1	$p_{r4}p_{r1}$	$q_{r2}q_{r3}$	$p_{r4}p_{r1}q_{r2}q_{r3}$	p^2q^2
	2	$p_{r4}p_{r1}$	$q_{r1}q_{r3}$	$p_{r4}p_{r2}q_{r1}q_{r3}$	p^2q^2
	3	$p_{r4}p_{r3}$	$q_{r1}q_{r2}$	$p_{r4}p_{r3}q_{r1}q_{r2}$	p^2q^2

$$P_{\Delta t}(A_d = 4|6) = \frac{1}{4!}\sum_{j_1=1}^{6}\sum_{j_2=1}^{6}\sum_{j_3=1}^{6}\sum_{j4=1}^{6} p_{rj_1}p_{rj_2}p_{rj_3}p_{rj_4}\prod_{j5=1}^{6} q_{rj_5}$$
$$j_2 \neq j_1\, j_3 \neq j_1, j_2\, j_4 \neq j_1, j_2, j_3 \quad j_5 \neq j_l, j_2, j_3, j_4$$

which has all the terms explicitly as

$$P_{\Delta t}(A_d = 4|6) = \frac{1}{1.2.3.4}\left(6 \times 5 \times 4 \times 3 \times p^4 q^2\right) = 15p^4q^2$$

On the other hand, Eq. 2.98 yields

$$P_{\Delta t}(A_d = 4|6) = \binom{6}{4} p^4 q^2 = 15p^4q^2$$

It is possible to try similar calculations with different sets of p, m, and n and the two formulations yield exactly the same results. This way is referred to induction methodology in the mathematical literature and hence the validity of Eq. 2.98 is confirmed.

The application of the presented methodology is presented for six precipitation stations on the northwestern part of Turkey as shown in Fig. 2.35 within the metropolitan boundaries of Istanbul City. These stations are distributed three and three on the European and Asian continental parts of the city. The records at each one of these stations are available from 1930 to 2010. In the area there are two different types of climate patterns. In the northern parts near the Black Sea where the mountainous areas extend, direct effect of the North Atlantic polar and maritime air movements are effective. In the southern art along the Marmara Sea, a rather temperate climate prevails with intrusions from the Mediterranean type of modified climate.

Table 2.13 indicates the location, averages, standard deviations, and regional as well as temporal dry and wet spell probabilities. Next to the international station numbers in the first column, the local station numbers are given in the parenthesis. In the dry and wet spell probability calculations, the annual averages are taken as threshold level.

Figure 2.36 is the template similar to the one in Fig. 2.33, where each subarea (site) is shown representatively by a square.

The regional and temporal dry spell probabilities are given in Fig. 2.37 for Δt time period and the same repetition will continue for many time periods in the future.

In order to carry out the numerical calculations, the author wrote a simple computer program. Input values are p_r and p_t and number of subareas, which is $n = 6$ in this case. Table 2.2 provides p_r and p_t values for these six stations in Istanbul. Figure 2.38 is the result of the program, which shows the change of areal drought coverage versus the cumulative probability of coverage area. It is observed

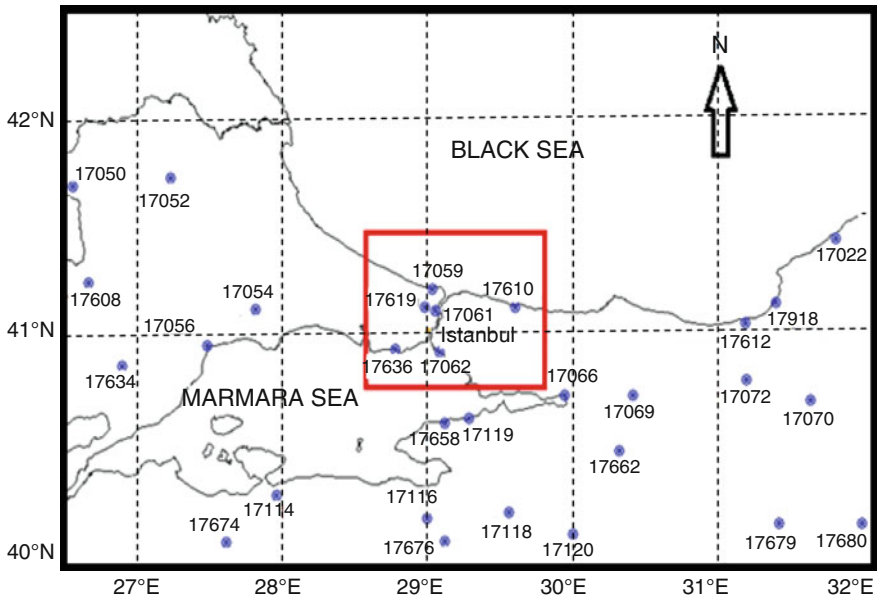


Fig. 2.35 Istanbul area meteorology station

that after the third year, the cumulative probability curve becomes a steady-state indication that in Istanbul Region, the average drought duration is about 3 years. This is in concurrence with the historical drought duration average which is referred to as 2.5–3 years (Şen 2010).

Detailed calculations are shown in Table 2.14 for the cases between $1\Delta t$ and $2\Delta t$ instances with marginal probabilities on the last column and row. Herein, subarea numbers 0 and 6 represent none of the subareas is under dry conditions, i.e., the whole region is covered by wet spell, whereas 6 indicates that the whole region is under dry conditions. The values on the last row indicate the drought-hit area probability after one time interval, which is 1 year in this work. One can calculate the cumulative probability for any given condition. Each cell value in this table shows conditional probability that if any subarea at the first instance is dry, then it is also dry in the next time instant. For instance, the probability that subarea 5 is dry given that the subarea 4 was dry initially is equal to 0.0526. If one is interested in the probability that subarea 1 is dry irrespective of initial subareas, he/she will end up with 0.0065.

Table 2.14 can also provide information to questions such as "if a given number of subareas are under dry conditions initially, what is the probability that more than this number will be under dry conditions?" Say, if three subareas are dry initially, what is the probability that five subareas will be under dry condition after one time interval? The answer can be found after fixing subarea 2 column in this table and by

Table 2.13 Station characteristics

Station no.	Longitude (E)	Latitude (N)	Precipitation (mm)		Regional probability		Temporal probability	
			Average	Std. dev.	Wet, p_r	Dry, q_r	Wet, p_t	Dry, q_t
17062 (18)	40.97	29.08	680.76	118.00	0.5584	0.4416	0.1711	0.8289
17636 (128)	40.98	28.78	596.06	199.85	0.3766	0.6234	0.4079	0.5921
17061 (17)	41.15	29.05	616.68	363.45	0.2987	0.7013	0.6318	0.3682
17619 (119)	41.17	28.98	854.84	490.94	0.3117	0.6883	0.6316	0.3684
17610 (116)	41.17	28.60	719.81	323.35	0.4545	0.5455	0.3816	0.6184
17059 (16)	41.25	29.03	572.62	323.35	0.3247	0.6753	0.6184	0.3816
Average					0.3874			

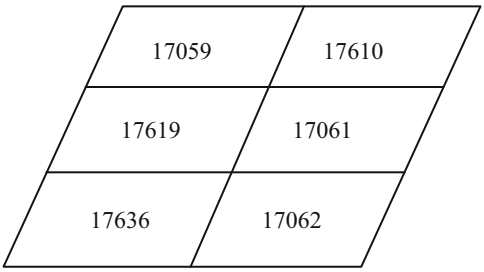


Fig. 2.36 Istanbul six station representative subareas

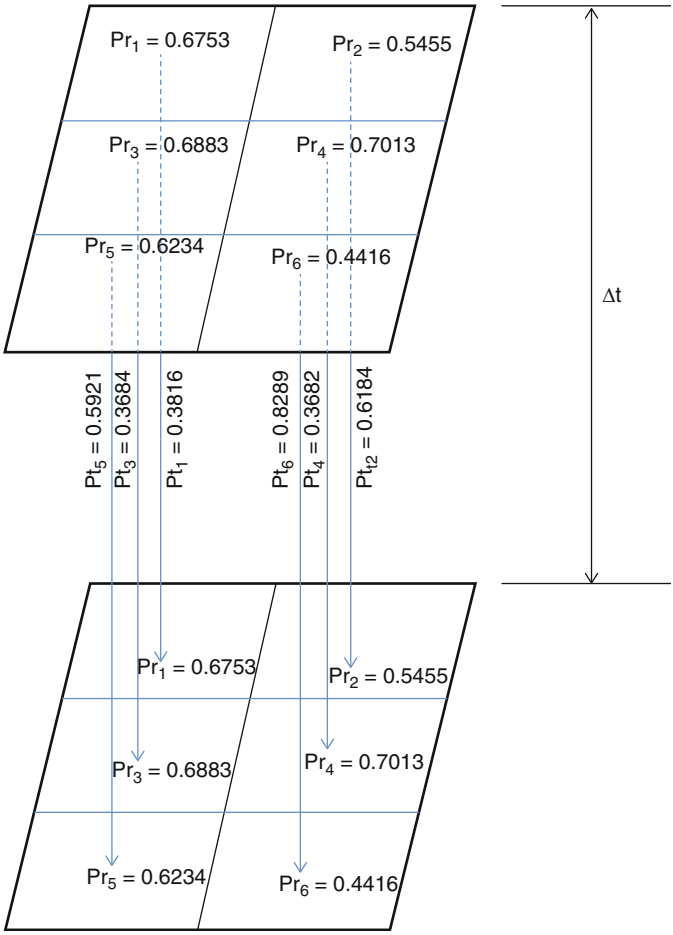


Fig. 2.37 Heterogeneous regional and temporal dry spell probabilities

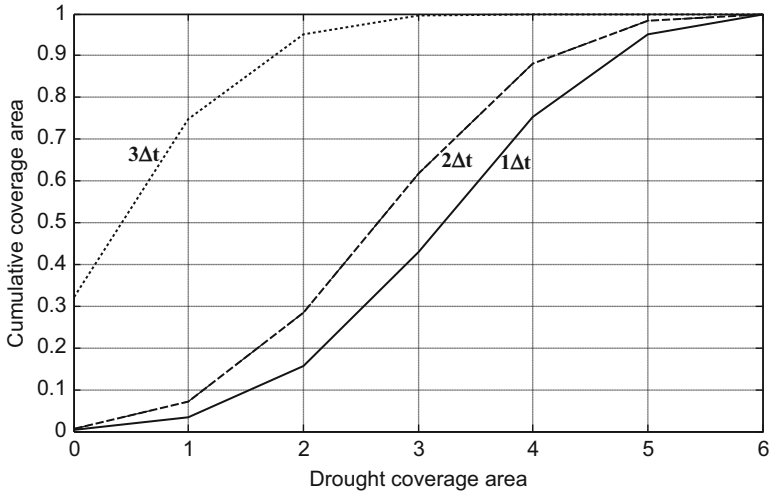


Fig. 2.38 Areal coverage versus cumulative probability of drought relationships

Table 2.14 Subareal probabilities between two successive time intervals

Subareas at $2\Delta t$ time instant	Subareas at $1\Delta t$ time instant							Marginal probability
	0	1	2	3	4	5	6	
0	0	0.0002	0.0006	0.001	0.0008	0.0003	0	0.0029
1	0.0002	0.0019	0.0063	0.0099	0.0079	0.0031	0.0005	0.0298
2	0.0008	0.0079	0.0265	0.0415	0.0329	0.0128	0.002	0.1244
3	0.0018	0.0173	0.0578	0.0905	0.0718	0.028	0.0043	0.2715
4	0.0021	0.0206	0.0689	0.1079	0.0856	0.0334	0.0051	0.3236
5	0.0013	0.0127	0.0423	0.0663	0.0526	0.0205	0.0031	0.1988
6	0.0003	0.0031	0.0104	0.0163	0.0129	0.005	0.0008	0.0488
Marginal probability	0.0065	0.0637	0.2128	0.3334	0.2645	0.1031	0.0158	1.0000

adding all the conditional probabilities in this column up to 5, inclusive. Hence, the answer is 0.2024. The reader can enumerate similar questions and find the corresponding answers from the same table.

The entrance of the last row probabilities into the same computer program provides the next successive probability expositions as in Table 2.15. The last row in this table yields the curve in Fig. 2.38 that corresponds to $3\Delta t$.

Table 2.15 Subareal probabilities between two successive time intervals

Subareas at $2\Delta t$ time instant	Subareas at $2\Delta t$ time instant							Marginal probability
	0	1	2	3	4	5	6	
0	0.0021	0.0205	0.0686	0.1073	0.0852	0.0332	0.0051	0.322
1	0.0028	0.0271	0.0905	0.1417	0.1124	0.0438	0.0067	0.425
2	0.0013	0.013	0.0436	0.0682	0.0541	0.0211	0.0032	0.2045
3	0.0003	0.0028	0.0094	0.0147	0.0117	0.0045	0.0007	0.0441
4	0	0.0003	0.0009	0.0014	0.0011	0.0004	0.0001	0.0042
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
Marginal probability	0.0065	0.0637	0.213	0.3333	0.2645	0.103	0.0158	1.0000

References

Akin JE (1971) Calculation of areal depth of precipitation. J Hydrol 12:363–376

Bayraktar H, Turalioglu FS, Şen Z (2005) The estimation of average areal precipitation by percentage weighting polygon method in Southeastern Anatolia Region, Turkey. Atmos Res 73:149–160

Box GFD, Jenkins GM (1976) Time series analysis, control and forecasting, Holden Day

Bras RL, Rodriguez-Iturbe I (1985) Random functions and hydrology. Addison- Wesley Publishing Co., Reading, 599p

Bruce JP, Clark RH (1966) Hydrometeorology. Pergamon Press, New York, 319 pp

Chow VT (1964) Handbook of applied hydrology. McGraw-Hill, New York

Chu PS, Karz RW (1985) A time-domain approach. MonWeather Rev 113:1876–1888

Cramer H (1946) Mathematical methods of statistics. Princeton University Press, Princeton, p 213

Davis A (2002) Statistics data analysis geology. Wiley, New York, 638 pp

Dirichlet GL (1850) Über die Reduktion der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen. J Reine Angew Math 40:209–227

Feller W (1957) An introduction to probability theory and its applications. Wiley, New York

Fisher RA (1912) On an absolute criterion for filtering frequency curves. Messenger Math 41:155

Gauss KF (1963) Theory of the motion of the heavenly bodies about the sun in conic section. Dover, New York

Hevesi JA, Istok JD, Flint AI (1992) Precipitation estimation in mountainous terrain using multivariate geostatistics, part I: structural analysis. J Appl Meteorol 31:661–676

Huff FA, Neill JC (1957) Areal representativeness of point rainfall. Trans Am Geophys Union 38 (3):341–351

Jackson IJ (1972) Mean daily rainfall intensity and number of rainy days over Tanzania. Geogr Ann A 54:369–375

Kedem B, Chiu LS, Karni Z (1990) An analysis of the threshold method for measuring area-average rainfall. J Appl Meteorol 29:3–20

Kendall MG (1954) Note on the bias in the estimation of autocorrelation. Biometrika 42:403–404

Koch S, Link RF (1971) Statistical analysis of geological data. Dover Publications, New York, 375 p

Kolmogorov AN (1941) Interpolation and extrapolation von stationaren zufalligen folgen. Bull Acad Sci USSR Ser Math 5:3–14

Mercalli G (1912) Sulle modifieazioni proposte alia scale sismica de Rossi-Forel. Boll Soc Sism ltal 8:184–191
Quenouille MH (1956) Notes on bias in estimation. Biometrika 43:353–360
Schubert SD, Saurez MJ, Schemm JK (1992) Persistence and predictability in a perfect model. J Atmos Sci 49(5):256–269
Şen Z (1974) Small sample properties of stationary stochastic processes and Hurst phenomenon in hydrology. Unpublished Ph. D. Thesis, Imperial College of Science and Technology, University of London, 256 pp
Şen Z (1980) Regional drought and flood frequency analysis: theoretical considerations. J Hydrol 46:265–279
Şen Z (1998a) Average areal precipitation by percentage weighted polygon method. ASCE J Hydrol Eng 3(1):69–72
Şen Z (1998b) Small sample estimation of the variance of time-averages in climatic time series. Int J Climatol 18:1725–1732
Şen Z (2002) İstatistik Veri İşleme Yöntemleri (Hidroloji ve Meteoroloji) (Statistical Data Treatment Methods – Hydrology and Meteorology). Su Vakfı yayınları, 243 sayfa (in Turkish)
Şen Z (2008) Kuraklik Afet ve Modern Modelleme Yontemleri (Drought disaster and modern modeling methods) Turkish Water Foundation (in Turkish)
Şen Z (2010) Fuzzy logic and hydrological modeling. Taylor and Francis Group, CRC Press, New York, 340 p
Şen Z (2014) Regional wet and dry spell analysis with heterogeneous probability occurrences. J Hydrol Eng. doi:10.1061/(ASCE)HE.1943-5584.0001144, 04014094
Smith K (2001) Environmental hazards. Assessing risk and reducing disaster. Routledge, New York, p 392
Stout GE (1960) Studies of severe rainstorms in Illinois. J Hydraul Div Proc ASCE, HY4, 129–146
Summer G (1988) Precipitation process and analysis. Wiley, New York, 455 pp
Tabios GO III, Salas JD (1985) A comparative analysis of techniques for spatial interpolation of precipitation. Water Resour Bull 21:365–380
Tase N (1976) Area-deficit-intensity characteristics of droughts. Colorado State University, Fort Collins, Colorado., Hydrology Paper 87
Tase N, Yevjevich Y (1978) Effects of size and shape of a region on drought coverage. Bull Hydrol Sci 23(2):203–213
Thiessen AH (1912) Precipitation averages for large areas. Mon Weather Rev 39:1082–1084
Voroni G (1907) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. J Reine Angew Math 133:97–178
Wiener N (1942) The extrapolation, interpolation and smoothing of stationary time series. OSRD 370, Report to the Services 19**,** Research Project DIC-6037, MIT
Wiener N (1949) The extrapolation, interpolation and smoothing of stationary time series. Wiley, New York
Wiesner CJ (1970) Hydrometeorology. Chapman and Hall Ltd., London, 232 pp
Wilson JW, Atwater MA (1972) Storm rainfall variability over Connecticut. J Geophys Res 77 (21):3950–3956
Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybern 3:28–44

Chapter 3
Point and Temporal Uncertainty Modeling

Abstract In earth sciences, there are regular data sets as time series and irregular data sets in the forms of uncontrollable variable measurements such as floods, droughts, earthquake occurrences, volcanic eruptions, stratigraphic thicknesses, fractures, fissures, or ore content. The uncertainty features in both data sets require probabilistic and statistical data processing techniques, which are the main topics of this chapter. For this purpose, prior to any methodology, the empirical frequency definition of various types is given, and their connections with the probability theory are explained with examples. Subsequently, probabilistic and statistical prediction modeling tools are presented for independent and dependent processes.

Keywords Regular and irregular data • Point modeling • Frequency • Histogram • Temporal modeling • Dependent and independent processes

3.1 General

Our minds are preconditioned on the Euclidian geometry, and consequently, ordinary human beings are bound to think in 3D spaces as length, width, and depth, but additionally the fourth dimension is time. Hence, the scientific formulations, differential equations, and others include space and time variability. All the earth, environmental, and atmospheric events vary along the four dimensions. If their changes along the time are not considered, then it is said to be frozen in time, and therefore, an instantaneous situation remains along the time axis but the variable has special variation in space. A good example for such a change can be considered as geological events, which do not change with human life time span. Iron, gold, magnesium, petroleum, and water contents of a rock mass vary rather randomly from point to another within the rock. Another example is the measurement of rainfall amounts at many irregularly located meteorology stations spread over an area, i.e., simultaneous measurement of rainfall amounts; again the time is frozen and the spatial variations are in view. Similar situation is valid for measurements taken from a set of oil or water wells.

Any natural event evolves in the 4D space-time media within the human visualization domain. Any record that has this property is referred to as the spatio-temporal variation.

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_3

On the other hand, there are timewise variations, which are referred to as the temporal variations in the natural phenomenon, and their records appear as a time series. For such a variation, it suffices to measure the event at a given location, which is the case at oil and water well or meteorology station cases. A time series is the systematic measurement of any natural event along the time axis at regular time intervals. Depending on this time interval, time series is called as hourly, daily, weekly, monthly, or annual time series.

Contrary to time variations, it is not possible to consider space series, because the records are not kept at regular distances except in very specific cases. For example, if water samples along a river are taken at every 1 km, then the measurements provide an equal-distance series in the regularity sense. In fact, distance series are very limited as if there are no such data sequences in practice. On the other hand, depending on the interest of event, there are series as time records but they are not time series due to irregularity or randomness in the time intervals between successive occurrences. Flood and drought occurrences in hydrology, earthquake, landslide, tsunami, and volcanic eruptions correspond to such cases. One cannot know the duration between any two successive occurrences of these events. Likewise, in meteorology the occurrence of precipitation or any dangerous levels of concentrations of air pollutants do not have time series characteristics.

Since the earth sciences data have uncertainties, they appear according to probabilistic or chaotic principles. In practical terms, probability is a percent value that is commonly used among people almost every day. Among the probabilistic questions is the oil hit possibility by a drill (wet or dry), earthquake occurrence rate (frequency) within time duration, fracture percentage along a scanline on an outcrop, relative frequency of rainfall occurrence, etc. None of these questions can be answered with certainty, but through the probabilistic, statistical, or stochastic principles. It is, therefore, very significant that earth scientist data treatment should be based on these principle concepts.

One of the most frequently used concepts in the probability calculations is the occurrence of an event. It is defined as the collection of uncertain outcomes in the forms of sets, classes, categories, clusters, or groups. Events might be elementary or compound depending on the decomposition. A compound event can be decomposed into at least two events, whereas an elementary event cannot be decomposed. For instance, if one considers a sequence of rainy and non-rainy days, non-rainy yesterday corresponds to a single elementary event, but non-rainy days, say, in the last 10 days, represent a compound event, because there may be 1, 2, 3, . . ., or 10 non-rainy days. In general, what is defined as elementary or compound event depends on the problem under consideration. Occurrence of precipitation in any day may include different events such as rainfall or snow or hail, and therefore, it is a compound event composed of these elementary events. If one says that the rainfall occurs or does not occur, these are mutually exclusive complementary elementary events. It is also possible to be interested in the amount of rainfall in addition to its occurrence; this is then a compound event with two parts, namely, the occurrence and the amount. For example, the flood disaster is a compound event since it must occur

with some damages on human life and/or property. In the probability calculations, it is preferred to decompose a compound event to underlying elementary events.

3.2 Regular Data Set

In practical studies, regular data are measurements at regular time, or space (distance, area, or volume) intervals, and most often time series are available almost in any discipline and abundantly in the earth sciences. In general, these data types are measured under controllable conditions, and the person in charge can decide about time or space intervals. In earth sciences, the following are among the regular data sets:

1. Hydrometeorology: Daily, monthly, or annual records of meteorological variables, such as rainfall, evaporation, and relative humidity.
2. Engineering geology: Scanline measurements and fracture spacing specification for rock quality designation calculations.
3. Geophysics: Electrical resistivity or seismic reflection measurements for subsurface geological prospecting.
4. Remote sensing: Pixel sizes are all in regular form so as to make the necessary interpretations and modeling studies.
5. Earth sciences modeling: Numerical solution of relevant differential equations by finite element approach, where the data are considered at the nodes of a mesh.
6. Earthquake: Seismogram measurements in the form of time series at a set of regular time intervals.
7. Hydrogeology: Groundwater quality monitoring at regular time intervals.

3.3 Irregular Data Set

Generally, these are the measurements of uncontrollable variables such as floods, droughts, earthquake occurrences, volcanic eruptions, stratigraphic thicknesses, fractures, fissures, ore content, etc. Among the irregular data sets, one can cite the following points in the earth sciences domain:

1. Flood occurrence times and also their magnitudes provide uncontrollable and irregular data.
2. Droughts are creeping phenomenon and one cannot know its duration and areal extends.
3. Earthquake epicenter locations and their depths vary from one to the next earthquake and, hence, appear in the form of irregular data.
4. Water or oil well locations are irregularly distributed in a potential area, and therefore, any measurement taken from each well appears in the form of spatially irregular data.

5. Pumping (pressure) test in water (oil) wells provides irregular data at a set of time instances, which are rather frequent at early times but with time passage they become rarer.
6. Fracture spacing measurements along a scanline for rock quality designation (RQD) assessments.
7. Temporal and spatial number of particulate matter or dust in a given volume over a certain study area.

3.4 Point Data Set Modeling

These are the set of measurements at a fixed location, and therefore, they reflect the temporal behavior of the concerned event at the point. The well logs are also reflections of stratigraphy at a given point. Since earth sciences events include uncertainties, they are in the form of most often regular and rarely irregular data types. Either type of data can be treated according to well-known frequency diagrams or their standard form of histograms leading to theoretical probability distribution functions.

3.4.1 Empirical Frequency Distribution Function

The question of any set of data without any preference attachment to any one of them can be searched whether it originates from a single phenomenon or more. This is referred to as the population belongingness in the statistical terms. In case of a single population origin, the shape of the distribution should have one of the standard forms known in the statistical literature (Davis 1986).

3.4.2 Relative Frequency Definition

For categorization as mentioned in the previous subsection, certain data values or design quantities must be considered. For instance, if a sequence of data X_i ($i = 1, 2, \ldots, n$) and a design level X_0 are given, then the classical frequency calculations can be done easily. The significance of the design value is for categorization of the given data into two parts, namely, those data values greater than the design value, ($X_i > X_0$) and others that are smaller. If the number of greater data values is n_G, then relative frequency, f_G, can be calculated as

$$f_G = \frac{n_G}{n} \tag{3.1}$$

Likewise, the relative frequency of the other data category being smaller than the design value can be written as

$$f_S = \frac{n_S}{n} \tag{3.2}$$

This definition is equivalent to the empirical probability of the categories. It is obvious that the summation of these two relative frequencies is equal to 1, because by definition, $f_G + f_S = 1$.

3.4.3 Classical Definition

This definition is based on the conceptual evolution of the phenomenon, and the percentage or relative frequency or probability can be defined even without data availability. Herein, again different categories of phenomenon outcomes are taken into consideration. For instance, the question of what is the relative frequency (probability) of major rock types (igneous, sedimentary, and weathered) in a given study area can be answered by considering the categories on a geology map, and the relative frequencies of igneous, f_I, sedimentary, f_S, and weathered, f_w, rocks are calculated. Again the summation of these three relative frequencies add to 1 as

$$f_I + f_S + f_W = 1.0 \tag{3.3}$$

In general, the frequency definition of probability approaches the classical definition for very large data numbers.

3.4.4 Subjective Definition

The relative frequency (probability) is regarded as the degree of belief or quantified judgment of an individual expert about the occurrence of different phenomologic parts. It is possible for an expert weather forecaster, who has been working on the weather prediction in an area for many years, that she/he can attach meaningful percentages to the rainfall occurrences for the next few days. Hence, two individuals can have different subjective probabilities (relative frequencies) for an event without either being necessarily wrong. However, the probability suggestions of these two individuals are close to each other. However, their suggestions depend on their long experiences about the phenomenon concerned. Subjective relative frequency (probability) does not mean that an individual is free to choose any number and call it probability. The subjective values must be consistent and confined within 0–1 range, inclusive.

In any data processing, the number of data that falls within a subinterval in the data variation range is very important, since it leads to the percentage, probability, or relative frequency of the data. In practice, the data range is divided into a number of adjacent non-overlapping equal class lengths, and the number of data in each class is evaluated from the given data. The following steps are necessary for such evaluations:

1. The total number of data, n, in a given time series or data series helps to decide about the number of classes. The larger the data, the more is the class numbers.
2. If the range is divided into two equal classes, the numbers of data that fall into these two categories, which are referred to herein as the frequency, and, in general, are not equal. In skewed frequency distribution functions, these two numbers are different from each other. If the frequencies in each class are n_1 and n_2 ($n_1 \neq n_2$), then it is necessary that $n_1 + n_2 = n$.
3. If the data range, R, which is the difference between the maximum, $X_{\max}$, and minimum, $X_{\min}$, data values, is divided into m classes, then the length, L, of each class can be calculated as

$$L = \frac{X_{\max} - X_{\min}}{m} = \frac{R}{m} \tag{3.4}$$

The sum of frequencies in each one of these classes is equal to the total number of data:

$$n_1 + n_2 + \ldots\ldots + n_m = n \tag{3.5}$$

If each class is represented by its mid-class value, then there will be m mid-class values as X_j ($j = 1, 2, \ldots, m$), with corresponding frequency values, n_j ($j = 1, 2, \ldots, m$). If these are plotted versus to each other on a graph as in Fig. 3.1, it is named as the frequency distribution diagram in data-processing affairs. In this evaluation, the key question is how to choose the number of classes, m. In practice, it is necessary that $5 < m < 15$. If the data number is between 10 and 20, the class number is adopted as 5, and for each 10 data number increase, m is increased by 1. For instance, if $n = 55$, then according to this rule, $m = 9$. Frequency distribution diagram is defined as the change of frequencies within the subclass midranges. It provides visually the change of frequency within the data values. The midpoints of each class are shown in Fig. 3.1.

Depending on the data behavior, there are different frequency diagrams in practical applications. In Fig. 3.2, nine of them are shown.

Logically, any data set can be thought initially and simply as comprising three different subsets, namely, "low," "medium," and "high." The frequency is the number of data in each subset. The subset can be a group, class, category, or cluster of similar property data. This means that there is a relationship between the categories and the number of data that falls within them. In general, as shown in Fig. 3.2, there are nine alternatives on the basis of three categories.

The frequency distribution functions in Fig. 3.2 can be generalized by increasing the number of categories. Based on the understanding on three subsets, the reader can then generalize the frequency functions according to his/her problem

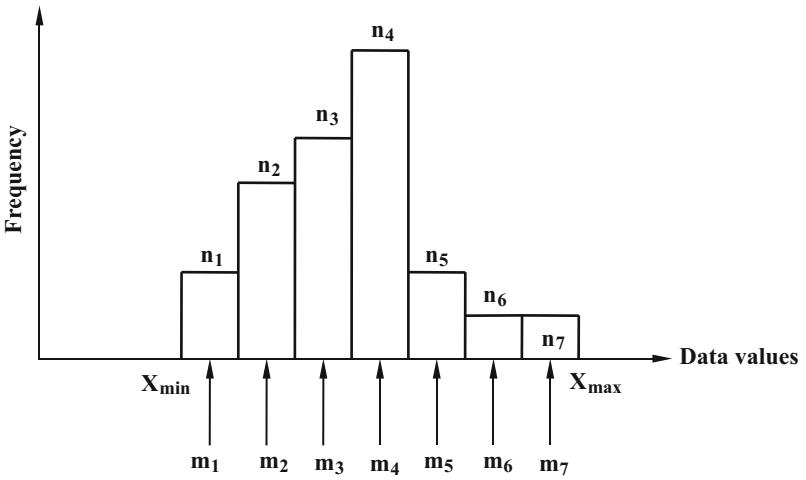


Fig. 3.1 Frequency diagram

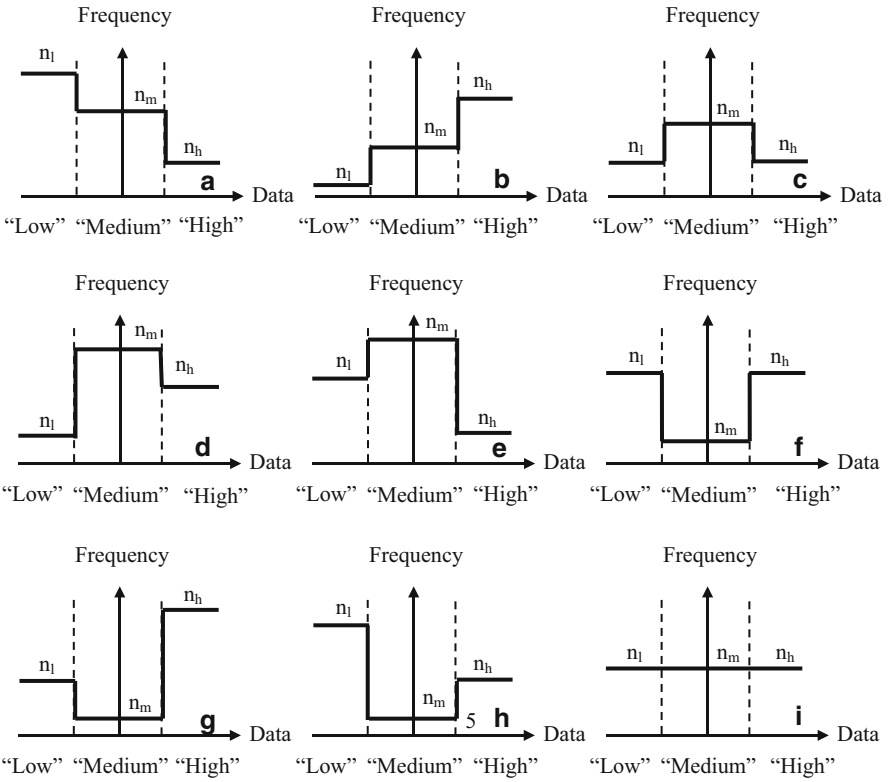


Fig. 3.2 Possible frequency distribution functions

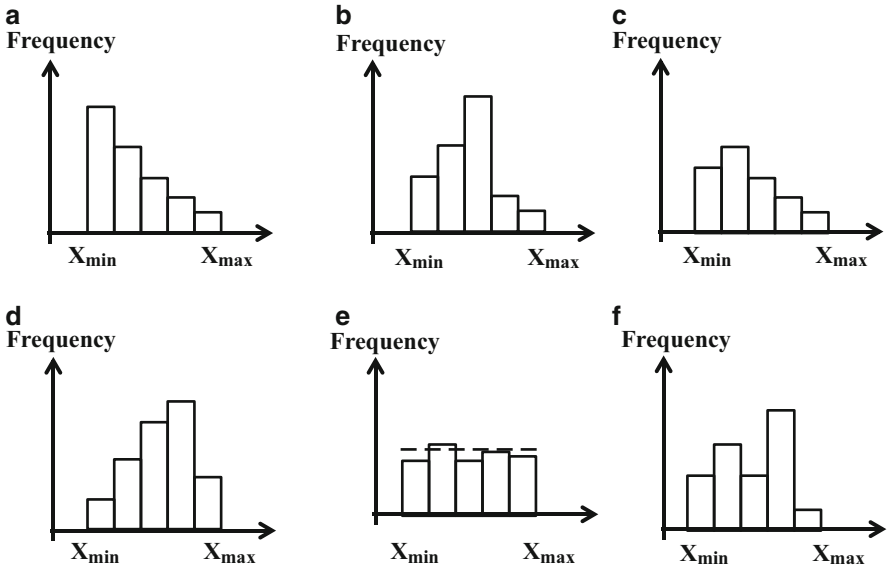


Fig. 3.3 Practical frequency distribution functions

and data at hand. Inspection of the frequency diagrams in Fig. 3.2 leads to the following main points:

1. Some of the frequency distribution functions are symmetrical, and they imply that the extreme "low" and "high" value occurrences are the same. Additionally, statistical parameters, arithmetic average, mode, and median are equal to each other in addition to skewness coefficient which is equal to zero or they may be practically different from each other at ± 5 % error limits.
2. Another set of frequency distribution functions are nonsymmetric, which means that they are skewed with a positive (right skewed) or negative (left skewed) with a skewness coefficient significantly different than zero.
3. Some of the frequency distribution functions have continuously increasing or decreasing forms, which imply that either the frequency of categories is "low" < "medium" < "high" or "high" > "medium" > "low." In statistical context, either mode < median < mean or mode > median > mode depending on the right or the left skewness values, respectively.

Figure 3.3 includes different frequency distribution functions that are frequently used in the practical applications.

Frequency functions show the change of frequencies within the subclass midpoints. It provides visually the change of frequency within the data structure. In Fig. 3.3e almost all class frequencies are equal and it is referred to as the uniform frequency function. It implies completely random variation of the phenomenon concerned. On the other hand, in Fig. 3.3f, there are two maximum frequencies, which imply that the data originate from two distinctive phenomena. In Fig. 3.3b, the frequency function is almost symmetrical, and this shows that the three

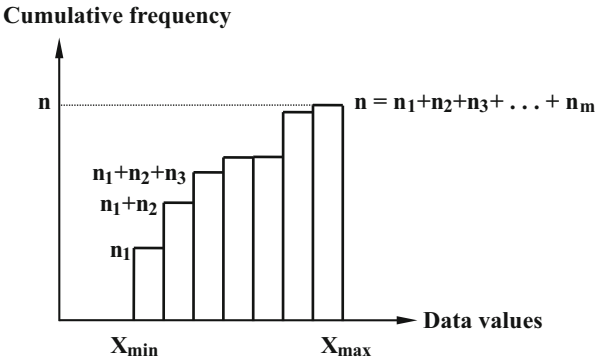


Fig. 3.4 Cumulative frequency diagrams

parameters are equal, mode = median = mean. The frequency functions in Fig. 3.3c, d are skewed to the right and left, respectively. In these two cases, mode < median < mean and mode > median > mean, respectively.

3.4.5 Empirical Cumulative Distribution Function

If the relative frequencies in Fig. 3.3 are added cumulatively, then the cumulative frequency diagram (CFD) is obtained as shown in Fig. 3.4.

The CFD does not decrease but increases always. If the same cumulative procedure is applied to frequency diagrams in Fig. 3.3, then similar cumulative frequency will emerge for each case. The benefits from the CFD are as follows:

1. The final value at the end of each CFD is equal to the number of data, n, as already given in Eq. 3.5.
2. Any value on the vertical axis that corresponds to data value on the horizontal axis in a CFD is the number of data that are smaller than the adopted data on the horizontal axis.
3. It is possible to find the number of data that falls within any desired data range.
4. The data value on the horizontal axis that corresponds to n/2 on the vertical axis is the median value of the data.

It is possible to bring the vertical axis in any frequency or cumulative frequency diagram into a data number independent form, after dividing the frequency numbers by the total number of data as

$$\frac{n_1}{n} + \frac{n_2}{n} + \ldots + \frac{n_m}{n} = 1 \tag{3.6}$$

If each term on the left-hand side is defined as the relative frequency, f_j ($j = 1, 2, \ldots, m$), then this last expression becomes

$$f_1 + f_2 + \ldots + f_m = 1 \tag{3.7}$$

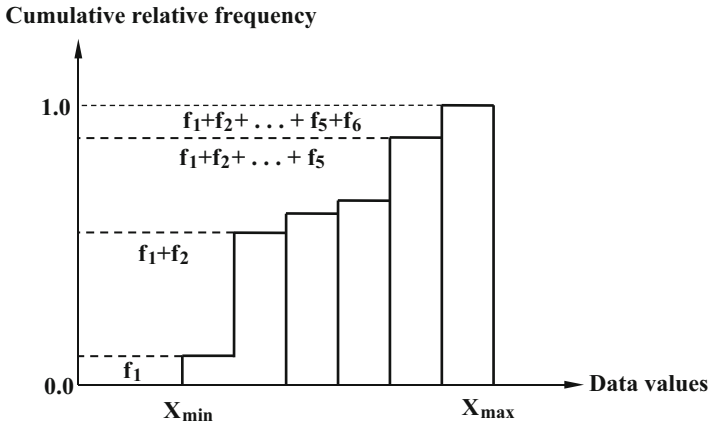


Fig. 3.5 Cumulative relative frequency diagrams

The relative frequencies are also percentages or the probability values. Therefore, their values are confined between 0 and 1. Similarly, the CFD can be converted into a data number independent case, and it is then called as the cumulative relative frequency diagram as show in Fig. 3.5.

This distribution function is the accumulation of the empirical relative frequency function. According to the categorizations in the previous subsection starting from the "low" category, the frequencies in the next category ("medium") are added, and finally, the "high" category frequencies are also added. Since in each category there is a certain number of data, and the frequency is equivalent to this number, the summation at the end reaches to the number, n, of data in the original set. For instance, if the data number (frequency) in each category "low," "medium," and "high" are n_l, n_m, and n_h, respectively, then the final number in an empirical cumulative distribution function can be expressed as follows:

$$n_l + n_m + n_h = n$$

The empirical cumulative distribution function has only one shape compared to the empirical probability distribution functions as in Fig. 3.6.

3.4.6 Histogram and Theoretical Probability Distribution Function

Although the probability is equivalent to daily usage of percentages, the problem in practice is how to define this percentage. For the probability definition, it is necessary to have different categories for the same data. In any probability study,

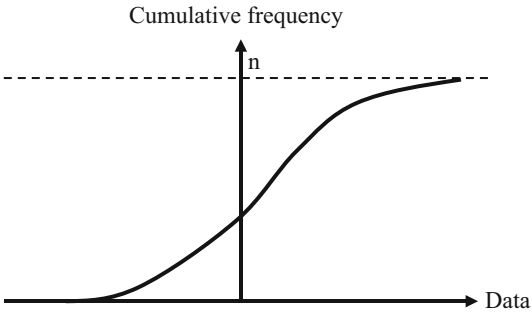


Fig. 3.6 Empirical cumulative distribution functions

the basic question is what is the percentage of data that belong to a given category? The answer to this question can be given through the three subsections depending on the data availability, conceptual model, enumeration, and personal belief about the phenomenon. Any compound probability statement can be obtained from these basic probability definitions. The conceptual understanding of these probability statements provides logical relationship derivations among them. Such relationships are possible in the case of at least two different events or phenomena.

3.4.6.1 Joint Probability

Basic probability definitions are concerned with one phenomenon; however, in practice, most often there are two or more events. The mutual consideration of these phenomena leads to some joint probability definitions. If the two events are shown as X_i ($i = 1, 2, \ldots, n_A$) and Y_j ($j = 1, 2, \ldots, n_B$), their joint probability, p_{XY}, can be defined in the light of the basic probability definitions. However, for the probability definition, categorization level for each event is required as X_0, and Y_0, respectively. Hence, each event can be categorized into two categories as $(X_i > X_0)$ and $(X_i > X_0)$ and similarly $(Y_i > Y_0)$ and $(Y_i > Y_0)$. On the basis of these four basic categories, four joint categories can be constituted as $[(X_i > X_0), (Y_i > Y_0)]$, $[(X_i > X_0), (Y_i < Y_0)]$, $[(X_i < X_0), (Y_i > Y_0)]$, and $[(X_i < X_0), (Y_i < Y_0)]$. Each pair represents a joint category from two different events, and hence, there are four different types of joint categories and probability questions. For instance, what is the joint probability of, say, $[(X_i > X_0), (Y_i < Y_0)]$ event? The answer can be given according to the classical probability distribution function definition. If n_{XY} is the number of joint events then the joint probability definition can be given as follows (Feller 1967)

$$p_{[(X_i \rangle X_0),(Y_i \langle Y_0)]} = \frac{n_{XY_x}}{n_X + n_Y} \tag{3.8}$$

3.4.6.2 Conditional Probability

Its definition also requires two different sequences as in the joint probability case. However, the question is different, as what is the probability of a categorization in one of the events given a categorization in the other event. With the notations in the previous subsection, what is the probability of $[(X_i > X_0)$ and $(Y_i < \mathrm{Y}_0)]$ given that $(Y_i < Y_0)$? This is the ratio of $[(X_i > X_0)$ and $(Y_i < Y_0)]$ event number to another event, $(Y_i < Y_0)$, number. In $[(X_i > X_0)$ and $(Y_i < Y_0)]$ statement, the number is the simultaneous occurrence of events $(X_i > X_0)$ and $(Y_i < Y_0)$, which can be shown as $n_{[(X_i \rangle X_0) \, and \, (Y_i \, \langle \, Y_0)]}$. If the number of $(Y_i < Y_0)$ event is $n_{(Y_i \langle Y_0)}$, then the conditional probability is defined as

$$p_{[(X_i \rangle X)/[(Y_i \, \langle Y_0)]]} = \frac{n_{[(X_i \rangle X_0) \mathrm{and} (Y_i \langle Y_0)]}}{n_{(Y_i \langle Y_0)}} \tag{3.9}$$

3.4.6.3 Marginal Probability

This is also based at least on two events, and the question is what percentage of a category of an event occurs by considering the other event completely. This is tantamount to saying that the marginal probability of $(X_i > X_0)$ category occurs during the whole event of Y irrespective of its categories. Hence, notationally this marginal probability can be written as

$$p_{(X_i \rangle X_o)/Y} = \frac{n_{(X_i \rangle 0)}}{n_Y} \tag{3.10}$$

3.4.6.4 Normal (Gaussian) Test

In the measurements of many uncertain events, the data had shown PDF that accords by a normal (Gaussian) distribution. Others can be transformed into a normal form by the application of various transformations (square root, logarithmic, etc.), if necessary. Many significance tests in data processing are achieved through the normal PDF application. Therefore, one should know the following properties of a normal PDF:

1. If the relative frequency distribution shows a symmetrical form, then the data are said to come from a possible normal PDF. In this case skewness coefficient is almost equal to zero.
2. In a normal PDF, there is only a single peak (mode) value, which is approximately equal to the median and the arithmetic average values.
3. A normal PDF can be represented by two parameters only, namely, the arithmetic average and the standard deviation.
4. Majority of the data are close to the maximum frequency band around the symmetry axis. Toward the right and left, the relative frequency of the data decreases continuously.

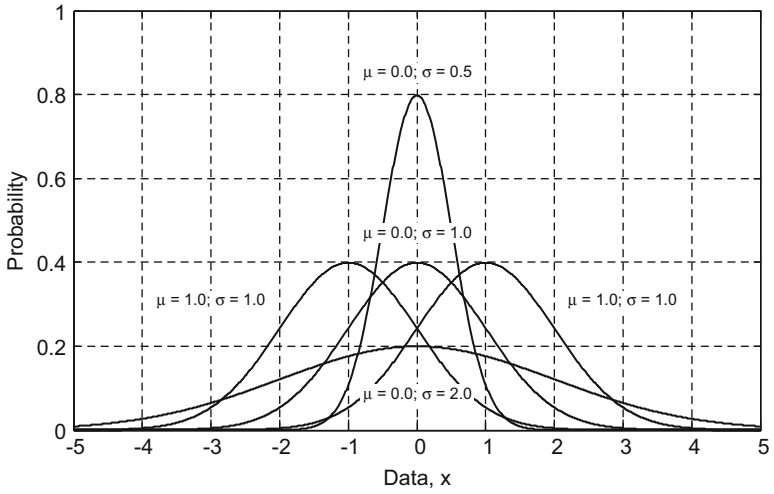


Fig. 3.7 Different normal distributions

5. Theoretically, the area under a normal PDF, as in all PDFs, is equal to 1.
6. There are two extreme regions as tails on the right and left. These tails extend to $-\infty$ and $+\infty$, but in practical works, the smallest and the biggest values are finite. Mathematically, a normal distribution function is given as

$$f(X) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right] \tag{3.11}$$

where μ and σ^2 are the arithmetic average and the variance of the data. The geometrical exposition of this expression is given in Fig. 3.7, where μ and σ are the arithmetic average and standard deviation parameters, respectively.

Since the area calculations from this expression are mathematically impossible, the areas are calculated by numerical techniques and given in Table 3.1. The areas in this table are from $-\infty$ up to the variable value, x. The subtraction of these areas from 1 leads to significance level values. By means of this table, if the significance level is given, then the corresponding area can be found.

In order to use this table, it is necessary that after the data confirmation with the normal distribution, the data values must be standardized according to the following expression:

$$x = \frac{X-\mu}{\sigma} \tag{3.12}$$

where X is the given data and x is the standard data. The arithmetic average is equal to 0 and variance to 1. For a normal distribution test, the following steps must be executed:

Table 3.1 Standard normal (Gaussian) distribution function

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.004	0.008	0.012	0.016	0.019	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706

1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499
3.1	0.499	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

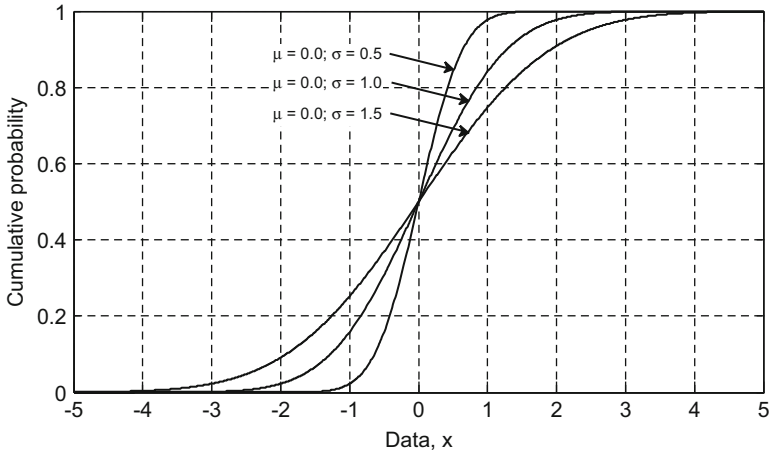


Fig. 3.8 Cumulative probability distribution functions

1. The data sequence must be standardized by the use of its arithmetic average and standard deviation value. This standard variate is referred to as the test quantity. Standardization gives rise to dimensionless data values such that they have zero mean and unit variance. The standard normal PDF has the following form with no explicit parameter values:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{3.13}$$

2. It is necessary to decide about the confidence interval. For this purpose, the significance level may be taken as 5 % or in some approximate works as 10 %.

3.4.7 Cumulative Probability Distribution Function

Any PDF integrated from the lower limit to the upper one leads to similar curves in the form of S shape, which varies between zero and one (see Fig. 3.6). In practical studies, zero and one correspond to impossible or completely deterministic events, respectively. Hence, their consideration in the earth sciences is out of question. All other values exclusively between zero and one are valid in any discipline that evolves temporally or spatially randomly. Some of the cumulative distribution functions, CDFs, corresponding to the PDFs in Fig. 3.7 are given in Fig. 3.8.

The CDF curve provides information of non-exceedance probability from the vertical axis for a given data value on the horizontal axis. In this way, it is also possible to know exceedance probability value also, because it is the complementary of the non-exceedance probability value. In practical terms, the probability of non-exceedance is numerically equivalent to safety or reliability measure, whereas

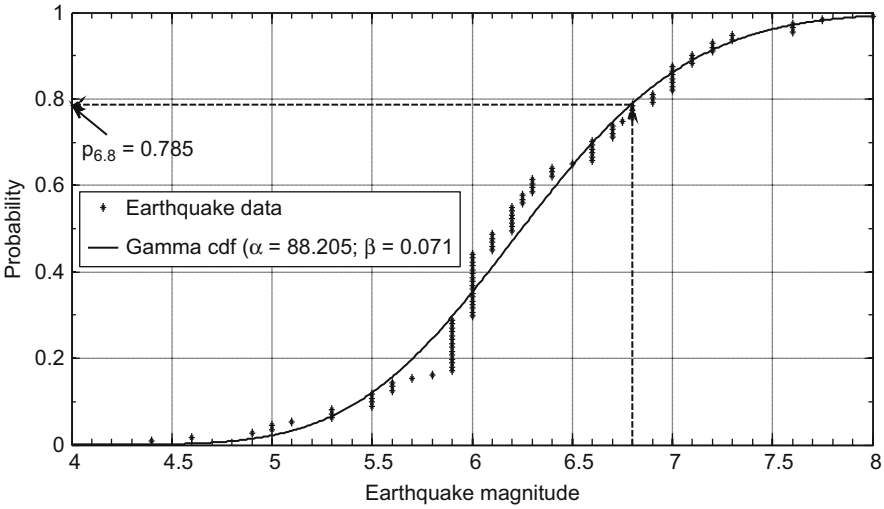


Fig. 3.9 Turkish earthquake probability models

the exceedance probability value reflects the risk quantity. The PDF or CDF provides the simplest modeling in any uncertainty event involvement. In the probabilistic modeling, the occurrence mechanism of the phenomenon is assumed completely independent. Such models are very important in many projects and engineering designs against natural or artificial uncontrollable events such as storms, floods, droughts, and earthquakes.

3.4.8 Prediction Methods

As already mentioned in the previous subsection, probability models are equivalents of independent event prediction models by considering a predesigned risk or reliability level. The following items suffice for the application of probability prediction models:

1. Data may be temporal or spatial and also regular or irregular.
2. Data sequence is ordered in ascending order from the smallest to the biggest; hence, each data has its location within the number of given data size, say, n.
3. Obtain scatter diagram of the ordered data as in Fig. 3.9.
4. The location in the ordered sequence is referred to as the rank of the data, say it is denoted by m.
5. Ask the question, what is the percentage (relative frequency, probability) of data that is smaller than any data value? The answer is the number of data less than the data considered divided by the original data size.

6. Since the rank, m, of each data indicates the number of data less than or equal to this data value, then the probability of non-exceedance, P_n, (safety, reliability) can be calculated as

$$P_n = \frac{m}{n}(m = 1, 2, \ldots, n) \tag{3.14}$$

As mentioned earlier certain events cannot enter the probability calculations, and therefore, when $m = n$, then $P_n = 1$, which is not acceptable, because this definition does not leave room for future probable occurrences. In order to alleviate this situation, Eq. 3.14 modifies as follows:

$$P_n = \frac{m}{n+1}(m = 1, 2, \ldots, n) \tag{3.15}$$

This final definition excludes all deterministic cases out and whatever the data size P_n can never reach 1.

6. P_n can never reach 1; big data sizes leave lesser uncertainty domain, which is plausible, because the more the data, the better is the information about the phenomenon and the better is the possibility to make reliable predictions.
7. The exceedance probability, P_e, (risk) based on the available data can be calculated empirically as the subtraction of the non-exceedance probability from 1, which gives

$$P_e = 1 - \frac{m}{n+1} = \frac{(n-m)+1}{n+1}(m = 1, 2, \ldots, n) \tag{3.16}$$

This expression leaves room left for the uncertainty as $1/(n+1)$, and it approaches to zero for $n \to \infty$, which is never possible in practical earth sciences events. This is the reason why nature is bound to break records.

8. After all these discussions, one can write simply that

$$P_n + P_e = 1 \tag{3.17}$$

9. The discussions so far are based on the data availability, and the scatter diagram is shown in Fig. 3.9. In order to generalize the scatter diagram to possible population of available data, it is necessary to fit a theoretical PDF. The theoretical PDF is taken as a two-parameter gamma function, and the final result is given in Fig. 3.9. On the same, figure the shape α and scale β parameter values are also given. This figure is based on 100 earthquake magnitude data from different locations in Turkey (see Table 3.2).

It is possible to make prediction from this figure for the non-exceedance (safety, reliability) probability corresponding to earthquake magnitude 6.8 from the vertical axis as $p_{6.8} = 0.785$.

Further information about the spatial modeling on the earthquake events can be obtained from a paper by Şen.

Table 3.2 Turkish earthquake data

Easting	Northing	Magnitude	Easting	Northing	Magnitude
40.30	38.40	5.10	40.80	33.20	7.60
38.00	27.00	6.00	39.90	29.26	6.00
39.00	39.00	6.30	39.31	26.70	6.80
40.50	42.70	6.20	37.11	35.70	6.00
38.00	26.50	6.00	38.41	43.76	4.90
40.18	38.10	6.75	38.30	32.20	5.60
42.50	26.40	5.90	39.33	41.10	5.90
41.00	34.00	6.20	38.56	26.21	6.60
36.00	30.00	6.25	39.40	40.90	6.70
40.65	27.20	7.75	37.11	35.69	6.00
40.60	27.10	6.40	40.60	33.60	7.00
40.10	26.80	6.90	39.99	41.50	6.00
38.00	30.00	6.90	40.01	27.49	7.20
40.27	36.38	7.10	40.94	33.13	6.00
39.26	26.71	7.00	34.80	32.50	6.30
35.50	34.00	5.80	37.65	27.26	7.00
39.70	42.80	5.30	39.82	30.51	6.00
39.96	41.94	6.80	36.43	28.63	6.80
38.55	30.78	5.90	36.42	28.68	7.10
41.33	43.41	6.00	36.22	28.87	5.90
38.00	30.50	5.90	40.58	31.00	7.00
37.03	29.43	6.10	40.61	30.85	6.20
35.84	29.50	7.00	40.63	30.93	6.00
36.54	27.33	7.30	36.94	28.58	5.90
40.94	43.88	6.00	39.16	41.53	5.00
38.10	27.10	6.50	36.70	28.49	6.60
40.50	26.50	6.10	35.45	33.44	6.00
40.20	37.90	6.10	39.96	44.13	5.30
37.98	44.48	7.60	37.96	29.14	5.50
37.97	45.00	6.30	40.71	29.09	6.20
39.72	39.24	5.60	37.33	29.82	4.40
36.77	27.29	6.70	38.06	38.35	5.50
38.20	29.70	6.00	40.10	40.90	4.60
39.00	41.00	5.90	40.20	28.20	6.60
40.67	27.44	6.40	37.85	29.32	5.70
40.60	27.49	6.30	39.20	41.60	5.50
39.50	43.00	6.00	39.19	41.48	6.80
39.53	33.95	6.60	39.42	40.98	6.20
39.00	26.90	6.60	40.67	30.69	7.20
39.70	40.40	5.90	39.50	40.40	5.60
39.80	39.38	8.00	39.20	44.30	5.30
39.66	35.83	6.25	41.79	32.31	6.60

(continued)

Table 3.2 (continued)

Easting	Northing	Magnitude	Easting	Northing	Magnitude
39.70	35.40	5.90	36.18	29.20	6.00
37.07	28.21	6.20	38.59	28.45	6.40
38.47	43.30	6.00	39.18	29.49	7.10
39.86	39.37	5.90	37.59	29.76	5.90
37.13	28.06	6.10	38.83	40.52	6.70
40.80	27.80	6.00	38.47	40.72	6.70
39.03	27.74	5.90	37.70	28.89	5.00
39.35	28.06	6.20	39.12	44.03	7.30
40.82	34.44	5.50	36.40	31.75	5.90
40.75	34.75	5.90	39.24	25.28	7.20
40.66	36.35	7.00	40.32	27.22	6.10
40.80	30.60	6.25	40.14	24.77	7.00
41.00	34.00	7.60	40.33	42.18	6.90

3.5 Temporal Data Set Modeling

Earth sciences branches data can be modeled temporally and spatially depending on the purpose of the study or spatio-temporal models can also be developed. In each one of these models, temporal or spatial dependences must be taken into consideration. If data are available at regular time instances, then time series models play the best role.

3.5.1 Time Series Analysis

In earth sciences, variable measurements show randomness and they cannot be predicted from the previous data values with certainty. This is the main problem in modeling as to which value to adopt in the final design. In the beginning of the twentieth century, Hazen (1924) has suggested the use of the following procedure without the availability of any computer and even calculator. His method is random drawing of pieces of paper mixed in a sack. If the past records of an event are denoted as a sequence, $X_1, X_2, \ldots, X_n$, it is possible to calculate its various statistical parameters such as the arithmetic average, standard deviation, serial correlation coefficient, empirical frequency distribution, histogram, etc. This sequence is the naturally ordered record of measurements, and the statistical parameters are dependent on the whole data, and they are valid for the record number, n. These historical data series can be used for the construction of simple future prediction values according to the following steps:

(a) Historical data record: There may be regular daily, monthly, or annual records of past variable measurements or irregular data values. Let the number of records be n.
(b) Each one of these measurements is written separately on equal size pieces of paper. They are folded and then put into a sack.

(c) These pieces of paper are drawn one after each other, and hence a new time series is constructed of the same duration with the same data values but at random times.
(d) After each drawing, the papers are either returned to the sack or not. In the former case, it is possible to generate sequences as long as desired. However, in the latter, the maximum length of the synthetic data can be equal to the length of the original data. After the completion of generation, whole papers are returned back into the sack, and likewise another sequence of random numbers can be generated.

After the completion of such a generation procedure, the input sequences are treated for the assessment of decision or design quantity. For instance, if the sequences are water discharge values, the necessary decision can be reached provided that the demand level is known. It is possible to decide about the water sufficiency or insufficiency over time durations. In the cases of insufficiency, additional water is withdrawn from other water storages.

The sequences obtained by this deterministic-uncertain method can be referred to as time series variable synthetic sequences, because data values are the same as in the measurement series. All these synthetic sequences have the following points in common:

1. Each synthetic sequence has the same arithmetic average as the original sequence.
2. Each synthetic sequence has the same variance and the standard deviation as the original sequence.
3. Other statistical parameters (mode, median, skewness, kurtosis, etc.) are also the same in addition to the relative frequencies, hence, the probability distribution function of the data variable.
4. The major assumption in such a withdrawal system is that each one of the generated sequence is regarded as independent from others. Each one of the generated sequence has its own serial correlation coefficients that are different from each other. The general procedural task of this deterministic-uncertain methodology is shown in Fig. 3.10.

3.5.1.1 Independent Process

If the successive observations are independent, then there are two simple and complementary events and their probabilities, namely, the probability of occurrence, $p = P(x > Q)$, and nonoccurrence, $q = 1 - p = P(x \leq Q)$, from which it is possible to evaluate various probabilities of any compound event of the phenomenon considered. In an independent process case, the safety, S, can be defined as the multiplication of n probability of non-exceedance, where n is the time series duration considered. In general, S is given as a compound event, and in terms of the simple event probabilities, it becomes

$$S = q^n \tag{3.18}$$

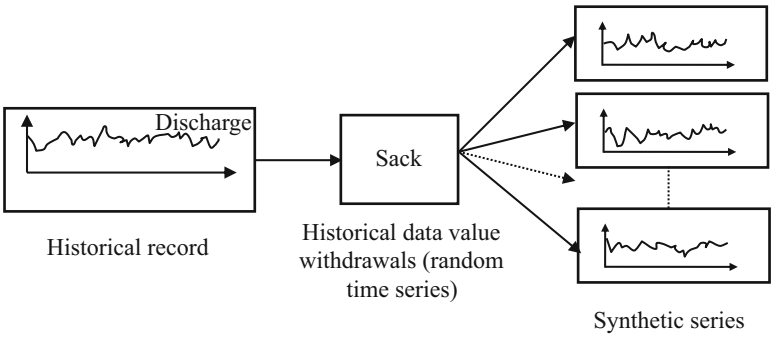


Fig. 3.10 Deterministic-uncertain method sequences

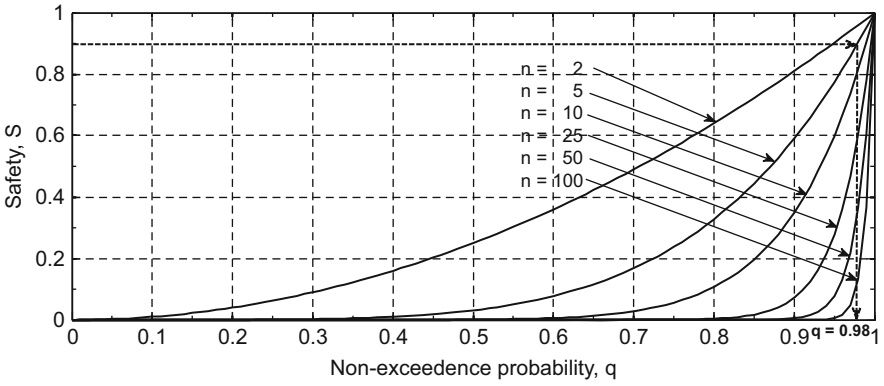


Fig. 3.11 Non-exceedance probability versus safety

The corresponding risk, R, is the complementary probability as

$$R = 1 - q^n \tag{3.19a}$$

Equation 3.18 yields to a set of curves as in Fig. 3.11. This graph is very useful in practical applications of systems design in calculating the probability of nonoccurrence of design variate, D, given the level of safety and the expected life, n, of the project.

For instance, if the planner is interested in designing his project for $n = 10$ years with 90 % safety, that is to say, in the long run the project will not fail 90 % of the time, then from Fig. 3.11, $q = 0.98$ is obtained; hence, the probability of occurrence is $p = 0.02$. The planner may be able to find the magnitude of design variate by adopting a suitable PDF to data at hand with $p = 0.02$. For example, if the adopted PDF is normal, then the standard design variate corresponding to $p = 0.02$ is $x \cong 2.0537$, whereas the actual design variate can be found as

$$Q = \mu + x\sigma \tag{3.19b}$$

where μ and σ are the mean and standard deviation of the available time series. Thus, the resulting design variate is safe at 99 % level.

On the other hand, if the expected life of the project is to be confined, say, to 20 years with $q = 0.99$, then the safety can be found from Fig. 3.11 as 80 %. Hence, as expected, project life increases as safety of the design decreases. This is logical, since as the project life increases, the probability for exceedance of design variate also increases.

Sometimes, it is desirable to know the safety of an already existing engineering structure. For instance, if the structure has been designed originally for an expected life of $n = 30$ years with $q = 0.99$, after its completion, the calculated safety will be 73.9 %. If the design value has not been exceeded, say for 20 years, then the safety of the same structure for the remaining 10 years will be 90.0 % which is considerably greater than the original safety. An important conclusion is that as the number of years without any hazard increases, the safety of structure will improve reducing the risk.

Probabilistic Modeling

In the design of any hydraulic structure such as dams, the lifetime is adopted for the maximum flood occurrence as 25 years, 50 years, 100 years, 500 years, or 1000 years. In the flood calculation by the probability method, there are two different approaches. The first one is theoretical, whereas the second approach is empirical. In the empirical approach, the floods are assumed to originate from the same population in an independent manner. For this purpose, the maximum floods of each year must be recorded. In order to attach probability value to each flood discharge, the historical flood records must be ranked in the ascending order. If there are n years, in such a ranking, there will be n measurements ordered in ascending sequence. Starting from the smallest flood discharge, the ranks are attached to each one in ascending order as 1, 2, 3, …, n. Hence, the smallest flood discharge has its rank as 1 and the greatest one as n. By considering the number of measurements and the order, each flood discharge can be given a certain probability value according to the Weibull formulation as

$$P_n(Q > Q_T) = \frac{m}{n+1} \tag{3.20}$$

According to this formulation, the probability of exceedance is big for small discharges and vice versa. The probability of exceedance is also referred to as the risk level. In Fig. 3.12 the relationship is given between the exceedance probability and the time series record values.

After the fitting of a straight line to the scatter of points on the risk-rank coordinate system, it provides the ability to extend the straight line along the big

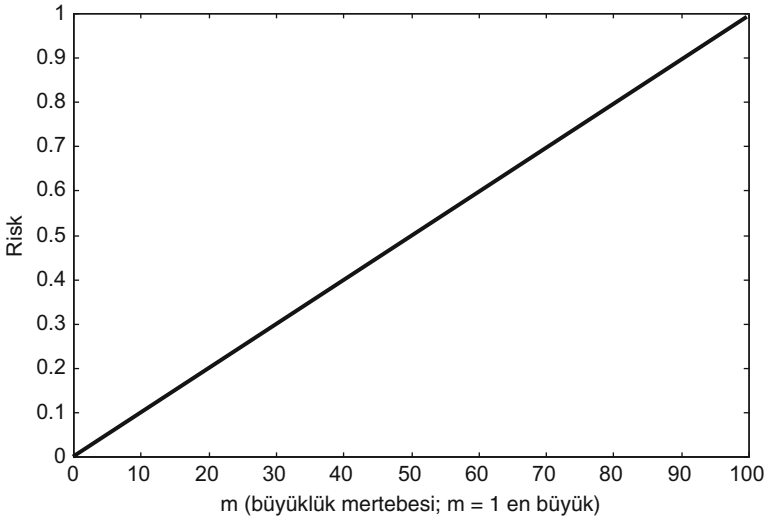


Fig. 3.12 Flood exceedance probability

and small discharge domains, and hence one can estimate the risk for any given design discharge and vice versa.

Flood frequency analysis is the name of this procedure in assessing the flood estimations in engineering projects. There are different theoretical probability distribution functions for such an assessment as Gumbel, gamma, lognormal, Pearson, and other functions.

Statistical Modeling

Statistics is concerned with a set of parameters of a given data set and also any model in the form of mathematical expressions and their parameter values. Statistics is the branch of mathematics, where reliable and significant relationships are sought among different causative and the consequent variables. As for the statistical uncertainties of parameters are concerned, they appear either as random variations or in the biased forms of the statistical parameters. Model parameters such as the intercept and slope values of any simple regression line are determined depending on a given sample of finite length. As the sample length increases, the reliability in the parameter estimations also increases.

This procedure is more developed than the previous one, and instead of using the same data values in synthetic sequence generation, the relative frequency distribution of the measured data is adopted. This histogram is then fitted with the most convenient theoretical PDF. This time the data are not drawn from the sack with historical data, but the synthetic sequence values are drawn from the theoretical PDF. This approach yields synthetic sequences that have in the long run almost the

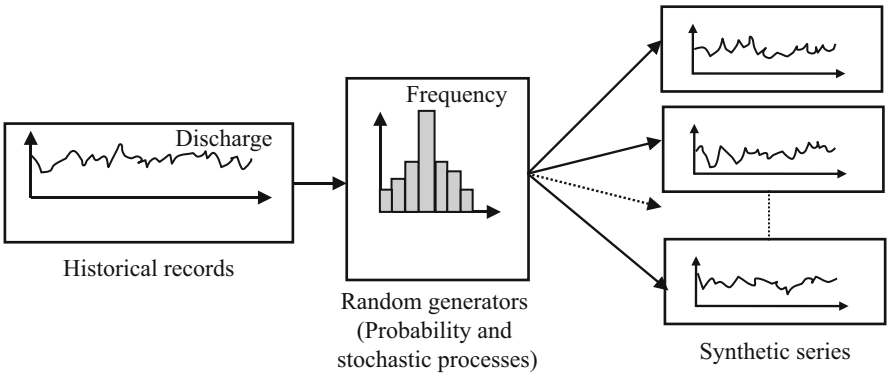


Fig. 3.13 Basic stages in the statistical generation procedure

same statistical parameters but it also brings the contribution of the extreme values into the generation procedure which is not possible with the deterministic-uncertainty method (Fig. 3.10). The fitting of a theoretical PDF to the relative histogram of the data can be achieved through the Chi-square test (Benjamin and Cornell 1970). The statistical synthetic time series generation basic steps are shown collectively in Fig. 3.13.

It is considered that the sequence of class intervals is the same at each time instant as shown in Fig. 3.13, but there are transition probabilities or relative frequencies also from one class interval at, say, t instant to $t+1$ instant. Hence, if there is m classes intervals, there will be m class interval relative frequencies, which are referred to herein as the state probabilities. Furthermore, there are mxm interclass interval transition probabilities, i.e., relative joint frequencies. Hence instead of the statistical parameters, the state and transition probabilities are used in the modeling of a given time series. These models are known in the literature as the Markov chain models. Their application requires the following steps:

(a) The histogram of the time series variable is obtained.
(b) Histogram implies that the state variables in each class interval are already calculated.
(c) The calculation of transition relative frequencies requires consideration of two successive time instances. As stated before, m intervals imply mxm transition probabilities, which can be considered in the form of a matrix, where rows are for time instant t and columns for $t-1$. Such a matrix is called as the transition probability matrix. The transition from state, say, s_i at time t, to state, s_j at time $t-s$, is the same as the transition from state s_j at time $t-1$ to s_i at time t, and the transition matrix will have a diagonal symmetric form. Hence, $m(m-1)/2$ transition probabilities are necessary for the construction of the transition probability matrix. The transition probabilities along the major diagonal are all equal to 1, because they represent the transition from a state to itself. The state and its transition matrix provide the basis of future phenomenon prediction. This procedure is presented schematically in Fig. 3.14.

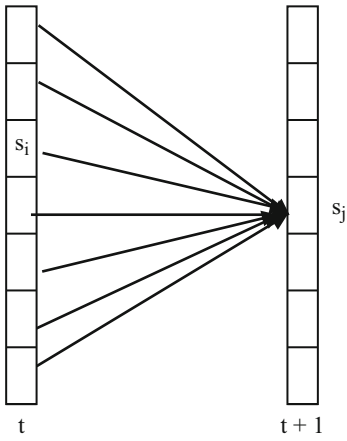


Fig. 3.14 State and transition probabilities

If the number of states is equal to n, then the transition probability matrix appears in the following form:

$$P_T = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & \cdot & \cdot & \cdot & \cdot & \cdot & p_{1n} \\ p_{21} & p_{22} & p_{23} & p_{24} & \cdot & \cdot & \cdot & \cdot & \cdot & p_{2n} \\ p_{31} & p_{32} & p_{33} & p_{34} & \cdot & \cdot & \cdot & \cdot & \cdot & p_{3n} \\ p_{41} & p_{42} & p_{43} & p_{44} & \cdot & \cdot & \cdot & \cdot & \cdot & p_{4n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{n1} & p_{n2} & p_{n3} & p_{n4} & \cdot & \cdot & \cdot & \cdot & \cdot & p_{nn} \end{bmatrix}$$

3.5.1.2 Dependent Processes

The general form of these processes has a deterministic functional form, $f_D(.)$, in addition to the random function, $f_R(.)$, as follows:

$$X_t = f_D(X_{t-1}, X_{t-2}, X_{t-3}, \ . \ .X_{t-k}) + f_R(\varepsilon_t) \tag{3.21}$$

If the implicit deterministic function is explicitly written in the linear term summation form, Eq. 3.21 takes the following form:

$$X_t = a_1 X_{t-1} + \mathrm{a}_2 X_{t-2} + a_3 X_{t-3}, + \ldots + \mathrm{a}_k X_{t-k} + \varepsilon_t \tag{3.22}$$

where a_i ($i = 1, 2, \ldots, k$) are the model parameters. Since the current value is a linear function of k previous values, this is referred to as lag-k Markov process. Its

simplest form appears when $k=1$, and this is the lag-1 or first-order Markov process as

$$X_t = a_1 X_{t-1} + \varepsilon_t \tag{3.23}$$

This is a single parameter model, which can be written in terms of given data statistical parameters of arithmetic mean, μ, standard deviation, σ, and the serial correlation function, ρ, as

$$(X_t - \mu) = (X_{t-1} - \mu) + \sigma\sqrt{1-\rho^2}\varepsilon_t \tag{3.24}$$

After the division of both sides by standard deviation, the model takes the standard form as,

$$x_t = x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t \tag{3.25}$$

Herein, x_t, $x_{t\text{-}1}$, and ε_1 have zero arithmetic average and unit variance; and ε_1 has independent (serial correlation zero) normal PDF variable. The serial correlation coefficient is estimated according to the Pearson correlation coefficient definition (Benjamin and Cornell 1970).

Markov processes are also known as autoregressive (AR) models because the deterministic part is in the form additive summations with parameters that can be calculated through the regression procedure.

3.5.1.3 Autoregressive Integrated Moving-Average (ARIMA) Processes

These have both the deterministic and error terms together. Its difference from the Markov processes is that the error term is not only a single value but should have at least two error terms. Hence, these models are a transition between the Markov and moving-average models. Another property of the ARIMA processes is that they are used for successive differences of the measured historical data.

Besides, the ARIMA (1,0,1) processes are capable for representing long memory effects in natural phenomenon. In general, two subsequent values, Z_i and $Z_{i\text{-}1}$, of the phenomena concerned are related recursively to each other as

$$Z_i = \phi Z_{i-1} + \varepsilon_i - \theta\varepsilon_{i-1} \tag{3.26}$$

in which ϕ and θ are model parameters and $\varepsilon_{i'}$s are random variables with Gaussian distribution. Notice that for $\phi=0$ Eq. 3.26 yields MA process, and when $\theta=0$, it reduces to an AR process which is commonly known as a first-order Markov process, and, finally, if $\phi=\theta=0$, then the resulting process is purely random which is usually referred to as white noise, i.e., IP. In the case of AR process, $\phi=\rho_1$ where ρ_1 is the lag-one autocorrelation coefficient of this process. A common property in all of these processes is that they are stationary. Last but not

least, when $\phi = 1$ and $\theta = 0$ Eq. 3.26 leads to random increments, $Z_i - Z_{i\text{-}1} = \varepsilon_i$, which is known as either random walk or Brownian motion process. As stated by Jenkins and Watt (1968), this process is nonstationary in both the mean and the variance.

The autocorrelation structure of ARIMA (1,0,1) process is given in terms of ϕ and θ by Box and Jenkins (1970) explicitly as

$$\begin{aligned} \rho_0 &= 1 \\ \rho_1 &= \frac{(\phi - \theta)(1 - \phi\theta)}{1 + \theta^2 - 2\phi\theta} \\ \rho_i &= \phi\rho_{i-1} (i > 1) \end{aligned} \tag{3.27}$$

in which ρ_0 and ρ_i are the lag-zero and lag-i autocorrelation coefficients.

3.5.1.4 Moving-Average Processes

If the deterministic part in Eq. 3.21 is not considered, then the explicit form of the stochastic process will be as

$$X_t = f_R(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \varepsilon_{t-m}) \tag{3.28}$$

This is the general form and again its explicit expression can be thought as a linear formulation as

$$X_t = b_1\varepsilon_t + b_2\varepsilon_{t-1} + b_3\varepsilon_{t-2}, \ldots b_m\varepsilon_{t-m} \tag{3.29}$$

which means that the data value is composed of independent random error term combination at different previous times. The parameters b_i ($i = 1, 2, \ldots, m$) show the contribution of each error on the data value. In general, this is an mth order moving-average stochastic model. Its simplest form is a first-order moving-average stochastic process which is referred to as the independent process or white noise stochastic process which can be expressed as

$$X_t = b_1\varepsilon_t \tag{3.30}$$

3.6 Empirical Correlation Function

Correlation coefficients are useful in the determination of the relationship strength between two time series. In a time series, the successive values might affect each other with significant dependence. For instance, an event of today might be affected at least partially yesterday's event occurrence. In general, high rainfall values

follow high values, and low values follow low values. This statement indicates that, so far as the event occurrences and successive values are concerned, there are successive relationships to a certain extent. Even a nonspecialist can feel such dependences through experience.

In mathematics, when two variables are related with each other, their variation on a Cartesian coordinate system does not appear as a horizontal or vertical line, but rather a line with a linear trend and slope or as a curve with many tangential slopes. The simplest form of dependence is of the linear type, which is used in the statistical or stochastic modeling works. For instance, ρ in Eq. 3.25 is of this linear type.

If there is one time series such as $X_1, X_2, \ldots, X_n$, then two time series can be generated from it so that they can be plotted against each other on a Cartesian system. In order to achieve this, the original time series is shifted, say for instance lag-one, and the two concurrent series are obtained as follows:

$$X_1, X_2, X_3, \ldots\ldots\ldots, X_n$$
$$X_2, X_3, X_4, \ldots\ldots\ldots, X_{n-1}$$

In each series there are $n-1$ corresponding points. The scatter diagram of these two time series gives rise to $n-1$ point on the Cartesian coordinate system as shown in Fig. 3.15.

If straight line appears through the points, then it is possible to conclude that there is dependence between the two time series, otherwise they are independent. The most suitable straight line matched through these scatter points gives the dependence measurement as its trend slope. The more the deviation of the slope from 45°, the smaller is the dependence. In Fig. 3.16 an independent scatter diagram is shown.

It is possible to make test with the correlation coefficient. In the statistical literature, this is known as Wald-Wolfowitz test. In this nonparametric test, standard data series is used and for this purpose, first of all, the average, μ, and the standard deviation, σ, of the given time series are calculated, and then the given time series is standardized according to Eq. 3.12. Subsequently the Pearson correlation coefficient, ρ, is calculated according to the following formulation:

$$\rho = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i x_{i+1} \tag{3.31}$$

The correlation coefficient takes values between -1 and $+1$. The closer the coefficient to zero, the more random, i.e., independent, is the internal structure of the time series; otherwise, close values to $+1$ imply positively strong or -1 negatively strong correlations. Positive (negative) correlation means direct (inverse) proportionality. In the case of positive dependence, high (low) values follow high (low) values, whereas in the case of negative dependence, high (low) values follow low (high) values. The dependence that is calculated through Eq. 3.31 is the serial correlation or autocorrelation coefficient. Similar to lag-one, lag-two or more lags

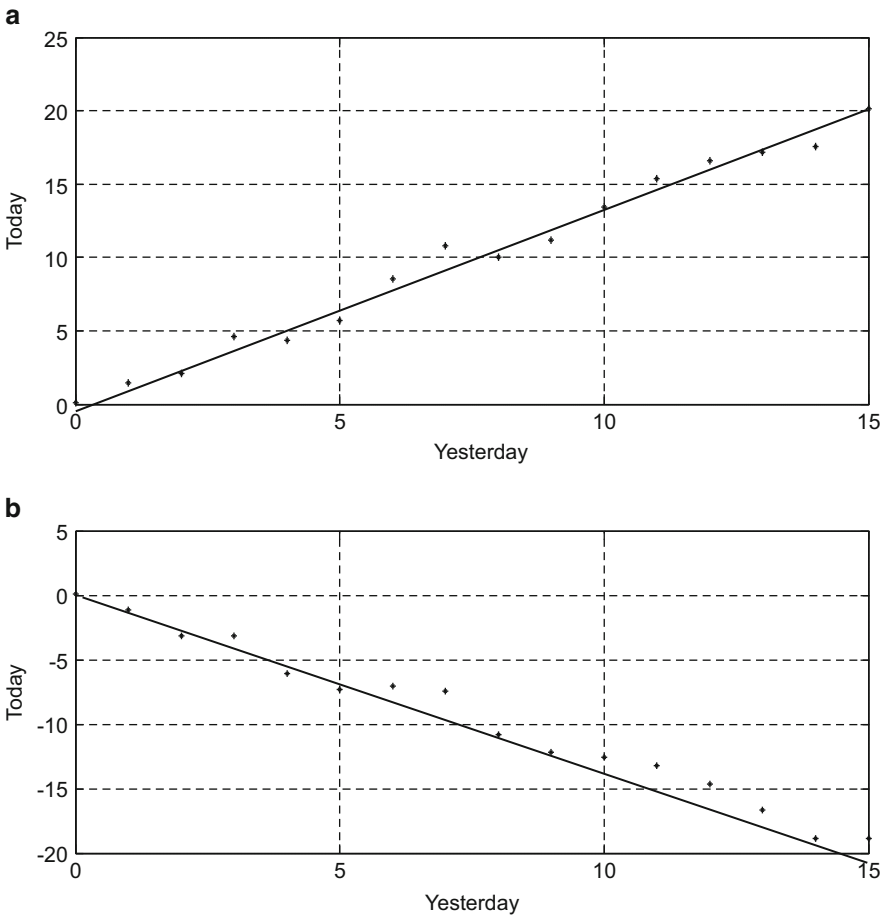


Fig. 3.15 Dependence scatter diagrams. (**a**) Directly proportional. (**b**) Inversely proportional

can also be calculated simply for a given time series. Hence, in general lag-k (in practice, care must be taken such that $k \leq$ n/3) can be calculated similar to Eq. 3.31 as

$$\rho_k = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i x_{i-k} \tag{3.32}$$

The qualitative linguistic interpretations of the numerical correlation coefficient values are presented in Table 3.3.

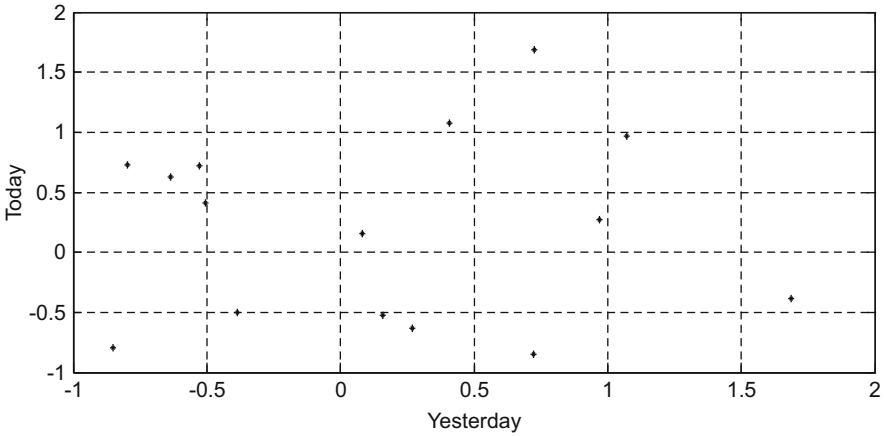


Fig. 3.16 Independent scatter diagrams

Table 3.3 Correlation coefficient numerical and linguistic interpretations

Numerical value intervals	Linguistic interpretations
$\rho = -1.0$	Perfect negative dependence
$-1.0 < \rho < -0.9$	Strong negative dependence
$-0.9 < \rho < -0.7$	Quite negative dependence
$-0.7 < \rho < -0.5$	Weak negative dependence
$-0.5 < \rho < -0.3$	Very weak negative dependence
$-0.3 < \rho < -0.1$	Insignificant negative dependence
$\rho = 0.0$	Complete independence
$0.1 < \rho < 0.3$	Insignificant positive dependence
$0.3 < \rho < 0.5$	Very weak positive dependence
$0.5 < \rho < 0.7$	Weak positive dependence
$0.7 < \rho < 0.9$	Strong positive dependence
$\rho = 1.0$	Perfect positive dependence

The empirical serial correlation coefficient is regarded as distributed according to a normal (Gaussian) PDF by Anderson (1942) with zero mean and variance equal to the inverse of $(n-1)$. Hence, with these parameters, the significance of the empirical correlation coefficient can be tested by a single-tail normal PDF.

References

Anderson RL (1942) Distribution of the serial correlation coefficient. Ann Math Stat 13:1–13
Benjamin JR, Cornell CA (1970) Probability, statistics and decisions for civil engineers. McGraw-Hill, New York

Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden Day, San Francisco, p 475
Davis JC (1986) Statistics and data analysis in geology. Wiley, New York
Feller W (1967) An introduction to probability theory and its application. Wiley, New York, p 509
Hazen A (1924) Flood flows: a study of frequencies and magnitudes. Wiley, New York
Jenkins GM, Watts DG (1968) Spectral analysis and its applications. Holden-Day, San Francisco, p 525

Chapter 4
Classical Spatial Variation Models

Abstract Men have tried to model spatial behavior of the natural phenomena since a long time with initiative simple models such as the weighting functions, which are supposed to represent regional dependence structure of the phenomenon concerned. Unfortunately, commonly employed weighting functions are not actual data dependent, and hence they are applicable invariably in each spatial prediction, which is not convenient since each spatial phenomenon will have its own spatial dependence function. Spatial data distribution can be uniform, randomly uniform, homogeneous, isotropic, clustering, etc. which should be tested by a convenient test as described in the text. Besides, statistically it is also possible to depict the spatial variation through trend surface fit methods by using least squares technique. Finally in this chapter, adaptive least squares techniques are suggested in the form of Kalman filter for spatial estimation.

Keywords Cluster sampling • Data group • Geometric weights • Kalman filter • Nearest neighbor • Random field • Spatial pattern • Trend surface • Uniformity test

4.1 General

The spatial nature of earth sciences phenomena expelled the researchers to explore spatial statistical procedures, whereas the classical statistics remained at the service as usual. In general, any phenomenon with spatial variations is referred to as the ReV by Matheron (1963). ReV fall between the random fields where the spatial variations have independence and the deterministic variability depending on the spatial correlation value. The most significant methodologies of the spatial analysis is Kriging, which is discussed in Chap. 5, but prior to their development, earth scientists were using different empirical and geometrical rational approaches in assessing the ReV. Hence, in this chapter these rather preliminary and simple but effective methods will be discussed with their drawbacks. The trend surface analysis that is an offshoot of statistical regression Kriging is related to time series analysis, and contouring is an extension of interpolation procedures.

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_4

4.2 Spatiotemporal Characteristics

Our minds are preconditioned on the Euclidean geometry, and consequently, ordinary human beings are bound to think in 3D space as length, width, and depth in addition to the time as the fourth dimension. Hence, all the scientific formulations including differential equations and others include space and time variability. Earth, environmental, and atmospheric variables vary along these four dimensions. If their changes along the time are not considered, then it is frozen in time, and therefore, a steady situation remains along the time axis but ReV in concern has variations in the space. A good example for such a change can be considered as geological events, which do not change appreciably during human lifetime. For instance, iron content of a rock mass varies rather randomly from a point to another within the rock mass, and hence spatial variation is considered. Another example is the measurement of rainfall amounts at many irregularly located meteorology stations spread over an area, i.e., simultaneous measurement of rainfall amounts, and again the time is kept constant and the spatial variations are sought.

On the other hand, there are time variations, which are referred to as the temporal variations in the natural phenomenon. For such a variation, it suffices to measure the event at a given location, which is the case in any meteorology station or in groundwater wells. Depending on the time evolution of the event whether it is continuous or not, time series records can be obtained. A time series is the systematic measurement of any natural event along the time axis at regular time intervals. Depending on this time interval, time series is called as hourly, daily, weekly, monthly, or yearly time series.

Contrary to time variations, it is not possible to consider space series where the records are kept at regular distances except in very specific cases. For example, if water samples along a river are taken at every 1 km, then the measurements provide a distance series in regularity sense. In fact, distance series are very limited as if there are no such data sequences in practical studies. On the other hand, depending on the interest of event, they appear as time series data but they are not time series due to irregularity or randomness in the time intervals between successive occurrences of the same event. Flood and drought occurrences in hydrology correspond to such cases. One cannot know the duration of floods or droughts. Likewise, in meteorology the precipitation occurrence or any dangerous levels of air pollutant concentrations do not have time series characteristics.

Any natural event evolves in the 4D human visualization domain, and consequently its records should involve the characteristics of both time and space variability. Record with this property is said to have spatiotemporal variation. It is the main purpose of this chapter to present the classical methodologies for spatial data treatment regional estimations and interpretations based on simple but effective methodologies.

4.3 Spatial Pattern Search

Human is biased toward regular pattern geometries, appearances, and presentation rather than irregular or random features. Although at the first instance he/she can distinguish visually between the regular and irregular patterns in the space, he/she may not know their quantitative assessment methodologies at the time. For instance, every day the pattern of clouds is different than the previous days, and the most regular ones are either overcast situation or cloudless patterns. However, in between there are millions of different irregular patterns that cannot be distinguished simply from each other in fine scales. It is, therefore, necessary to sample such situations at a set of irregular measurement sites and then use quantitative methodologies for the identification of a set of parameters or a spatial map for each case, which can be then compared and classified into different categories.

Usually the pattern is invisible such as the oil, gold and groundwater resources, and sub-surface geological features but based on the sampling and thereafter the data treatment with objective methodologies as well as personal expertise and intuition, the general spatial features may be depicted in the form of maps.

The worth of data in earth sciences and geology is very high since most of the interpretations and decisions are based on their qualitative and quantitative information contents. This information is hidden in representative field samples, which are analyzed for extraction of numerical or descriptive characteristics, which are referred to as data. Data collection in earth sciences is difficult and expensive and requires special care for accurately representing the geological phenomenon. After all, various parameters necessary for the description and modeling of the geological event, such as bearing capacity, fracture frequency, aperture, orientation, effective strength, porosity, hydraulic conductivity, chemical contents, etc., are hidden within each sample, but they represent a specific point in the space and time. Hence, it is possible to attach with data temporal and three spatial reference systems as shown in Fig. 4.1.

In geological sciences and applications, the concerned phenomenon can be examined and assessed through the collection of field data, and accordingly meaningful solutions can be proposed. It is, therefore, necessary to make the best use of available data from different points. Geological data are collected either directly in the field or field samples are transferred to laboratories in order to make necessary analyses and measurements. For instance, in hydrogeology domain, among the field measurements are the groundwater table elevations, pH, and total dissolved solution (TDS) readings, whereas some of the laboratory measurements are chemical elements in parts per million (ppm). There are also office calculations that yield hydrogeological data such as hydraulic conductivity, transmissivity, and storage coefficients. In the mean time, many other data sources from soil surveys, topographic measurements, geological prospection, remote sensing evaluations, and others may also support data for further investigation. The common property of these measurements and calculations of laboratory analysis is that they include uncertainty attached at a particular time and

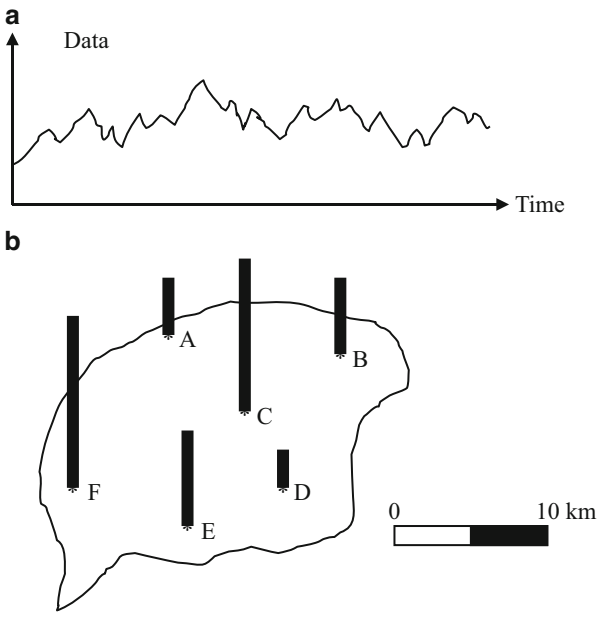


Fig. 4.1 Data variation by (**a**) time, (**b**) space

sampling point. Hence, the first question is how to deal with rather uncertainly (randomly) varying data. At times, the data is random, sometimes chaotic, and still in other cases irregular or very regular. These changes can be categorized into two broad classes as systematic and unsystematic. Systematic data yields mathematically depictable variations with time, space, or both. For instance, as the depth increases, so does the temperature, and this is an example for systematic variation. Especially, if there is only one type of geological formation, then systematic variation becomes more pronounced. Otherwise, on the basis of rather systematic variation on the average, there are unsystematic deviations, which might be irregular or random. Systematic and unsystematic data components are shown in Fig. 4.2.

Explainable deterministic components are due to systematic natural geography, astronomy and climatology factors, but unsystematic unexplainable variations have random parts that need special treatment.

It is obvious that the spatial patterns or arrangements of individual sampling locations in any study area have different patterns. The problem is to visualize qualitatively and quantitatively the change of a ReV over these points and then over the area. Appreciation of some differences between the set records of concerned variable at a set of measurement stations is the first step prior to any formal model application. It helps in the final modeling stage if the variation feature of the ReV variable is deduced first with simple but effective methods. In order to achieve such a preliminary work, the following steps become helpful:

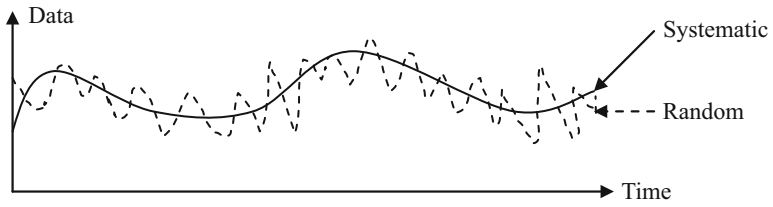


Fig. 4.2 Systematic and unsystematic components

1. Data Collection: Since the sampling locations are considered on a plane, their locations can be described by abscissas, X, and ordinates, Y. This procedure is referred to finding the easting and northing from the longitudes and latitudes. Although it is not possible to meet always in the practical studies, it is preferable that the spatial distribution of sampling sites should be rather evenly distributed over the study area and if possible constitutes points of a regular grid. The first descriptive exposition of sampling sites may be their location distribution, but without its relationship to the ReV, it does not mean much in spatial analysis. The sampling locations may already have been fixed for the researchers such as the existing well locations (water or oil), meteorology stations, urban areas, etc. If the distribution of a variable Z with respect to easting and northing is given as in Table 4.1, then its frequency distribution can be calculated for various classes and presented in Fig. 4.3.

   It is obvious from this figure that as the number of classes increases, the categories change both in class limits and in frequency values.

   On the other hand, it is possible to see the scatter of these data as in Fig. 4.4. If the ReV is denoted in general as Z, then the points in this scatter diagram can be grouped depending on some criterion as follows with respective symbols.

   These symbols help to make two types of interpretations, namely, interclass and smaller or less than a given class limits two mutually exclusive but collectively exhaustive groups. According to the attached symbols, one can deduce the following interpretations:

   (a) The majority of small ReV are gathered at high easting but low northing regions. Two exceptional clusters of the same type appear at low easting and northing regions and also in high easting but medium and high northing regions. In short, there are three clusters of small ReV. Each cluster has its special dependence feature different than others. The frequency of low ReV occurrences falls within this type of data.
   (b) The ReV variability with values between 189 and 536 takes place also at high easting and low northing region with the most concentrated cluster of ReV less than 189. In fact these two groups can be assumed as one cluster in larger scale as ReV values less than 536.
   (c) The ReV values that lie between 536 and 936 are scattered almost all over the sampling area but they do not fall within the major sampling area where easting is high and northing is low. This group of sampling points has the maximum ReV variability over the region in an independent manner. In

Table 4.1 Easting, northing, and ReV

Easting	Northing	ReV (m)	Easting	Northing	ReV (m)
40.78	30.42	31	41.25	29.02	30
40.52	30.3	100	41.73	27.23	232
39.72	40.05	1631	41.10	29.06	114
37.00	35.33	20	40.60	43.08	1775
41.17	29.04	130	41.40	27.35	46
40.73	31.60	742	37.93	41.95	896
39.62	27.92	120	41.18	29.62	31
40.18	29.07	100	42.03	35.17	32
40.32	27.97	58	40.98	27.48	4
40.15	29.98	539	40.65	29.27	4
38.40	42.12	1578	39.83	34.82	1298
40.13	26.40	3	38.50	43.5	1671
41.17	27.8	183	38.75	30.53	1034
40.55	34.97	776	38.78	35.48	1053
37.88	40.18	677	37.75	30.55	997
39.58	28.63	639	37.97	32.55	1032
39.60	27.02	21	37.20	28.35	646
39.92	41.27	1869	41.28	36.33	4
41.67	26.57	51	36.70	30.73	50
39.78	30.57	789	40.92	38.4	38
40.98	28.80	36	41.37	33.77	799
40.20	25.90	72	39.75	37.02	1285
40.97	29.08	33	39.73	39.5	1156
40.93	26.40	10	41.17	29.05	56
40.78	29.93	76	39.93	44.03	858
40.90	29.18	28	38.43	38.08	862

other words, although ReV less than 189 have high regional variability, but at least they are fathered in three cluster.

(d) Those ReV values between 536 and 1310 show a single cluster within the medium easting and northing region without intermixing with other classes. They show a directional extension along the northeast-southwest trend in Fig. 4.4.

(e) The high values of ReV more than 1310 are gathered at medium easting but high northing region of the sampling area.

(f) If the symbols are followed from the smallest to moderate and then to the biggest ReV values in Fig. 4.4, it is clear that there is an increasing trend from low northing sampling locations toward high northing direction. In the middle northing region along the easting direction, there is a hill with low values at low and high easting regions.

The above interpretations of the ReV data scatter based on the sampling locations yield clues that become very useful in actual modeling scheme. So far only

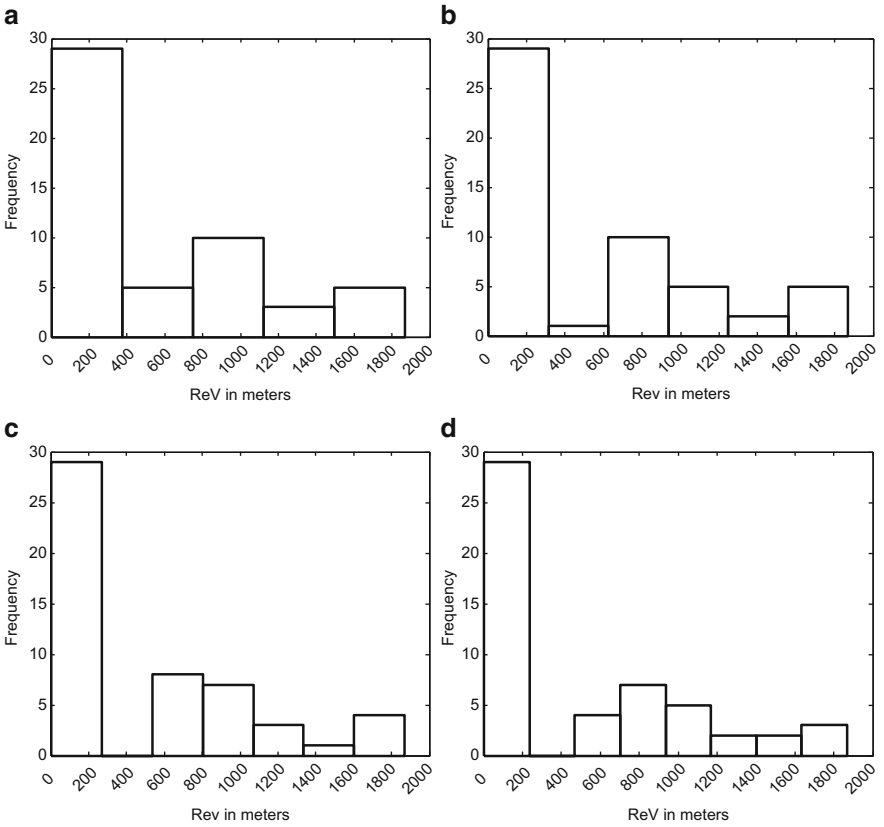


Fig. 4.3 Relative frequency diagram of ReV with (**a**) 5 class; (**b**) 6 class; (**c**) 7 class; (**d**) 8 class

linguistic interpretations are deduced, and they are the basic ingredients or rules in fuzzy logic system modeling (Zadeh 1968; Şen 2004).

It is important to notice from the qualitative information that distinctive features are depicted by considering the differences or dissimilarities between the scatter points and attached ReV values at each point. The first difference implies the distances between the sampling points and the second is the difference between the ReV values at two sites. Hence, in any quantitative study, these two differences should be taken into consideration. These differences are considered independently from each other but there may be regional dependence between them. For instance, visualization of a trend in step f implies such a regional dependence. As will be explained in Chap. 5, the relationship between the distance of two sampling points and the difference in their ReV values can be depicted by semivariogram (SV) concept.

2. Univariate Data Description: It is necessary and helpful to explore the uni-, bi-, and multivariate statistical properties of the ReV data irrespective of sampling point locations. The second step in geostatistics (or, for that matter, any statistical

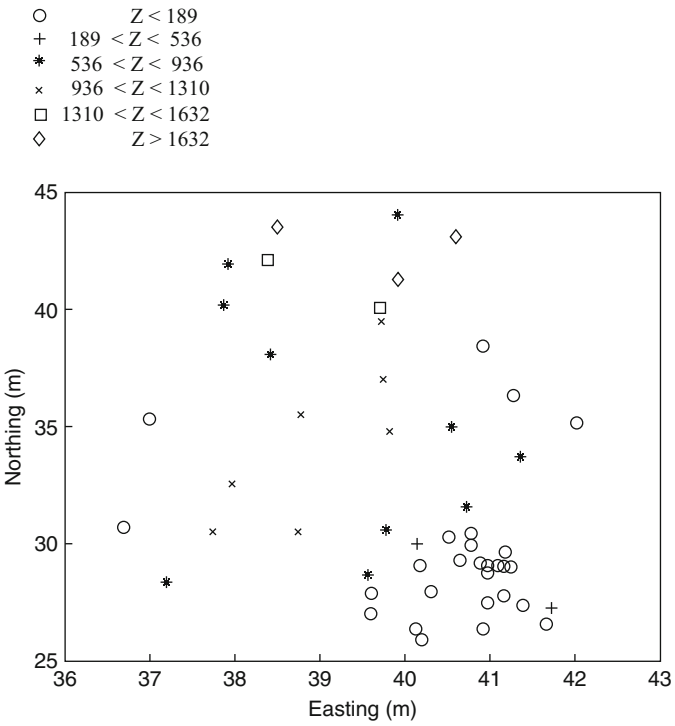


Fig. 4.4 Spatial pattern of each class

analysis) is to describe the data using univariate (and if possible, bivariate or multivariate) descriptive statistics. Among the univariate descriptive statistics are the arithmetic mean, median, mode, standard deviation, skewness, kurtosis, etc.

These parameters can be obtained either from the frequency diagram (as in Fig. 3.3) or through the classical mathematical expression of the parameters (Koch and Link 1971). Often used parameters are presented in Table 4.2 for the ReV data in Table 4.1.

In spatial analysis, one should note that there is only one univariate parameter value for any given ReV variable. As will be explained later in more detail, the univariate parameter values do not change with direction in the sampling area. Whatever the direction, although, the projection of sampling point sequence and distances will change but the univariate statistical parameters will remain the same. This explains why the univariate parameters cannot be useful in the spatial dependence search directly. The spatial variation can be measured by considering comparatively the properties of at least two ReV values at two distinctive sampling locations.

3. Bivariate Data Description: This can be achieved by considering at least two sampling points simultaneously by comparing their ReV values. It can be achieved simply either by considering a direction or without any direction. In order to familiarize the reader with each one of these approaches, six ReV values

Table 4.2 Univariate statistical parameters

Parameter	Value
Minimum	3
Mode	4
Median	125
Arithmetic mean	506.05
Maximum	1869
Standard deviation	571.11
Skewness	0.89
Kurtosis	0.41

Table 4.3 Representative ReV values

Easting (m)	Northing (m)	ReV (m)
40.78	30.42	31
40.52	30.3	100
41.17	29.04	130
39.62	27.92	120
40.18	29.07	100
40.32	27.97	58
	Parameters	Value
	Minimum	31
	Mode	100
	Median	100
	Arithmetic average	89.83333
	Maximum	130
	Standard deviation	37.96007
	Skewness	37.96007
	Kurtosis	0.363526

with their easting and northing values are considered from Table 4.1 and presented in Table 4.3. The scatter of sampling points in this case is given in Fig. 4.5.

On the basis of the data in this table, the spatial relationship between the distance and ReV values can be thought in three categories, namely, punctual, directional, and global assessments.

In the directional case, all the sampling points are projected onto desired number of directions, and the ReV values are considered not at the sampling points but at the projection points. In this manner, 2D sampling point scatter is transformed into 1D samplings as shown in Fig. 4.6.

A close inspection of these figures indicates that the sequence of sampling points and the distances between two successive points differ from each case. This is the first indication that depending on the sequential properties, the spatial variations show different appearances. The directions correspond to cross sections of 3D ReV along these directions. In order to confirm this situation, Fig. 4.7 presents the variation of ReV values with distance along the direction. It is obvious that although

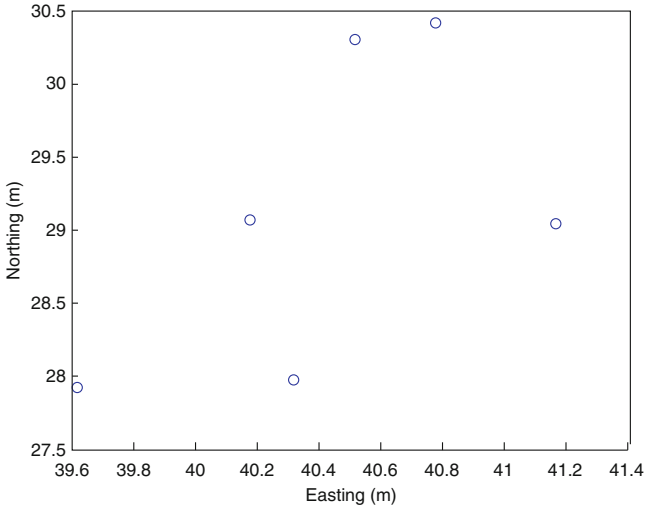


Fig. 4.5 ReV sampling locations

the univariate statistical parameters remain as they are in Table 4.3 whatever the direction, they expose different patterns.

Comparison of these three cross sections shows different patterns. Especially, in section C-C, points 4 and 5 fall on the same directional point, and in this case the representative ReV value is taken as the average of the two values. The directional graphs show the change of ReV with distance, but the distances among the points are irregular. Had it been that they are regular, then the classical time series analysis could be applied for their statistical analysis. In practice, it is time-consuming to rely on such directions, and the question is whether there is an easy statistical way to inherent the common features of all the directions and then to express the regional variability by a mathematical formulation, which helps to make ready calculations provided that the spatial data of ReV is available. This can be achieved by SV or covariogram analysis (Chap. 5).

4.4 Simple Uniformity Test

In the case of sampling point scatter as in Fig. 3.4, it is necessary to decide whether the points in each subquadrangle (subarea) arranged in a manner have more or less the same uniformity. In the case of uniformity, there is no superiority among subareas, and each subarea has the same likelihood of sampling point occurrences. If there are n sampling points over the whole area and the number of subareas is k, then the expected average uniform sampling number, U_E, for each subarea is

$$U_E = \frac{n}{k} \tag{4.1}$$

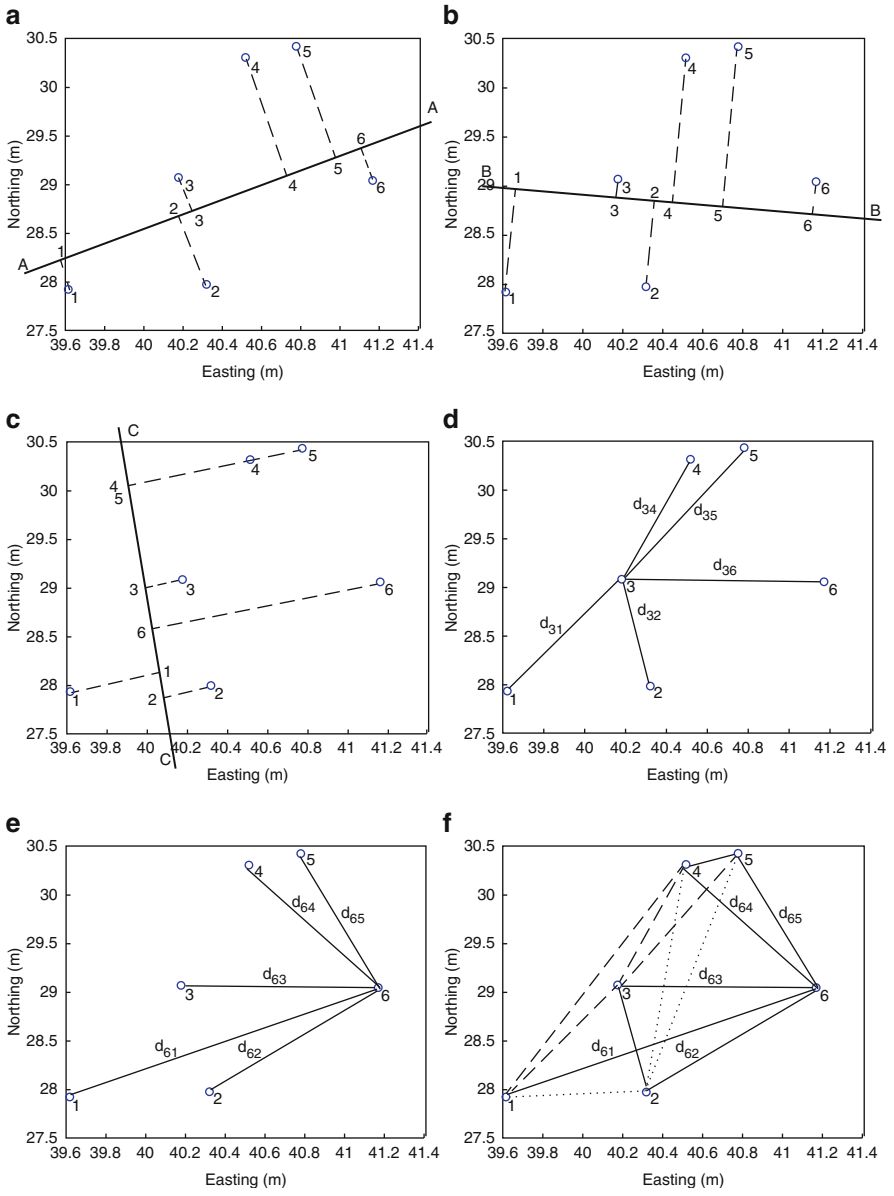


Fig. 4.6 (**a**) A-A direction, (**b**) B-B direction, (**c**) C-C direction, (**d**) partial directional (point 3) configuration, (**e**) partial directional (point 6) configuration, (**f**) unidirectional (global) configuration

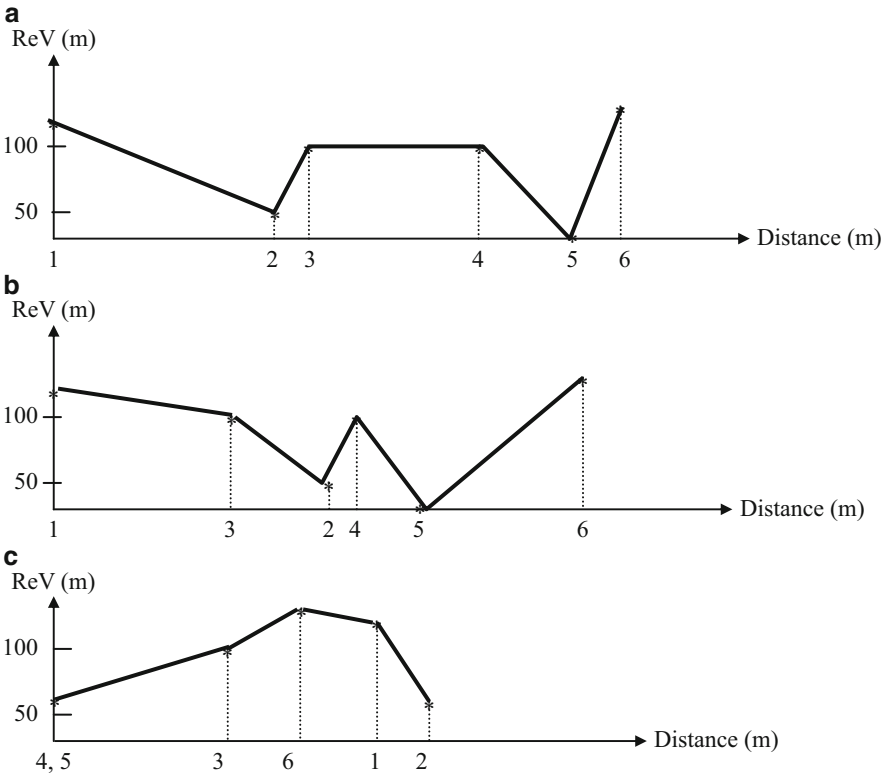


Fig. 4.7 Directional variations (**a**) A-A direction, (**b**) B-B direction, (**c**) C-C direction

This indicates the average number of sampling points per subarea. However, the actual sampling point count, C_i, ($i = 1, 2, \ldots, k$) in ith subarea is different from each other and in general from the uniform sampling number. It is possible to check the uniformity by the chi-square test as

$$\chi^2 = \sum_{i=1}^{k} \frac{(C_i - U_E)^2}{U_E} \tag{4.2}$$

The chi-square distribution has $\nu = k - 2^\circ$ of freedom, and one can find from the chi-square distribution tables in any statistical text book (Benjamin and Cornell 1970) the critical chi-square value, χ^2_{cr}, that corresponds to this degree of freedom. If $\chi^2 \leq \chi^2_{cr}$, then the distribution of the points in each subarea over the whole area is uniform.

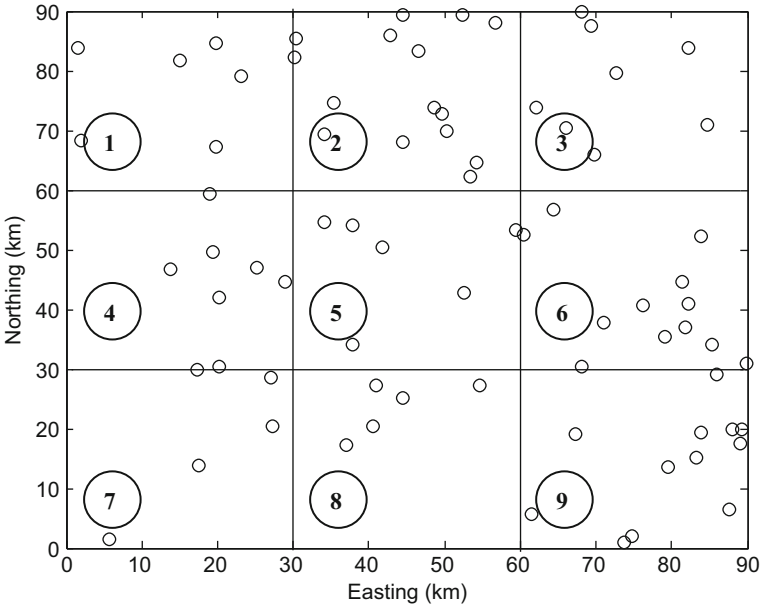


Fig. 4.8 Earthquake station location scatter

Example 4.1 The positions of many earthquake measurement stations are given in Fig. 4.8, and it has been divided into nine equal subareas. The question is whether the spatial scatter of station locations is uniformly distributed in the area or not.

Table 4.4 indicates the number of stations in each subarea where the station locations on the boundary are considered as belonging to the right and upper subareas.

According to Eq. 4.1, expected average uniform number of stations in each subarea is $75/9 = 8.33$, which can be taken as a round number equal to 8. The application of Eq. 4.4 leads to $\chi^2 = 14$. Since the degree of freedom is $\nu = 9-2 = 7$, the critical χ^2_{cr} value at 5 % significance level appears as 14.2 (i.e., $\chi^2 \leq \chi^2_{cr}$), and hence the distribution of the points is almost uniform.

4.5 Random Field

If the distribution of sampling points in the study area is not uniform, then one starts to suspect whether they constitute a random field or not. The random field is the collection of many randomly scattered points in an area where there is no spatial correlation between the points. If the points are completely random without any spatial (regional) dependence, then the only way of their spatial treatment is the probability axioms.

Table 4.4 Subareal stations

Subareal no.	Station number
1	7
2	15
3	8
4	7
5	6
6	12
7	4
8	5
9	12
Total	75

Let us consider the whole study area, A, with m sampling locations. If they are random, then the probability of one sampling point occurrence can be expressed as a percentage

$$\lambda = \frac{m}{A} \tag{4.3}$$

which lies between 0 and 1 exclusive. Now the total area is considered in terms of very small subareas where the number of subareas, n, is much larger than the number of sampling point, m, ($n >> m$). This means that m/n goes to zero as the number of subareas increases. Each one of the n subareas can be considered as a pixel (see Fig. 2.2), which is very small, compared to total area. This makes it possible to consider that each pixel is almost equal to the influence area of one sampling point. It is, therefore, impossible to have two sampling points within each pixel. Hence, the area of influence for each sampling point can be calculated as

$$a = \frac{A}{n} \tag{4.4}$$

which is the pixel area. Since each sampling point has the occurrence probability of λ per area, then in the influence area, the probability of sampling occurrence becomes

$$p_o = \lambda a = \lambda \frac{A}{n} \tag{4.5}$$

Accordingly, the probability of nonoccurrence of the sampling point in a pixel can be obtained simply as

$$p_n = 1 - p_0 = 1 - \lambda \frac{A}{n} \tag{4.6}$$

If the question is to find k sampling sites in n pixels, then in the remaining $n-k$ pixels, there will not be any sampling point. Since the spatial sampling point occurrences and nonoccurrences are independent from each other, the production operation of the probability theory gives the probability of k sampling point occurrence according to Binomial PDF as

$$p_k = \binom{n}{k} (p_o)^k (p_n)^{n-k} = \binom{n}{k} \left(\lambda \frac{A}{n}\right)^k \left(1 - \lambda \frac{A}{n}\right)^{n-k} \tag{4.7}$$

After this line, the remaining is the application of pure mathematical principles, and as n goes to very large numbers (mathematically positive infinity), then Eq. 4.7 takes its new shape as

$$p_k = e^{-\lambda A} \frac{(\lambda A)^k}{k!} \tag{4.8}$$

which has the name of Poisson distribution in common statistics. This equation requires only the rate of sampling points, λ, (Eq. 4.3); sampling point number, k; and the area of the concerned region. The combined value λA indicates the mean number of stations per quadrant. All these values are practically observable or calculateable, and, therefore, it is possible to calculate p_k from Eq. 4.8. If for a given confidence interval (90 % or 95 %) the critical probability value, p_{cr} is found from the Poisson distribution tables in any textbook on statistics (Benjamin and Cornell 1970) and if $p_k < p_{cr}$ then the sampling points are randomly distributed in the area.

However, in earth sciences the distribution of sampling points is exactly neither random nor uniform.

Example 4.2 It is now time to search for the spatial random character of the same example given in the previous section. As shown in Fig. 4.9, the whole area is divided into $9 \times 9 = 81$ quadrants.

The mean number of stations in each quadrant is $\lambda A = 75/81 = 0.926$. The first column in Table 4.5 shows the number of quadrants with 1, 2, 3, etc. stations. The probability of occurrence calculations is given in the second column.

According to Eq. 4.2, one can calculate that $\chi^2 = 1.064$. Since there are five categories, the degree of freedom is $\nu = 5-2 = 3$. The critical value of χ^2 for $\nu = 3$ and the significance level of 5 % is 7.81. The test statistics is less than this critical value, and so the hypothesis of randomness is acceptable.

The mean number of stations per quadrant λA and its variance can be estimated as

$$s^2 = \frac{\sum_{i=1}^{T} (r_1 - m/T)^2}{T-1} \tag{4.9}$$

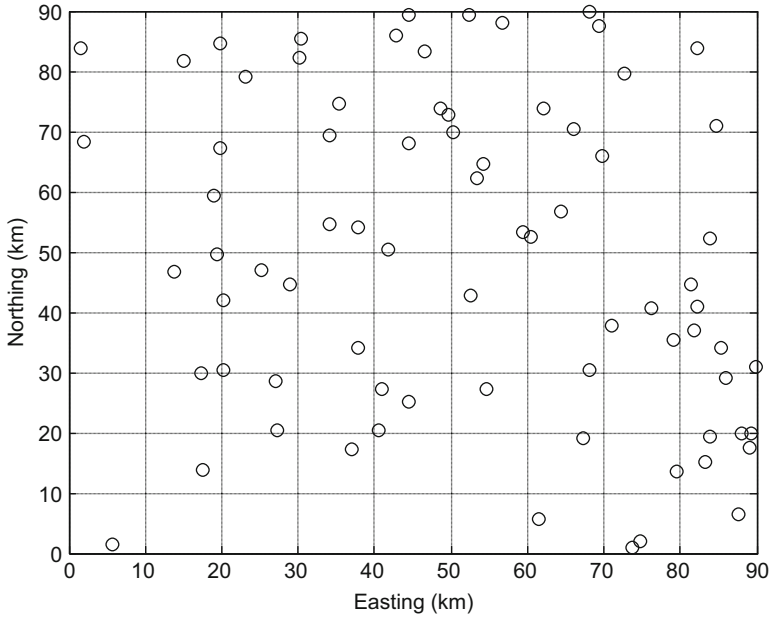


Fig. 4.9 Quadrants and station locations

Table 4.5 Applied and theoretical numbers of quadrants

Number of stations	Eq. 3.8 calculations	Number of quadrants	
		Theoretical	Actual
0	0.3961	32	31
1	0.3668	30	29
2	0.1698	14	14
3	0.0524	4	6
4	0.0121	1	1
Total	0.9972	81	81

where r_i is the number of stations per quadrant and T is the number of quadrants. It is well known that in Poisson distribution the arithmetic average is equal to variance, and by using this rule, it is possible to make further interpretations. For instance, if the arithmetic average is greater (smaller) than the variance, the scatter of stations is more uniform (clustered) than random. If the two parameters are equal to each other, the scatter of stations accords with a complete random behavior. However, at this stage it must be kept in mind that some sampling differences occur between these two parameters, and in practice it is not possible to have them equal in an exact manner.

4.6 Cluster Sampling

Due to stratification of many earth sciences events, the sampling procedure also shows grouping pattern in different non-overlapping clusters (classes) as in Fig. 3.3. In the modeling of cluster sampling, there are two stages. First is the modeling of cluster centers and then the second is the modeling of sampling points within each cluster relative to its center. The first is inter-cluster modeling and the second in within cluster modeling. As stated in many publications, the negative exponential distribution can be used to model the occurrence of clustered points in space in a manner equivalent to the use of the Poisson model to represent random field points. The cluster point distribution is modeled according to the Poisson distribution and the points in each cluster by the logarithmic distribution. Hence, the negative exponential distribution function provides the occurrence of k sampling point probability as (Davis 2002)

$$P_k = \binom{n+k-1}{k}\left(\frac{p_o}{1+p_o}\right)^k\left(\frac{1}{1+p_o}\right)^n \tag{4.10}$$

where k is the number of sampling points, p_o is the probability of event occurrence at a sampling site, and n is the degree of clustering of the occurrences. If n is large the clustering is less spelled and the sampling point distribution is close to random field distribution of the points. However, as n approaches to zero, the clustering pattern becomes more pronounced over the area. The density of sampling points is

$$\lambda = np_o \tag{4.11}$$

The following mathematical approximation is considered for solving Eq. 4.10.

$$P_0 = \frac{1}{(1+p_o)^n}, \tag{4.12}$$

and the following recursive formula

$$P_k = \frac{(n+k-1)\left(\frac{1}{p_o}+p_o\right)}{k}P_{k-1} \tag{4.13}$$

helps to calculate the subsequent probabilities. The clustering parameter can be calculated as

$$n = \frac{\left(\frac{m}{T}\right)^2}{s^2 - \frac{m}{T}} \tag{4.14}$$

where s^2 is the variance in the number of occurrences per tract, which is defined as

Table 4.6 Negative exponential distribution expected number of quadrants with r stations

Number of stations	Probability	Number of quadrants	
		Theoretical	Actual
1	0.4140	33.5	31
2	0.3593	29.1	29
3	0.1616	13.1	14
4	0.0501	4.0	6
5	0.0120	1.0	1
Total	1.0020	81.0	81

$$\lambda A = \frac{m}{T}$$

or

$$T = \frac{m}{\lambda A} \tag{4.15}$$

Likewise, the probability of occurrence can be calculated as

$$p_0 = \frac{\lambda}{n} = \frac{\frac{m}{T}}{n}$$

Example 4.3 The same example given in the previous section can be adopted for the application of cluster test. It is already calculated that $\lambda A = 0.926$. The variance, s^2, in the number of occurrences per quadrangle (subarea) is 0.897. With these values at hand, the clustering effect from Eq. 4.14 becomes $n = 27.74$. Hence, from Eq. 4.12 $p_0 = 0.0323$, and it is possible to calculate from Eq. 4.13 the probability that a given quadrant will have 1, 2, 3, etc. stations as shown in Table 4.6.

In order to test the theoretical values with the actual correspondences, it is necessary to apply chi-square test which yields $\chi^2 = 1.249$. Since there are five categories, the degree of freedom is $\nu = 5–2 = 3$. The critical value of χ^2 for $\nu = 3$ and the significance level of 5 % is 7.81. The test statistics is less than this critical value, and so the hypothesis of randomness is acceptable.

4.7 Nearest Neighbor Analysis

Rather than considering quadrangles in the subdivision of the total area into subareas and then to make calculations on the basis of the sampling points within the subarea, it is preferable to consider the neighbor points next to each other. Logically, if there are n sampling points within a total area, A, then the size of each equal subarea, a, is simply calculated according to Eq. 4.4. This area is supposed to

include two nearest points on the average. It is considered as a square, and therefore the side length, L, of the square is the square root of the area

$$L = \sqrt{a} \tag{4.16}$$

Hence, the mean distance between two points is

$$\overline{d}_T = \frac{1}{2}L \tag{4.17}$$

The variance of the average distance between two points can be calculated rationally as

$$\sigma_T^2 = \frac{(4-\pi)}{4\pi n^2}A \tag{4.18}$$

In the derivations of Eqs. 4.17 and 4.18, the area is assumed without boundary, i.e., very extensive. However, this is not the situation in practice, and, therefore, these statistical parameters are without areal extent restrictions, and hence they provide underestimations. If the constants are worked and the standard error of estimate is calculated, the result becomes

$$s_{Te} = \frac{0.26136}{\sqrt{\frac{a}{n^2}}} \tag{4.19}$$

If the number of sampling points is more than six, then the distribution of the average distance between nearest neighbors comply with a normal PDF. The mean and the variance of this PDF are given by Eqs. 4.17 and 4.19, respectively. Hence, for decision the standardized normal PDF value, x, can be obtained as

$$x = \frac{\overline{d} - \overline{d_T}}{s_{Te}} \tag{4.20}$$

As mentioned earlier, the theoretical Eqs. 4.17 and 4.18 are in underestimations, and, therefore, a correction factor less than one must be imported. Many researchers suggested different adjustments but the one given by Donnelly (1978) found frequent use. According to his extensive numerical simulation studies, the theoretical mean distance and its variance can be expressed as (Davis 2002)

$$\overline{d_T} \approx \frac{1}{2}\sqrt{\frac{A}{n} + \left(0.514 + \frac{0.412}{\sqrt{n}}\right)\frac{P}{n}} \tag{4.21}$$

and

$$\sigma_T^2 \approx 0.0070\frac{A}{n^2} + 0.035\mathrm{P}\frac{\sqrt{A}}{n^{2.5}} \tag{4.22}$$

where P is the perimeter of the regular map. The ratio, R_d, of the expected and observed mean nearest-neighbor distances can be used to indicate the spatial pattern as

$$R_d = \frac{\overline{d}}{\overline{d_T}} \tag{4.23}$$

This ratio approaches to zero where all the sampling points are very close to each other with almost negligible average distance. Another extreme point appears as R_d approaches one, where the random field scatter of sampling points takes place. When the mean distance to the nearest neighbor is maximized, then R_d takes its maximum value as 2.15. Figure 4.10 indicates the sampling point distribution with different distance ratio indices.

The distance ratio indices for Fig. 4.8a–f are equal to 2.15, 1.95, 1.20, 0.89, 0.31, and 0.11, respectively.

4.8 Search Algorithms

One of the critical differences between various local-fit algorithms for interpolating to a regular grid is the way in which "nearest neighbors" are defined and found. Some may depend on the number of nearest points and others adopt a radius of influence and depend on the points within this area of influence (Chap. 2). Some search techniques (such as the optimum interpolation technique) may be superior to others in certain situations especially concerning the data point arrangement. There are a variety of search procedures that may be selected by the user. The simplest method finds the n nearest neighboring data points, in a Euclidean distance sense, regardless of their irregular distribution around the estimation node, which is fast and satisfactory if observations are distributed in a comparatively uniform manner, but provides poor estimates if the data are closely spaced along widely separated traverses.

One of the objections to a simple nearest-neighbor search is that all the nearest points may lie, for instance, in a narrow wedge on one side of the grid node (Fig. 4.11). Hence, the estimation is essentially unconstrained, except in one direction. It is possible to avoid such a situation by restricting the search in some way, which ensures that the points are equitably distributed about the estimation point.

The simplest method introduces a measure of radial search as a quadrant. Some minimum number of points must be taken from each of the four quadrants around the estimation point. An elaboration on the quadrant search is an octant search,

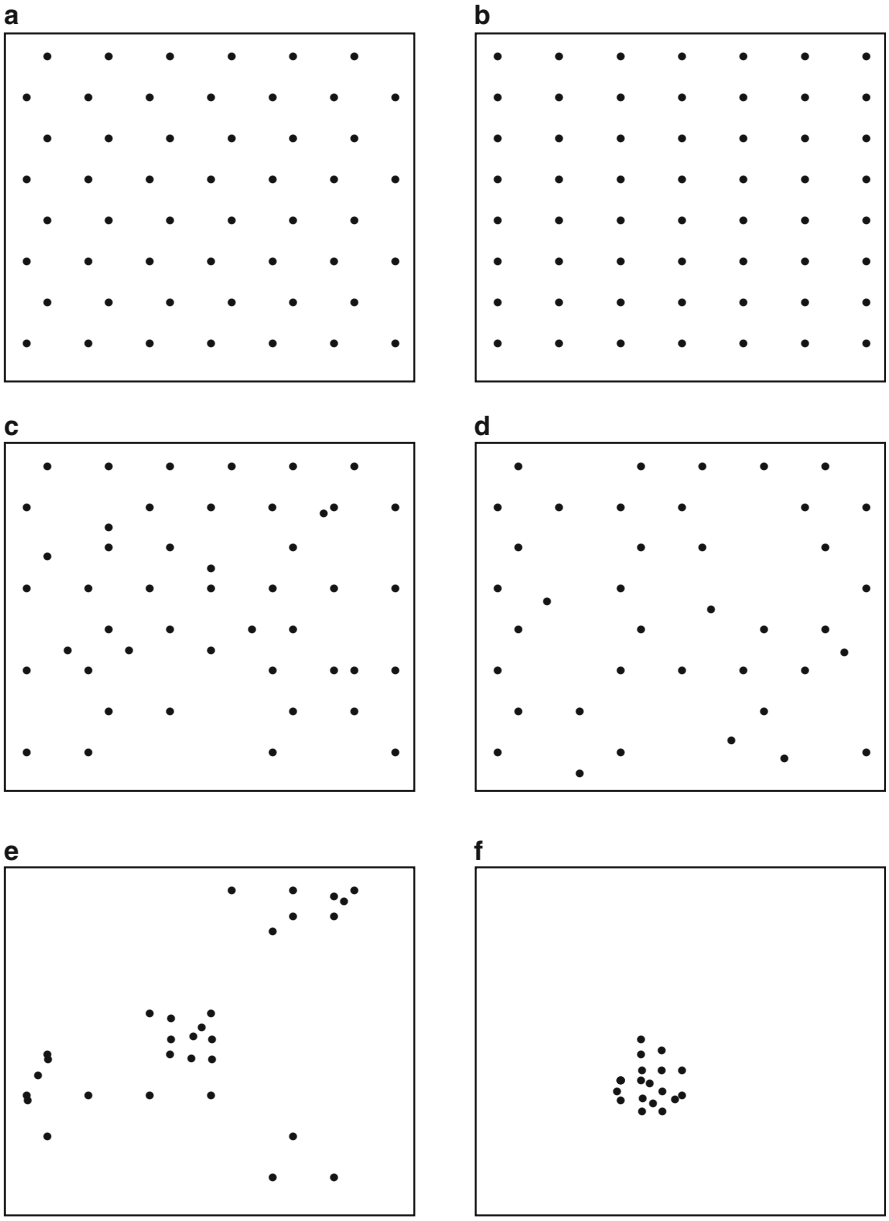


Fig. 4.10 Nearest-neighbor ratio statistics

which introduces a further constraint on the radial distribution of the points used in the estimation procedure. A specified number of control points must be found in each of the 45° segments surrounding the estimation point. These constrained search procedures require finding and testing more neighboring points than in a

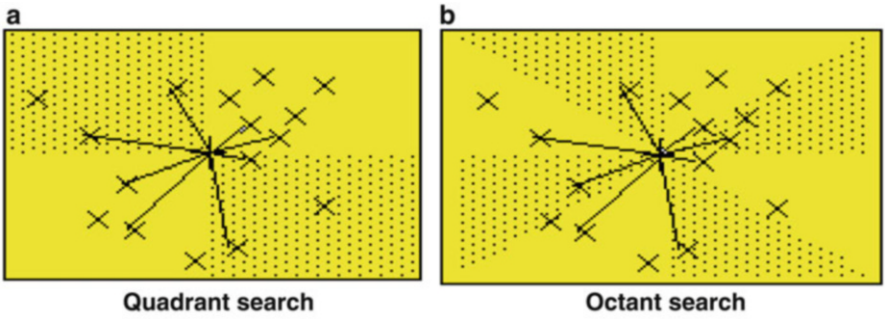


Fig. 4.11 Regular directional nearest point systems, (**a**) quadrant, (**b**) octant

simple search, which increases the time consumption. It is also possible to consider all points within a radius R of the estimation point. Estimates made using the other search procedures are based on a fixed number of points collected at variable distances from the grid node. This search algorithm uses a variable number of points within a fixed distance of the estimation point.

Any constraints on the search for the nearest points, such as a quadrant or octant requirement, will obviously expand the size of the neighborhood around the estimation point. This occurs because some nearby control points are likely to be passed over in favor of more distant points in order to satisfy the requirement that only a few points may be taken from a single sector. Unfortunately, the autocorrelation of a surface decreases with increasing distance, so more remote points are less closely related to the estimation point. This means the estimate may be poorer than if a simple nearest-neighbor search procedure is used.

4.8.1 Geometric Weighting Functions

Operational objective analysis procedures are suggested for interpolating in a practical way information from unevenly distributed observations to a uniformly distributed set of grid points.

The earliest studies were started by Gilchrist and Cressman (1954) who reduced the domain of polynomial fitting to small areas surrounding each node with a parabola. Bergthhorsson and Döös (1955) proposed the basis of successive correction methods which did not rely only on interpretation to obtain grid point values but also a preliminary guess field is initially specified at the grid points. Cressman (1959) developed a number of further correction versions based on reported data falling within a specified distance R from each grid point. The value of R is decreased with successive scans, and the resulting field of the latest scan is taken as the new approximation. Barnes (1964) summarized the development of a convergent weighted-averaging analysis scheme which can be used to obtain any desired amount of detail in the analysis of a set of randomly spaced data. The

scheme is based on the supposition that the 2D distribution of an atmospheric variable can be represented by the summation of an infinite number of independent waves, i.e., Fourier integral representation. A comparison of existing objective methods up to 1979 for sparse data is provided by Goodin et al. (1979). Their study indicated that fitting a second-degree polynomial to each subregion triangular in the plane with each data point weighted according to its distance from the subregion provides a compromise between accuracy and computational cost. Koch et al. (1983) presented an extension of the Barnes method which is designed for an interactive computer scheme. Such a scheme allows real-time assessment both of the quality of the resulting analyses and of the impact of satellite-derived data upon various meteorological data sets. Both Cressman and Barnes methods include power parameter and radius of influence values which are rather subjectively determined in the current meteorological practice.

In any optimum analysis technique, the main idea is that the estimation at any point is considered as a weighted average of the measured values at irregular sites. Hence, if there are $i = 1, 2, \ldots, n$ measurement sites with records Z_i then the estimated site solar irradiation, Z_C, can be calculated according to Eq. 2.27, which is most commonly used in different disciplines because of its explicit expression as the weighted average. Weighting functions proposed by Cressman (1959), Gandin (1963), and Barnes (1964) also appear as sole functions of the distances between the sites only. Unfortunately, none of the weighting functions are event dependent, but they are suggested on the basis of the logical and geometrical conceptualizations of site configuration. Various geometrical weighting functions are shown in Fig. 4.12.

The weighting functions that are prepared on a rational and logical basis without consideration of regional data have the following major drawbacks:

(a) They do not take into consideration the natural variability of the regional variability features. For instance, in meteorology, Cressman (1959) weightings are given as

$$W(r_{i,E}) = \begin{cases} \dfrac{R^2 - r_{i,E}^2}{R^2 + r_{i,E}^2} & \text{for} \quad r_{i,E} \le R \\ 0 & \text{for} \quad r_{i,E} \ge R \end{cases} \tag{4.24}$$

where $W(r_{i,E})$ corresponds to W_i in Eq. 2.27, $d_{i,E}$ is the distance between estimation point and other points, and R is the radius of influence, which is determined subjectively by personal experience.

(b) Although weighting functions are considered universally applicable all over the world, they may show unreliable variability for small areas. For instance, within the same study area, neighboring sites may have quite different weighting functions.

(c) Geometric weighting functions cannot reflect the morphology, i.e., the regional variability of the phenomenon. They can only be considered as practical first approximation tools.

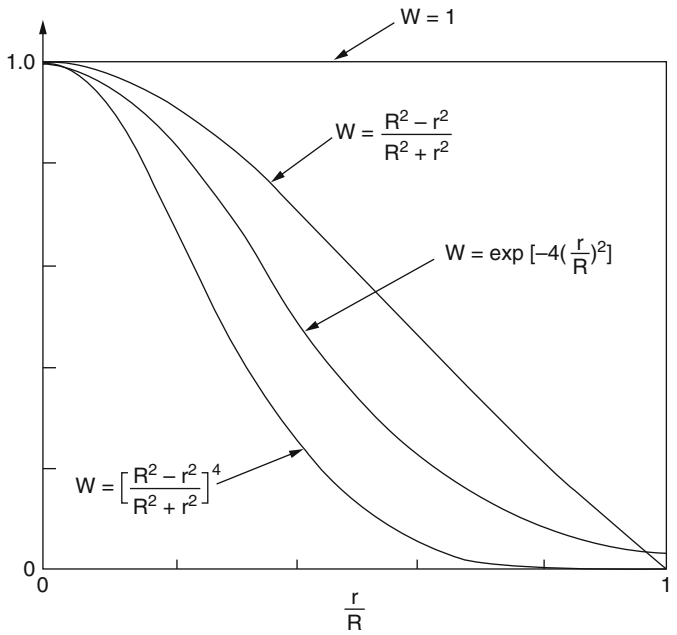


Fig. 4.12 Weighting functions

A generalized form of the Cressman model with an extra exponent parameter α is suggested as

$$W(r_{i,E}) = \begin{cases} \left(\frac{R^2 - r_{i,E}^2}{R^2 + r_{i,E}^2}\right)^{\alpha} & \text{for } r_{i,E} \leq R \\ 0 & \text{for } r_{i,E} \geq R \end{cases} \tag{4.25}$$

The inclusion of α has alleviated the aforesaid drawbacks to some extent but its determination still presents difficulties in practical applications. Another form of geometrical weighting function was proposed by Sasaki (1960) and Barnes (1964) as

$$W(r_{i,E}) = exp\left[-4\left(\frac{r_{i,E}}{R}\right)^2\right] \tag{4.26}$$

In reality, it is expected that weighting functions should reflect the spatial dependence behavior of the phenomenon. To this end, regional covariance and SV functions are among the early alternatives for the weighting functions that take into account the spatial correlation of the phenomenon considered. The former method requires a set of assumptions such as the Gaussian distribution of the

regionalized variable. The latter technique, semivariogram (SV), does not always yield a clear pattern of regional correlation structure (Şen 1989).

4.9 Trend Surface Analysis

The main purpose of trend surface analysis is to summarize data that are numerous to grasp readily by eye from the list of numbers. A spatial trend is a systematic change usually in a geographical (longitude and latitude) direction in the ReV. The trend surface analysis is also called a "global interpolator" because it uses all data. Most interpolators (called "local interpolators") use only a sample of points to estimate the value at each grid node location. It fits the entire surface with a polynomial equation by the ordinary least squares method. Once trend surfaces were the only basic tool as maps for communication in earth sciences. In trend analysis, the statistical model coefficients are estimated, and the equation is presented as a contour map. Trend surface presentation in the form of contour maps is the same as the preparation of topographic maps, where the spatial dependence is not considered explicitly. The mathematical model of the trend surface is ready in the form of polynomial equation which relates the geographic variables x and y to the ReV, z, as

$$z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \tag{4.27}$$

where a_i's are the model coefficients to be estimated from available data recorded at a set of geographical points within the study area. Although there are many statistical models, this model is appropriate for a smooth concave and convex surface. The locations of local maximum and minimum can be obtained from Eq. 4.27 by taking partial derivative with respect to x and y, which yields

$$\frac{\partial z}{\partial x} = a_1 + 2a_3 x + a_4 y$$

and

$$\frac{\partial z}{\partial y} = a_2 + a_4 y + 2a_5 y$$

When these derivatives are set equal to zero, the solution for the extreme points becomes

$$x_e = \frac{a_2 a_4 - 2a_1 a_5}{4a_3 a_5 - a_4^2}$$

and

$$y_e = \frac{a_1 a_4 - 2a_2 a_3}{4a_3 a_5 - a_4^2}$$

Different interpretations can be made of an extreme case depending on the analysis. Among the possibilities are that the extreme point may be a structural high, low, or inflection point.

In any trend surface analysis, the only limit on the number of coefficients is that one cannot have more than the number of sample points. It is good for approximating the general form of the surface. Estimates of individual values are not very good (including at sample points). The interpolated surface will always have the same value as the sample value where there are sample points and assigns each point to be interpolated that same value as the nearest sample point. It gives a stairstep shape to the surface discontinuities.

Trend surface analysis is a popular numerical technique in trying to explore the 3D features of any phenomena based on the geographical coordinates, but due to a set of restrictive assumptions, they cannot be regarded as successful at the time. Hence, it is frequently misused, and consequently the conclusions based on such uses may also be in great error. Majority of problems arise from the measurement site distribution, lack of real data fit, extensive computational requirements, and at times inappropriate applications. In general, trend surface fitting methodologies are multivariate statistical model fitting to a set of regional measurement data with restrictive assumptions. A uniform density of data points is important in many types of analysis, including trend surface methods.

Trend analysis separates the ReV into two complementary components, namely, regional nature of deterministic variations and local fluctuations around the regional component. The regional and local components are dependent on the scale of the ReV. In any trend analysis, there are three variables:

1. The basic variables are the geographical coordinates, which give the exact locations of the sampling points in the study area. For trend analysis the longitude and latitude are converted to easting and northing values with respect to a common reference point. In trend analysis the choice of the reference point does not make any difference in further calculations.
2. The trend component is the regional representation of the event rather deterministic manner, which has the form

$$\hat{z} = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \tag{4.28}$$

   where x and y are some function or combination of the geographical coordinates of the sample locations. This expression yields the trend component as $\hat{z}_i$ at the ith location of the ReV. Although there are nonlinear terms, this equation is called linear because the terms are added together.
3. The trend coefficients are chosen such that the squared deviations of measurements from the trend surface are minimized. The sum of the squares from the

trend surface is equivalent to the variance of the ReV. The multiple linear regressions are also defined similarly, and therefore the trend analysis is similar to the multivariate analysis in the classical statistics.

4.9.1 Trend Model Parameter Estimations

It is possible to obtain the coefficient estimations in Eq. 4.27 through the application of regression methodology with the criterion of least squares analysis. This is a very formal way of deriving the coefficient estimation expressions in terms of the data values. However, in this book a more practical expression is obtained by successive independent variable term multiplication and average taking procedure. The procedure says the following steps.

1. Consider the main Eq. 4.27 and then take the arithmetic average of both sides which leads to

$$\overline{z} = a_0 + a_1\overline{x} + a_2\overline{y} + a_3\overline{x^2} + a_4\overline{xy} + a_5\overline{y^2} \tag{4.29}$$

   This is the first estimation expression where all the arithmetic averages can be obtained from an available data set. Since there are six coefficients as unknowns, it is necessary to obtain five more expressions.
2. Multiply both sides of the main equation by the first term variable on the right-hand side and then take the arithmetic averages. This procedure yields finally

$$\overline{zx} = a_0\overline{x} + a_1\overline{x^2} + a_2\overline{yx} + a_3\overline{x^3} + a_4\overline{x^2y} + a_5\overline{xy^2} \tag{4.30}$$

3. Apply the same procedure as in the previous step but this time by multiplying both sides by the second independent variable, i.e., y, which gives

$$\overline{zy} = a_0\overline{y} + a_1\overline{xy} + a_2\overline{y^2x} + a_3\overline{yx^2} + a_4\overline{xy^2} + a_5\overline{y^3} \tag{4.31}$$

4. Multiply both sides by x^2 and then take the arithmetic average of both sides, which leads to

$$\overline{zx^2} = a_0\overline{x^2} + a_1\overline{x^3} + a_2\overline{x^2y} + a_3\overline{x^4} + a_4\overline{x^3y} + a_5\overline{x^2y^2} \tag{4.32}$$

5. This time the independent variable is xy, and accordingly its use under the light of the aforementioned procedure results in

$$\overline{xyz} = a_0\overline{xy} + a_1\overline{x^2y} + a_2\overline{xy^2} + a_3\overline{x^3y} + a_4\overline{x^2y^2} + a_5\overline{xy^3} \tag{4.33}$$

Finally, considering the last independent term variable as y^2, the same rule yields the following expression:

$$\overline{y^2 z} = a_0\overline{y^2} + a_1\overline{y^2 x} + a_2\overline{y^3} + a_3\overline{y^2 x^2} + a_4\overline{xy^3} + a_5\overline{y^4} \tag{4.34}$$

Hence, the necessary equations are obtained for the trend surface model coefficient estimation. In matrix notation these equations can be written explicitly as

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 1 & \overline{x} & \overline{y} & \overline{x^2} & \overline{xy} & \overline{y^2} \\ \overline{x} & \overline{x^2} & \overline{xy} & \overline{x^3} & \overline{x^2 y} & \overline{xy^2} \\ \overline{y} & \overline{xy} & \overline{y^2} & \overline{x^2 y} & \overline{xy^2} & \overline{y^3} \\ \overline{x^2} & \overline{x^3} & \overline{x^2 y} & \overline{x^4} & \overline{x^3 y} & \overline{x^2 y^2} \\ \overline{xy} & \overline{x^2 y} & \overline{xy^2} & \overline{x^3 y} & \overline{x^2 y^2} & \overline{xy^3} \\ \overline{y^2} & \overline{xy^2} & \overline{y^3} & \overline{x^2 y^2} & \overline{xy^3} & \overline{y^4} \end{bmatrix}^{-1} \begin{bmatrix} \overline{z} \\ \overline{xz} \\ \overline{yz} \\ \overline{x^2 z} \\ \overline{xyz} \\ \overline{y^2 z} \end{bmatrix}$$

The right-hand side of this last expression includes various averages that are calculable from a given set of data, and, therefore, the parameter estimation is possible after the inverse of square matrix. This matrix is symmetrical around the main diagonal. It is also possible to construct the titles of calculation table by considering these averages as in Table 4.7 the following table titles.

The arithmetic average of each column gives the elements of each matrix on the right-hand side of Eqs. 4.31, 4.32, 4.33, 4.34, 4.35, 4.36, and 4.37.

Example 4.4 A set of earthquake records are given for some part of Turkey, and the magnitude is required to be related to easting (x) and northing (y) coordinates for the prediction of Richter magnitude. The first three columns of Table 4.8 give the easting, northing, and earthquake magnitude values, respectively. For the sake of argument, simple linear trend surface equation is adopted as

$$z = a + bx + cy \tag{4.35}$$

where a, b, and c are the trend surface parameters. According to the steps given above, the rest of the table is prepared accordingly.

There are three unknowns and it is necessary to obtain three equations in the light of practical trend surface calculations. By considering the last row of averages from Table 4.8, one can write the necessary equations simply as

$$\begin{aligned} 6.38 &= a + 39.12b + 33.30c \\ 249.64 &= 39.12a + 1533.62b + 1306.02c \\ 211.97 &= 33.30a + 1306.02b + 1150.11c \end{aligned}$$

The simultaneous solution of these equations yields $a = 5.63$, $b = 0.0314$, and $c = -0.0143$, and hence the final linear trend surface expression is given as

$$z = 5.56 + 0.0314x - 0.0143y$$

Table 4.7 Trend surface calculation table

x	y	z	x^2	xz	xy	yz	y^2	x^3	x^2y	x^2z	xyz	y^2z	xy^2	y^3	x^4	x^3y	x^2y^2	xy^3	y^4

Table 4.8 Earthquake data and trend calculations

Number	x	y	z	xz	x^2	xy	yz	y^2
1	40.3	38.4	5.1	205.53	1624.09	1547.52	195.84	1474.56
2	38	27	6	228	1444	1026	162	729
3	39	39	6.3	245.7	1521	1521	245.7	1521
4	40.5	42.7	6.2	251.1	1640.25	1729.35	264.74	1823.29
5	38	26.5	6	228	1444	1007	159	702.25
6	40.18	38.1	6.75	271.215	1614.432	1530.858	257.175	1451.61
7	42.5	26.4	5.9	250.75	1806.25	1122	155.76	696.96
8	41	34	6.2	254.2	1681	1394	210.8	1156
9	36	30	6.25	225	1296	1080	187.5	900
10	40.65	27.2	7.75	315.0375	1652.423	1105.68	210.8	739.84
11	40.6	27.1	6.4	259.84	1648.36	1100.26	173.44	734.41
12	40.1	26.8	6.9	276.69	1608.01	1074.68	184.92	718.24
13	38	30	6.9	262.2	1444	1140	207	900
14	40.27	36.38	7.1	285.917	1621.673	1465.023	258.298	1323.504
15	39.26	26.71	7	274.82	1541.348	1048.635	186.97	713.4241
16	35.5	34	5.8	205.9	1260.25	1207	197.2	1156
17	39.7	42.8	5.3	210.41	1576.09	1699.16	226.84	1831.84
18	39.96	41.94	6.8	271.728	1596.802	1675.922	285.192	1758.964
19	38.55	30.78	5.9	227.445	1486.103	1186.569	181.602	947.4084
20	41.33	43.41	6	247.98	1708.169	1794.135	260.46	1884.428
21	38	30.5	5.9	224.2	1444	1159	179.95	930.25
22	37.03	29.43	6.1	225.883	1371.221	1089.793	179.523	866.1249
23	35.84	29.5	7	250.88	1284.506	1057.28	206.5	870.25
24	36.54	27.33	7.3	266.742	1335.172	998.6382	199.509	746.9289
25	40.94	43.88	6	245.64	1676.084	1796.447	263.28	1925.454
26	38.1	27.1	6.5	247.65	1451.61	1032.51	176.15	734.41
27	40.5	26.5	6.1	247.05	1640.25	1073.25	161.65	702.25
28	40.2	37.9	6.1	245.22	1616.04	1523.58	231.19	1436.41
29	37.98	44.48	7.6	288.648	1442.48	1689.35	338.048	1978.47
Averages	39.12172	33.30483	6.384483	249.6336	1533.642	1306.022	211.9668	1150.113

4.10 Multisite Kalman Filter (KF) Methodology

Trend surface analysis depends on continuous and regular mathematical surfaces, which does not consider individual features in any site separately, but globally fits the model to available spatial data at fixed time. It is necessary to repeat the same procedure for any subsequent time interval independently from the previous step. It is, therefore, necessary to adopt a spatial model, which digests the spatial data at a given time and then updated the model output with the new recordings at the same locations in the region. It is possible to achieve such a task through Kalman filter (KF) procedure. KF is an optimal state estimation process applied to a dynamic system that involves random perturbations. The atmosphere can be regarded as an uncertain dynamical system (Kalman 1960; Dee 1991). More precisely, KF gives a linear, equitable, and minimum error variance recursive algorithm to optimally estimate the unknown state of a dynamic system from noisy data taken at continuous or discrete real-time intervals. The KF supplies a measure of the accuracy of an analysis in the form of its error covariance (Kalman and Bucy 1961). Therefore, it allows the impact of different observation sets to be compared.

KF is the most general approach to statistical estimation and prediction. It has been shown by Harrison and Stevens (1975) that all forecasting methods are special cases of KF. This filter can deal with changes in the model, the parameters, and the variances. The difficulty with KF is that many technical questions have not yet been answered satisfactorily. The approach itself has grown out of engineering practices. Consequently, many statisticians and operation researchers know little about it, or find it difficult to understand, because it is most often described in state-space notation. Furthermore, many practical difficulties still exist as the initial estimates for parameters, variances, covariance, and the transition matrix. As shown in Fig. 4.13, the discrete KF is an iterative procedure containing several elements which are described in the next sections.

The filter is supplied with initial information, including the prior estimate of initial parameters, which is based on all the knowledge about the process, and the error covariance associated with it, and these are used to calculate a Kalman gain. The error between the parameter estimation and the measured data is determined and multiplied by Kalman gain to update the parameter estimate and estimation error. The updated error and parameters are used as inputs to a model, in order to predict the projected error and parameters at the next time instance. The first derivation of the KF recursive equations is the "filter" equation. The equations used in the discrete KF are given in detailed by Brown and Hwang (1992). Latif (1999) has performed multisite KF development and application for Turkish precipitation in detail.

The objective of this section is to investigate and develop a KF model approach to multisite precipitation modeling and prediction in addition to the assessment of associated errors. In order to have an online prediction operation, it is desirable to be able to deal with a multitude of rainfall events. The precipitation predictor should not be fixed with time and space, but adapt itself to the evolving

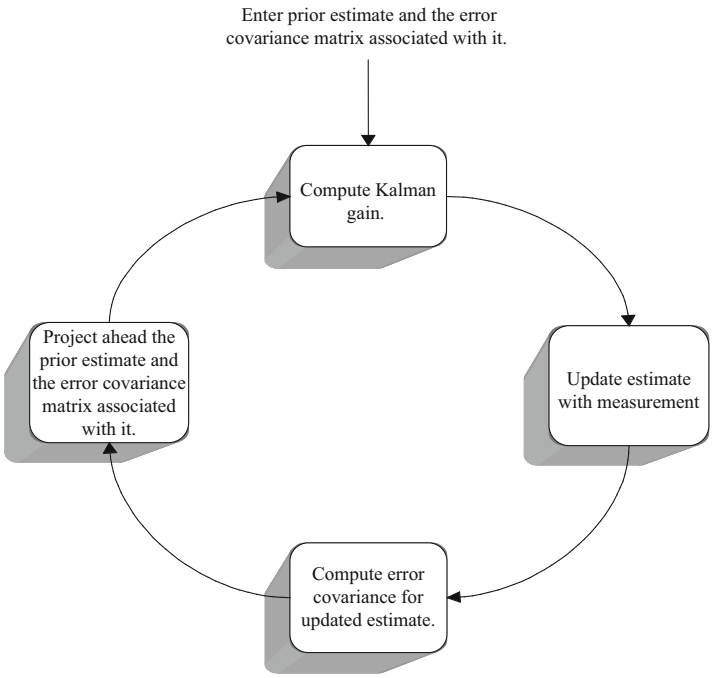


Fig. 4.13 KF iterative procedure

meteorological conditions. Any stochastic model is associated with various uncertainties. KF consists of combining two independent estimates to form a weighted estimate or prediction. One estimate can be a prior prediction or an estimate based on prior knowledge and the other a prediction based on new information (new data). The purpose of the KF is to combine these two pieces of information to obtain an improved estimate.

4.10.1 1D KF

Suppose that there is a random variable, X_k, whose values should be estimated at a set of certain times t_0, t_1, t_2, …, t_n and that X_{k-1} satisfies the following dynamic system equation:

$$X_k = \varphi X_{k-1} + W_{k-1} \tag{4.36}$$

In this expression, φ is a known parameter relating X_k to X_{k-1}, and W_{k-1} is a pure independent random number with zero mean $\overline{W}_{k-1} = 0$ and the variance equals to Q. Let the measurement, Z_k, at time k is given as

$$Z_k = H_k X_k + V_k \tag{4.37}$$

where V_k is another white noisy with zero mean and variance, R and H, is measurement parameter.

In order to improve the prior estimate $\hat{\mathrm{X}}_{k/k-1}$, the noisy measurement at time k, Z_k is used as

$$\hat{\mathrm{X}}_{k/k} = \hat{\mathrm{X}}_{k/k-1} + K_k\left(Z_k - H_k\hat{\mathrm{X}}_{k/k-1}\right) \tag{4.38}$$

where $\hat{\mathrm{X}}_{k/k}$ is updated estimate and K_k is the Kalman gain. Notice that $\left(Z_k - H_k\hat{\mathrm{X}}_{k/k-1}\right)$ is just the error in estimating Z_k. For deciding on the value of K, the variance of the error be computed as

$$\begin{aligned} E\left[\left(X_k - \hat{\mathrm{X}}_{k/k}\right)^2\right] &= E\left[\left(X_k - \hat{\mathrm{X}}_{k/k-1} - K_k\left(Z_k - H_k\hat{\mathrm{X}}_{k/k-1}\right)\right)^2\right] \\ &= E\left[\left(1 - K_kH_k\right)\left(X_k - \hat{\mathrm{X}}_{k/k-1}\right) + K_kV_k\right]^2 \\ &= P_{k/k-1}\left(1 - K_kH_k\right)^2 + R\,K_k^2 \end{aligned} \tag{4.39}$$

where the cross product terms drop out because V_k is assumed to be uncorrelated with X_k and $\hat{\mathrm{X}}_{k/k-1}$. So, the variance of updated estimation error is given by

$$P_{k/k} = P_{k/k-1}\left(1 - K_kH_k\right)^2 + R\,K_k^2 \tag{4.40}$$

If it is necessary to minimize the estimation error, then minimization of $P_{k/k}$ is required by differentiating $P_{k/k}$ with respect to K_k and setting the derivatives equal to zero. A little algebra shows that the optimal K_k is obtained as

$$K = H\,P_{k/k-1}\left[P_{k/k-1}(K_k)^2 + R\right]^{-1} \tag{4.41}$$

In these calculations, X_k is a column matrix with many components. Then Eqs. 4.35, 4.36, 4.37, 4.38, 4.39, 4.40, and 4.41 become matrix equations, and the simplicity as well as the intuitive logic of the KF becomes obscured. The covariance matrices for W_{k-1} and V_k vectors are given by

$$\begin{aligned} E\left[W_kW_i^T\right] &= \left\{\begin{array}{ll} Q_k, & i = k \\ 0, & i \neq k \end{array}\right\} \\ E\left[V_kV_i^T\right] &= \left\{\begin{array}{ll} R_k, & i = k \\ 0, & i \neq k \end{array}\right\}. \\ E\left[W_kV_i^T\right] &= 0, \qquad \text{for all k and } i\,. \end{aligned}$$

where subscript T denotes transpose of the vector.

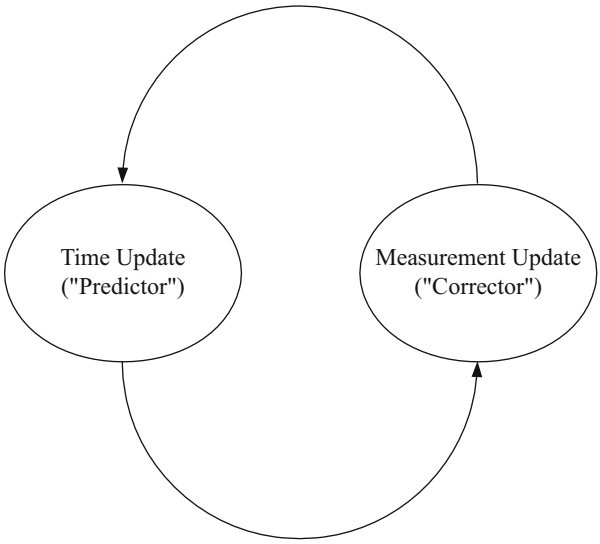


Fig. 4.14 The ongoing discrete KF cycle

These equations fall into two groups, time and measurement update equations. The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the a priori estimates for the next time step. The measurement update equations are responsible for incorporating a new measurement into the a priori estimate to obtain an improved estimate. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. Indeed the final estimation algorithm resembles that of a predictor algorithm for solving numerical problems as shown in Fig. 4.14.

The specific equations for the time updates are as follows.

$$\hat{\mathrm{X}}_{k+1/k} = \Phi_{k+1k}\hat{\mathrm{X}}_{k/k} \tag{4.42}$$

and

$$P_{k+1/k} = \Phi_{k+1k}P_{k/k}\Phi_{k+1k}^{T} + Q_k \tag{4.43}$$

On the other hand, the measurement update (corrector) equations are as follows.

$$K_k = P_{k/k-1}H_k^{T}\left(H_kP_{k/k-1}H_k^{T} + R_k\right)^{-1} \tag{4.44}$$

$$\hat{\mathrm{X}}_{k/k} = \hat{\mathrm{X}}_{k/k-1} + K_k\left(Z_k - H_k\hat{\mathrm{X}}_{k/k-1}\right) \tag{4.45}$$

and

$$P_{k/k} = P_{k/k-1} - K_kH_kP_{k/k-1} \tag{4.46}$$

It is possible to summarize KF estimation execution steps under the light of previous derivations as follows:

1. Enter prior estimate $\hat{\mathrm{X}}_{k/k-1}$ which is based on all our knowledge about the process prior to time t_{k-1} and also suggest the error covariance matrix associated with it $P_{k/k-1}$.
2. Compute the Kalman gain as $K_k = P_{k/k-1}H_k^T\left(H_kP_{k/k-1}H_k^T + R_k\right)^{-1}$.
3. Update estimate with measurement Z_k as $\hat{\mathrm{X}}_{k/k} = \hat{\mathrm{X}}_{k/k-1} + K_k\left(Z_k - H_k\hat{\mathrm{X}}_{k/k-1}\right)$.
4. Compute error covariance for updated estimate as $P_{k/k} = P_{k/k-1} - K_kH_kP_{k/k-1}$.
5. Project ahead the updated estimate $\hat{\mathrm{X}}_{k/k}$ and the error covariance matrix associated with it $P_{k/k}$, to use it as a prior estimation for the next time step $\hat{\mathrm{X}}_{k+1/k} = \Phi_{k+1k}\hat{\mathrm{X}}_{k/k}$ and finally $P_{k+1/k} = \Phi_{k+1k}P_{k/k}\Phi_{k+1k}^T + Q_k$.

Once the loop is entered as shown in Fig. 3.12, then it can be continued as much as necessary. Initially, when the model parameters are only rough estimates, the gain matrix ensures that the measurement data is highly influential in estimating the state parameters. Then, as confidence in the accuracy of the parameters grows with each iteration, the gain matrix values decrease, causing the influence of the measurement data in updating the parameters and associated error to reduce.

4.10.2 KF Application

To investigate and develop a KF model approach to multisite precipitation modeling, 30-year records (1956–1985) of annual rainfall for the 52 stations are used. As shown in Fig. 4.15, these stations are distributed over an area approximately covering all of Turkey with more concentration in the northwestern part.

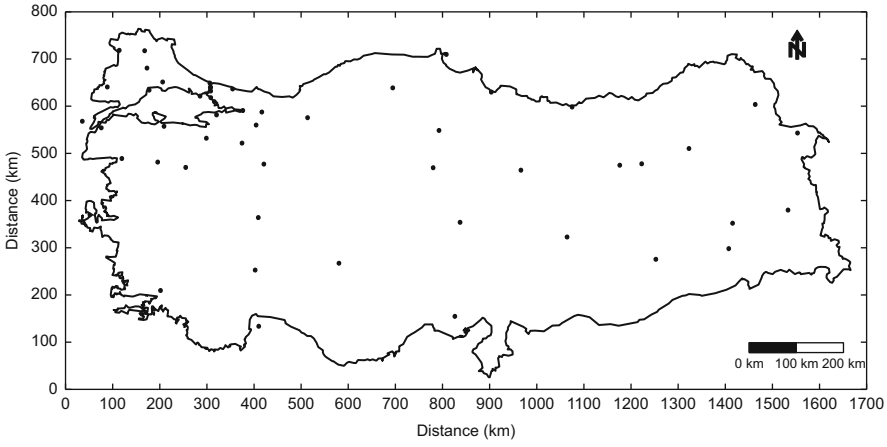


Fig. 4.15 Distribution of rainfall stations over turkey

The study area lies between latitudes 36°:70′N and 42°:03′N and longitude 25°:90′E and 44°:03′E, extended over an area of about 780,576 km^2. The geographic locations (latitude, longitude, and elevation in meter above mean sea level) of precipitation stations are represented in Table 4.9.

Precipitation is characterized by variability in space and time. In addition, there are many factors affecting the magnitude and distribution of precipitation, elevation of station above mean sea level, various air mass movement, moisture, temperature, pressure, and topography. The magnitude and distribution of precipitation vary from place to place and from time to time, even in small areas. The application of multisite KF model as developed in this section approach to multisite precipitation modeling which illustrates some interesting points in the annual precipitation pattern.

Figure 4.16 provides the observed and estimated annual rainfall values at Adapazarı from 1956 to 1984.

It is to be noticed from this figure that the observed and estimated values follow each other closely, which indicates that KF provides an efficient method for modeling of annual rainfall. Some statistical parameters of annual observed and estimated rainfall values during the time period (1956–1985) are summarized in Table 4.10.

From another point of view, Fig. 4.16 provides the observed and estimated annual rainfall values at 52 selected stations in Turkey for 1985. From Fig. 4.17 and Table 4.10, again as noticed before, in the case of one station (Adapazarı), the observed and estimated values follow each other closely, which indicates that KF provides an efficient method for modeling of annual rainfall in both time and space dimension.

Contour maps of observed and estimated annual rainfall for 1956–1985 and percentage errors of estimated annual rainfall are presented in Figs. 4.18 and 4.19, respectively. In Fig. 4.18 dashed lines indicate estimated annual average precipitation contours.

According to the areal values of observed and estimated annual rainfalls, the multisite KF method has a slight tendency toward underestimation. Standard deviation of estimated value is smaller than that of observed one (Fig. 4.20). Its mean has less variability and, therefore, more smoothed than observed values.

Furthermore, Fig. 4.17 proves that the estimated values of annual rainfall at most of the sites in the study area are close to the observed values, especially in the part where more stations are available such as in the northwestern part of Turkey. The percentage error of estimated values vary from −6 in station number 52 (underestimation) to 6 in station number 49 (overestimation) with overall average about 0.121 %.

The magnitude and distribution of precipitation vary from place to place and from time to time even in small areas. Describing and predicting the precipitation variability in space and/or time are fundamental requirements for a wide variety of human activities and water project designs. In this paper, a KF technique has been developed for the prediction of annual precipitation amounts at a multiple site. In

Table 4.9 Station locations and elevation above mean sea level

No.	Station no.	Station name	Latitude (N)	Longitude (E)	Elevation (meter)
01	068	Adapazarı	40°: 78′	30°: 42′	31
02	662	Ali Fuat Paşa – Adapazarı	40°: 52′	30°: 30′	100
03	102	Ağrı	39°: 72′	40°: 05′	1631
04	350	Adana	37°: 00′	35°:33′	20
05	620	Bahçeköy – İstanbul	41°: 17′	29°: 05′	130
06	070	Bolu	40°: 73′	31°: 60′	742
07	150	Balıkesir	39°: 62′	27°: 92′	120
08	116	Bursa	40°: 18′	29°: 07′	100
09	115	Bandırma – Balıkesir	40°: 32′	27°: 97′	58
10	122	Bilecık	40°: 15′	29°: 98′	539
11	164	Bitlis	38°: 40′	42°: 12′	1578
12	112	Çanakkale	40°: 13′	26°: 40′	3
13	054	Çorlu – Tekirdağ	41°: 17′	27°: 80′	183
14	084	Çorum	40°: 55′	34°: 97′	776
15	280	Diyarbakır	37°: 88′	40°: 18′	677
16	651	Dursunbey – Balıkesir	39°: 58′	28°: 63′	639
17	653	Edremit – Balıkesir	39°: 60′	27°: 02′	21
18	096	Erzurum	39°: 92′	41°: 27′	1869
19	050	Edirne	41°: 67′	26°: 57′	51
20	124	Eskişehir	39°: 78′	30°: 57′	789
21	058	Florya – İstanbul	40°: 98′	28°: 80′	36
22	673	Gökçeada – Çanakkale	40°: 20′	25°: 90′	72
23	062	Göztepe – İstanbul	40°: 97′	29°: 08′	33
24	674	İpsala – Edirne	40°: 93′	26°: 40′	10
25	010	İzmit	40°: 78′	29°: 93′	76
26	063	Kartal – İstanbul	40°: 90′	29°: 18′	28
27	059	Kumköy – İstanbul	41°: 25′	29°: 02′	30
28	601	Kırklareli	41°: 73′	27°: 23′	232
29	011	Kandilli	41°: 10′	29°: 06′	114
30	098	Kars	40°: 60′	43°: 08′	1775
31	052	Luleburgaz – Kirklareli	41°: 40′	27°: 35′	46
32	210	Siirt	37°: 93′	41°: 95′	896
33	020	Şile – İstanbul	41°: 18′	29°: 62′	31
34	026	Sinop	42°: 03′	35°: 17′	32
35	056	Tekirdağ	40°: 98′	27°: 48′	4
36	118	Yalova – İstanbul	40°: 65′	29°: 27′	4
37	132	Yozgat	39°: 83′	34°: 82′	1298
38	170	Van	38°: 50′	43°: 50′	1671
39	190	Afyon	38°: 75′	30°: 53′	1034
40	195	Kayseri – Erkilet	38°: 78′	35°: 48′	1053
41	240	Isparta	37°: 75′	30°: 55′	997

(continued)

Table 4.9 (continued)

No.	Station no.	Station name	Latitude (N)	Longitude (E)	Elevation (meter)
42	244	Konya	37°: 97′	32°: 55′	1032
43	292	Muğla	37°: 20′	28°: 35′	646
44	030	Samsun	41°: 28′	36°: 33′	4
45	300	Antalya	36°: 70′	30°: 73′	50
46	034	Giresun	40°: 92′	38°: 40′	38
47	074	Kastamonu	41°: 37′	33°: 77′	799
48	090	Sivas	39°: 75′	37°: 02′	1285
49	092	Erzincan	39°: 73′	39°: 50′	1156
50	061	Sarıyer – İstanbul	41°: 17′	29°: 05′	56
51	100	Iğdır	39°: 93′	44°: 03′	858
52	200	Malatya – Erhav	38°: 43′	38°: 08′	0862

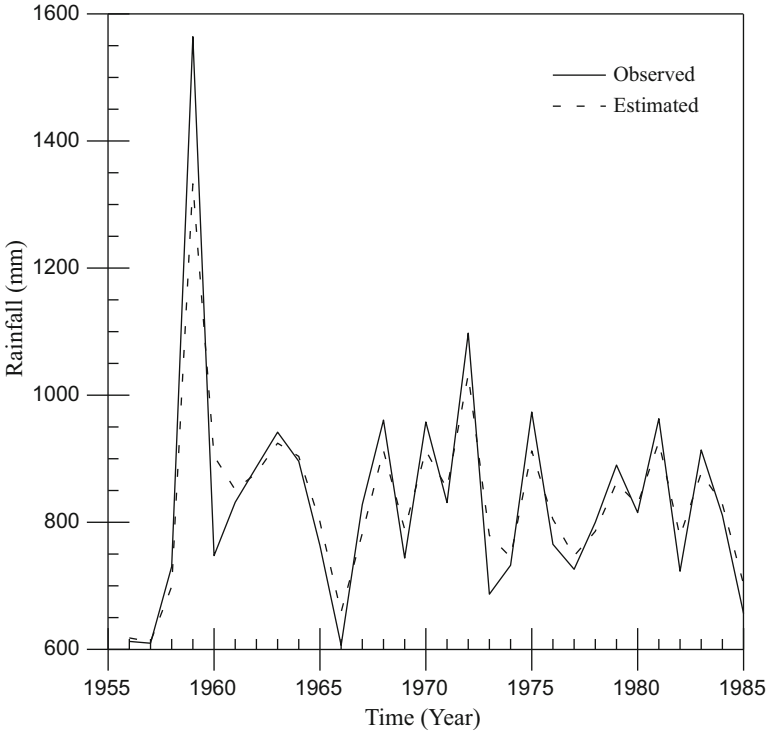


Fig. 4.16 Observed and estimated annual rainfall values at Adapazarı from 1956 to 1984

Table 4.10 Statistical parameters of observed and estimated annual rainfall values (1956–1985)

No	Station name	Mean (mm)		St. dev. (mm)		Minimum (mm)		Maximum (mm)		Range (mm)		Skewness	
		Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.
1	Adapazarı	835.5	834.6	183.4	135.3	606.3	611.7	1564	1332	957.7	720.3	2.18615	1.50665
2	Ali Fuat Paşa – Adapazarı	660.6	661	107.7	85.1	484.9	519.1	922.5	851.1	437.6	332	0.36170	0.13996
3	Ağrı	506.7	504.7	91.9	66.6	352	389.1	693.1	612.7	341.1	223.6	0.13271	−0.0978
4	Adana	690.4	688.7	221.3	184.4	319.4	360	1265	1149	945.6	789	0.86862	0.66023
5	Bahçeköy – İstanbul	1287	1284	478.6	354.7	848.1	866.2	2698	2325	1850	1459	1.73873	1.56783
6	Bolu	549.2	549.2	83.3	59.3	377.4	429.1	716.8	659.7	339.4	230.6	0.14283	0.16372
7	Balıkesir	606.7	606.1	163.6	126.3	362.6	385.6	1192	1036	829.4	650.4	1.59256	1.25458
8	Bursa	680.6	680.1	109.1	82.7	447.8	521.1	871	835	423.2	313.9	0.16262	−0.0062
9	Bandırma – Balıkesir	719	718.9	149.3	115	456.2	494.7	1086	988.2	629.8	493.5	0.71735	0.55033
10	Bilecik	445.1	444.5	75.4	57.5	320.6	332.8	624.7	563.1	304.1	230.3	0.62528	0.39746
11	Bitlis	1164	1160	287.5	235.8	564.5	734.2	1866	1738	1301	1004	0.27270	0.51462
12	Çanakkale	636.6	640.4	139.9	108.9	414.8	459.8	977.7	868.5	562.9	408.7	0.64583	0.45769
13	Çorlu – Tekirdağ	570.6	571.1	104.1	81.5	431.2	456.4	801.3	759.9	370.1	303.5	0.58876	0.64617
14	Çorum	432.1	430.7	79.7	59.1	285.5	296.8	606.7	542.9	321.2	246.1	−0.05457	−0.2570
15	Diyarbakır	490	489	134.5	99.3	146.3	276	748.8	671	602.5	395	−0.10185	0.24572
16	Dursunbey – Balıkesir	779.1	782.2	171.6	130.6	479.3	575	1199	1079	719.7	504	0.36409	0.15470
17	Edremit – Balikesir	696.2	697.2	144.1	111.6	512.9	548.4	1175	1061	662.1	512.6	1.36238	1.2798
18	Erzurum	410.1	409.8	85.3	61.3	291.1	308.1	638.5	551.8	347.4	243.7	0.92707	0.49036
19	Edirne	593.2	594.4	102.2	72.1	430.5	496.6	863.6	792.2	433.1	295.6	0.85735	0.87714
20	Eskişehir	393.1	391.8	94.8	51.7	215.8	220.1	524	486.8	326.2	266.7	−0.18598	−0.9474
21	Florya – Istanbul	656.4	655.5	117.3	89.3	500.8	540	1026	942.6	525.2	402.6	1.02088	1.07883
22	Gökçeada – Çanakkale	792.5	796.4	212.2	162.8	483	543.8	1451	1235	968	691.2	1.28517	0.92635
23	Göztepe – Istanbul	698.7	696.7	127.4	98.3	538.8	559	1047	990.9	508.2	431.9	0.87054	0.94386
24	İpsala – Edirne	612.5	616.3	98.7	77.5	386.5	434.7	808.9	763.9	422.4	329.2	0.15228	0.11054
25	İzmit	766.6	764.9	150.2	119.9	554.9	557.2	1088	1012	533.1	454.8	0.37969	0.12170
26	Kartal – İstanbul	651.1	649.5	118.5	90.9	475.6	496.9	871.9	832.4	396.3	335.5	0.16123	0.05385
27	Kumköy – İstanbul	796.1	793.2	185.8	143.7	519.5	581.9	1278	1145	758.5	563.1	1.02989	0.88983

(continued)

Table 4.10 (continued)

No	Station name	Mean (mm)		St. dev. (mm)		Minimum (mm)		Maximum (mm)		Range (mm)		Skewness	
		Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.	Obs.	Est.
28	Kirklareli	589.4	592.5	133.9	97.5	371.5	417.5	1000	876.9	628.5	459.4	1.17951	0.99290
29	Kandilli	827.7	825.5	148.7	117.5	600.5	631.5	1230	1143	629.5	511.5	0.54675	0.42092
30	Kars	470.5	472.2	106.1	78.2	298.5	361.7	718.8	661.1	420.3	299.4	0.75373	0.81958
31	Luleburgaz – Kırklareli	652.3	655.4	182.2	138.8	399.7	437.4	1360	1159	960.3	721.6	2.00788	1.57797
32	Siirt	684.5	682.4	183.4	141.5	420.8	474.5	1229	1060	808.2	585.5	1.02095	0.91065
33	Şile – İstanbul	800.8	796.2	251.5	211.6	454.6	491.5	1697	1537	1242	1045	1.74568	1.66792
34	Sinop	640.2	639.7	134.2	109.6	414.7	455.2	990.4	917.1	575.7	461.9	0.83072	0.77820
35	Tekirdağ	604.2	606.6	192.6	141.2	405.2	434.8	1464	1192	1059	757.2	3.1807	2.50102
36	Yalova – İstanbul	770.4	769.2	266.9	202.7	473.2	484.7	1959	1617	1486	1132	3.216	2.60607
37	Yozgat	565.8	562.6	109.2	86	391	412.9	858.2	771.3	467.2	358.4	0.72179	0.44662
38	Van	377.9	377.9	62.7	47.6	267.9	292.5	486	461.1	218.1	168.6	−0.1035	0.04482
39	Afyon	408.6	407.7	89.5	69	239	259.3	618	576.6	379	317.3	0.14329	−0.0192
40	Kayseri – Erkilet	364.8	365.2	60	45.3	263	278.6	535	475.3	272	196.7	0.75670	0.2949
41	Isparta	557.6	557.8	160	122.7	332	371.4	968	891.5	636	520.1	0.70164	0.85167
42	Konya	326.1	325	77.6	63.7	193	225	545	483.7	352	258.7	0.95732	0.66484
43	Muğla	1165	1164	298.3	229.3	658	767.4	1805	1672	1147	904.6	0.53785	0.44423
44	Samsun	693.7	694.2	129.4	103.2	442.2	503.5	1011	955	568.8	451.5	0.45975	0.72549
45	Antalya	1074	1072	293.4	228.3	554	608.7	1914	1747	1360	1138	0.29616	0.28473
46	Giresun	1240	1244	153.2	125.5	1057	1090	1679	1668	622	578	1.17704	1.63857
47	Kastamonu	457.6	455	82.2	67.3	308	312.1	617	574.9	309	262.8	0.17054	−0.1840
48	Sivas	411.8	410.1	79.2	56.4	286	298.8	575	519.8	289	221	0.03905	0.09282
49	Erzincan	368.8	367.6	71.7	56.8	257	277.9	563	504.1	306	226.2	0.86580	0.62331
50	Sarıyer – İstanbul	796.6	794.3	150.8	130.6	574.9	597.8	1171	1102	596.1	504.2	0.91685	0.72847
51	Iğdır	251.9	253.2	77.3	55.5	114.5	154.9	501.2	422.9	386.7	268	1.0145	1.00863
52	Malatya – Erhav	402.7	403.1	100	72.1	248	292.1	593	533.9	345	241.8	0.42067	0.33824

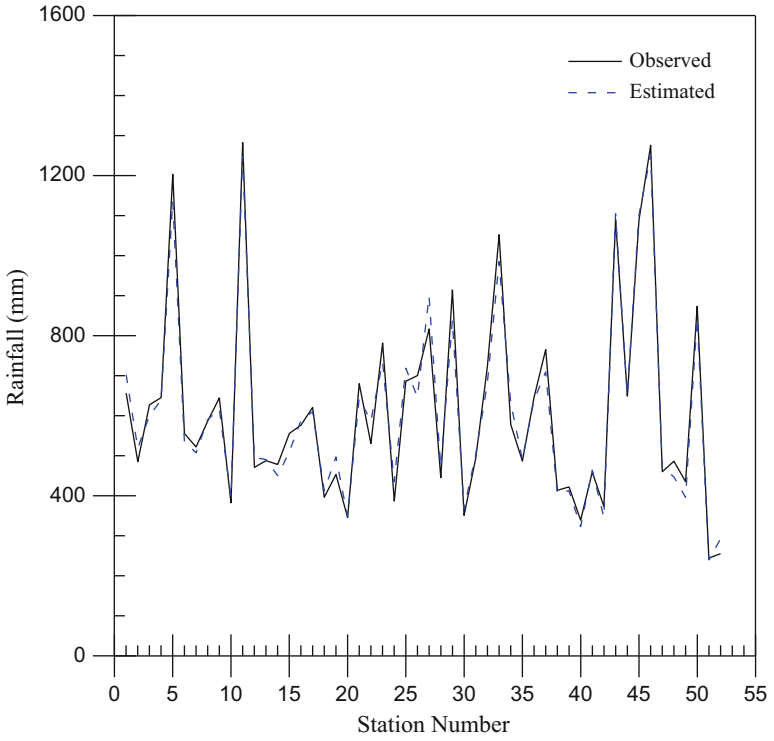


Fig. 4.17 Observed and estimated annual rainfall values at selected stations in Turkey for 1985

this meaner the precipitation amount at any year and stations is predicted by considering all the available station interrelationships.

Once a model has been selected, then KF processing requires specification of initial state vector, error covariance matrix associated with this initial state vector, system noise covariance, measurement noise covariance, state transition matrix, and connection matrix. Most of this information should be based on physical understanding and all the previous knowledge about the process prior to t_{k-1}. If little historical information is available to specify the above matrices, then KF may be started with very little objective information and adapted as data becomes available. However, the less the initial information, greater diagonal elements should be selected in the covariance matrices. In this manner, the algorithm will have flexibility to adjust itself to sensible values in a relatively short space of time.

The average amount of rainfall values at the selected stations are used as the elements of initial state vector. Sufficiently great diagonal elements of error covariance matrix are needed with initial state vector provided in the initial moment. Then the prediction error covariance steadily decreases with time and arrives at a stable value after some steps, indicating the efficiency of the prediction algorithm.

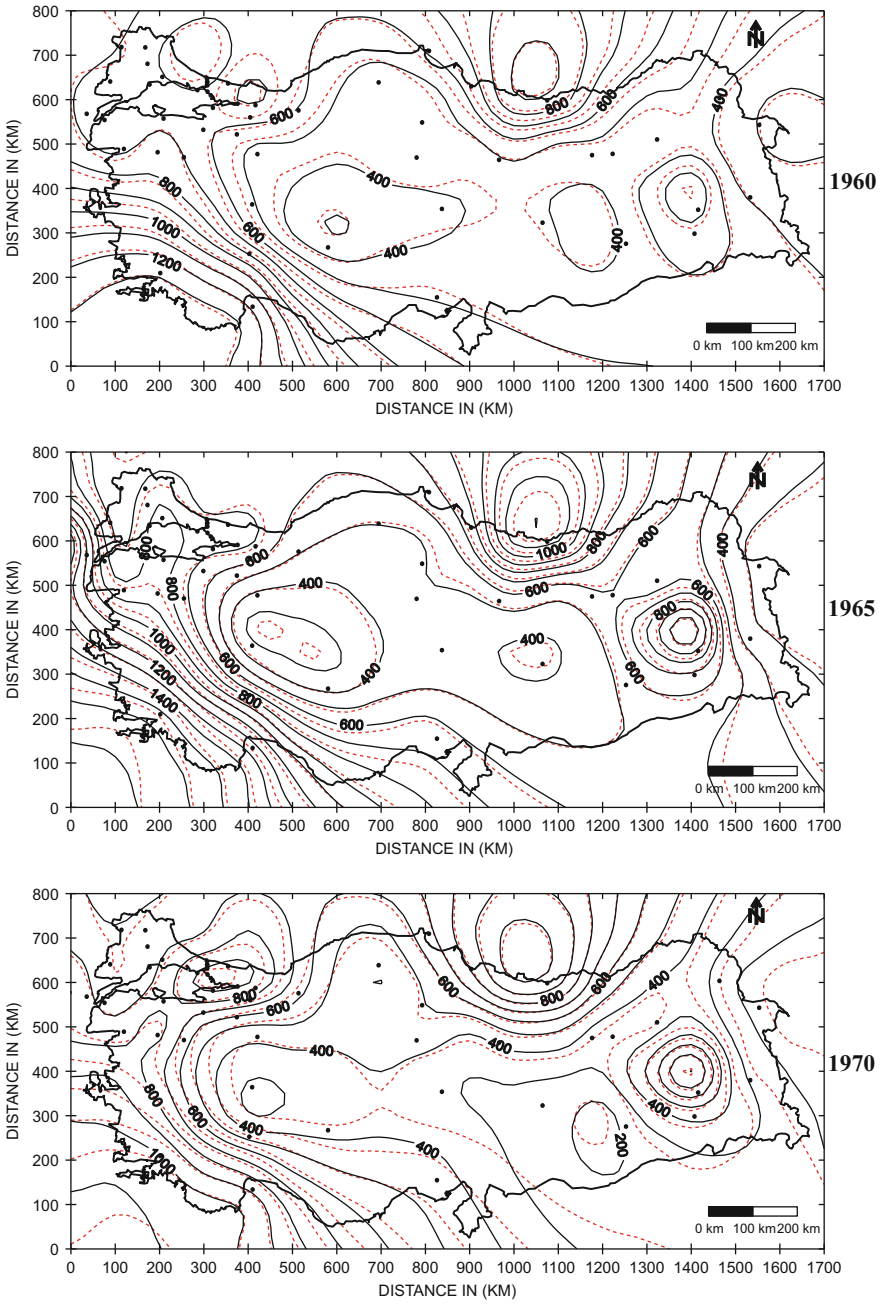


Fig. 4.18 Contour map of observed and estimated annual precipitation 1960–1985

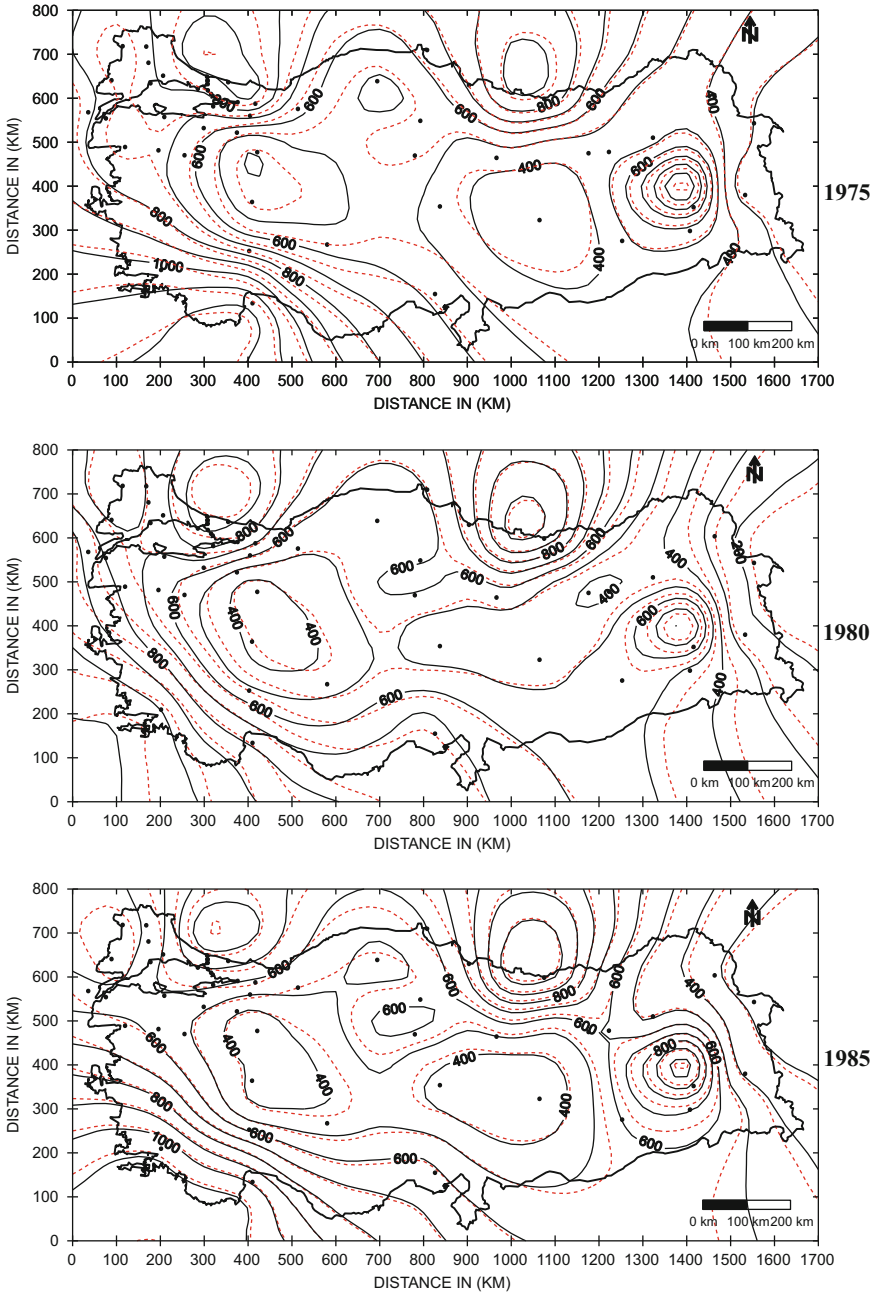


Fig. 4.18 (continued)

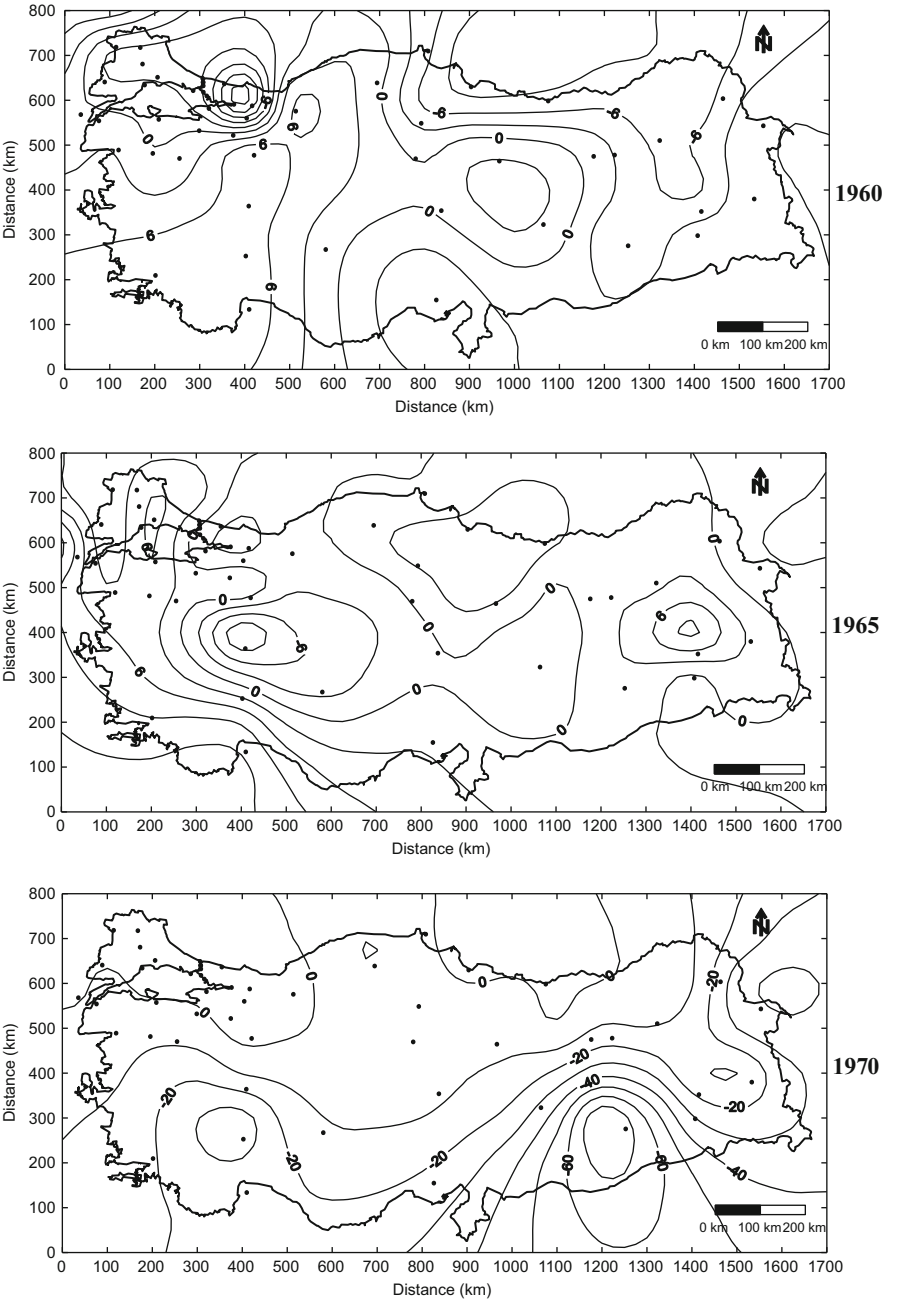


Fig. 4.19 Contour map of percentage error of estimated annual rainfall 1960–1985

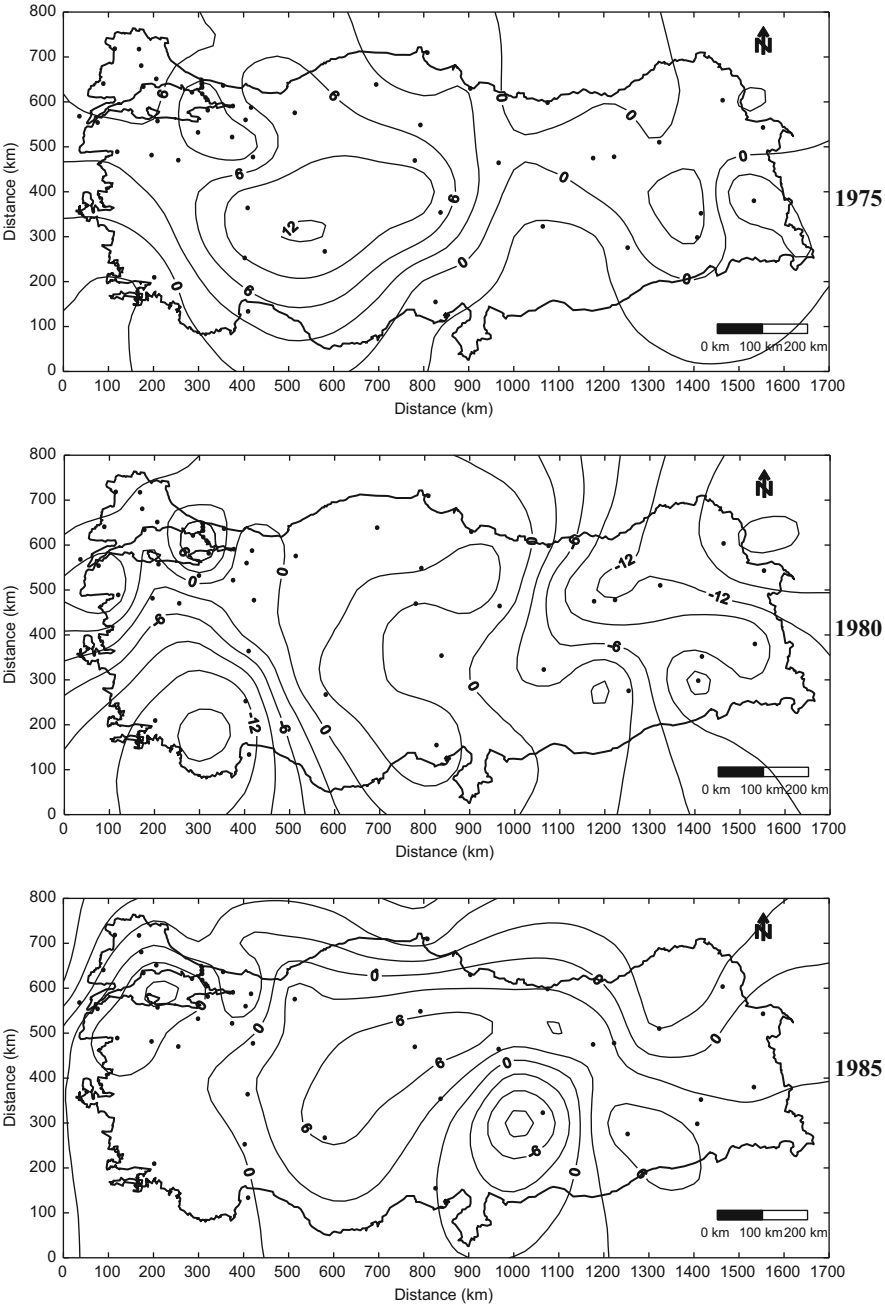


Fig. 4.19 (continued)

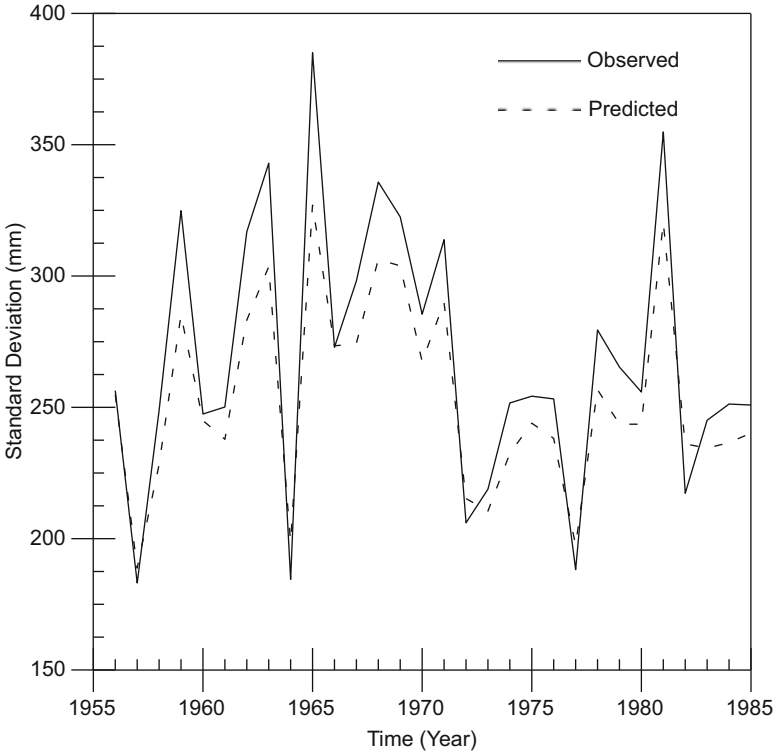


Fig. 4.20 Standard deviation of observed and estimated areal rainfall values for 1956–1985

After the initial state descriptions are read as first step, then the Kalman gain matrix for the one-step prediction can be computed, with necessary assumptions, as the connection matrix is unity, i.e., all stations are reporting their observations. The diagonal elements of measurement noise covariance matrix are taken smaller than those of the system covariance matrix because the observed values are relatively noise-free compared with the errors which result from the system.

Initially, when the model parameters are only rough estimates, with little objective information, the Kalman gain matrix ensures that the measurement data is highly influential in estimating the state parameters. Then, as confidence in the accuracy of the parameters grows with each iteration, the gain matrix values decrease, causing the influence of the measurement data in updating the parameters and associated error.

With the assumption of an initial estimate, the measurement is used to improve the initial estimate by a linear blending of the noisy measurement and the prior estimate. The covariance matrix associated with the updated estimate is computed by using the error covariance matrix associated with initial state vector, connection, and Kalman gain matrices.

The updated estimates can be projected ahead via the transition matrix, where the contribution of system error can be ignored, because it has zero mean. However, the estimate of transition matrix may be difficult, but the system is quite robust to the transition matrix values, and that they can, therefore, be set to fixed values. They have a minimal effect on the results. However, for simplicity transition matrix is assumed to be unity.

It is recommended that one must depend on his/her expert judgement for initial error covariance and system noise covariance matrices' values, so that the updated error covariance matrix values can be computed, and for the system noise covariance matrix, one can use his judgment to select appropriate values. He can then examine the actual operation of the filter and adjust these values online if the situation changes at a later time. Similar to the covariance of initial values, KF is started with large diagonal elements in the system noise covariance matrix.

The projected ahead estimation and the error covariance matrix are used as initial estimation for the next time step. Once the multisite KF loop is entered, then it can be continued as much as necessary.

It is to be noticed that for one station the observed and estimated values follow each other closely, which indicates that KF provides an efficient method for modeling of annual rainfall. The observed and estimated annual rainfall values at 52 selected stations in Turkey for 1985 follow each other closely, which indicates that KF provides an efficient method for modeling of annual rainfall in both time and space dimensions.

The estimated values of annual rainfall at most of the sites in the study area are close to the observed values, especially in the part where more stations are available such as in the northwestern part of Turkey. The percentage error of estimated values vary from -6 (underestimation) to 6 (overestimation) with overall average about 0.121 %.

References

Barnes SL (1964) A technique for maximizing details in numerical weather map analysis. J Appl Meteorol 3:396–409

Benjamin JR, Cornell CA (1970) Probability statistics and decision making in civil engineering. McGraw-Hill Book, New York

Bergthorsson P, Döös BR (1955) Numerical weather map analysis. Tellus 7:329–340

Brown RG, Hwang YC (1992) Introduction to random signals and applied Kalman filtering, 2nd edn. Wiley, New York

Cressman GP (1959) An operational objective analysis system. Mon Weather Rev 87:367–374

Davis A (2002) Statistics data analysis geology. Wiley, New York, 638 pp

Dee DP (1991) Simplification of Kalman filter for meteorological data assimilation. Q J Roy Meteorol Soc 117:365–384

Donnelly KP (1978) Simulation to determine the variance and edge effect of total nearest-neighbour distances. In: Hodder I (ed) Simulation methods in archeology. Cambridge Press, London, pp 91–95

Gandin LS (1963) Objective analysis of meteorological fields. Hydromet Press, New York, 242 pp

Gilchrist B, Cressman GP (1954) An experiment in objective analysis. Tellus 6:309–318
Goodin WR, McRea GJ, Seinfeld JH (1979) A comparison of interpolation methods for sparse data: application to wind and concentration fields. J Appl Meteorol 18:761–771
Harrison PJ, Stevens CF (1975) Bayesian forecasting. University of Warwick, working paper No. 13
Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME Ser D J Basic Eng 82:35–45
Kalman RE, Bucy R (1961) New result in linear filtering and prediction theory. Trans ASME Ser D J Basic Eng 83:95–108
Koch S, Link RF (1971) Statistical analysis of geological data. Dover Publications, New York, 375 p
Koch SE, DesJardins M, Kocin PJ (1983) An iterative Barnes objective map analysis scheme for use with satellite and conventional data. J Appl Meteorol 22:1487–1503
Latif AM (1999) A Kalman filter approach to multisite precipitation modeling in meteorology. Unpublished Ph.D. thesis, Istanbul Technical University, 125 pp
Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266
Sasaki Y (1960) An objective analysis for determining initial conditions for the primitive equations. Tech Rep 60-16T
Şen Z (1989) Cumulative semivariogram models of regionalized variables. Math Geol 21:891–903
Şen Z (2004) Fuzzy logic and system models in water sciences. Turkish Water Foundation Publication, Istanbul, 315 pp
Zadeh LA (1968) Fuzzy algorithms. Inf Control 12:94–102

Chapter 5
Spatial Dependence Measures

Abstract Rather than geometrical weighting functions as in Chap. 2, it is preferable to obtain spatial dependence function from a set of measurement points. Prior to such a functional derivation, it is necessary to examine the isotropy and homogeneity of the spatial data directionally and pointwise features of the regionalized variable (ReV). The basics of semivariogram (SV) with its different components such as sill, nugget, and radius of influence are presented in descriptive and application manners. Similar to SV, cumulative SV (CSV) and point CSV (PCSV) concepts are explained with applications to groundwater quality data. It is emphasized that PCSV helps to depict the spatial behavior features around any sampling point by taking into consideration the contribution from the surrounding measurement points. It is shown that for each location of measurement, it is possible to obtain the radius of influence, if necessary along any direction, and their regional contour maps provide the radius of influence at non-measurement locations. Once the radius of influence is known, then it is possible to depict which nearby measurement locations should be taken into consideration in the calculation of unknown data value. The validity of any method can be decided on the basis of cross-validation error minimization. A new concept of spatial dependence function (SDF) is developed without need that the regional data has normal (Gaussian) probability distribution. The application of SDF is presented for earthquake and precipitation data from Turkey.

Keywords Cumulative semivariogram • Homogeneity • Isotropy • Nugget • Point cumulative semivariogram • Range • Spatial correlation • Sample semivariogram • Sill • Spatial dependence function

5.1 General

Uncertainty within earth sciences data and geological maps is rarely analyzed or discussed. This is undoubtedly due, in part, to the difficulty of analyzing large sets of paper files. The increasing use of computers for storing, retrieving, and serving large data sets has made it easier to systematically analyze these data and preparation of maps and models from them.

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_5

General error theory can be a useful method for characterizing the uncertainty within spatial data. For most data, errors can be due to inaccuracies in record keeping, description, identification of behavior and location, generalization, and correlation. While analysis by error theory can be employed to evaluate the uncertainty within the data used in mapping, other methods can be used to more fully evaluate the uncertainty within specific maps. As already mentioned in Chap. 2, the area of influence method was developed to use three map characteristics (longitude, latitude, elevation) for estimating uncertainty within maps of earth sciences. This method uses the spatial distribution of data points, the probability of misidentification of the targeted unit, and the size of the targeted geological feature to calculate the probability that additional, unidentified targets can exist. Insight gained from the use of error theory and the area of influence methods can be used to describe the uncertainty included in spatial maps.

Earth scientists need maps based even on few samples for extensive decisions in executing certain projects. It is, therefore, necessary for him/her to be able to visualize the subject through limited amount of data preliminarily, which can then be expanded to more effective directions with the coming of additional data. Although there is ready software for mapping but if the user is not familiar with some basic principles of the methodology implemented, then the conclusions may be at bias at the minimum. At the early stages of any study, maps are needed for efficient spatial relationships without any involved mathematical formulations. In general, maps relate three variables (triplets as mentioned in Chap. 2) two of which are the longitude and latitude (geographical coordinates) of the sample points, and therefore, they show the spatial variability of the third variable. However, in this chapter, imaginary maps of three different earth sciences variables are also presented for better logical arguments, interpretations, and conclusions. Maps help to perceive large-scale spatial relationships easily on a small piece of paper.

Maps are based on the point data measurements, the distances between the points, and the density of points. Since most natural phenomenon is of continuous type, maps are representations of finite number of measurement sites and their continuous surface expressions. Hence, the more and better scattered the measurement points within the study area, the better is the map representation of the natural phenomenon. The most common ones are the topographic maps with continuous contour lines of elevations that are derived from a set of discrete location surveys. In the drawing of maps, not only the measurements but also the artistic talent and skills of the expert are also taken into consideration. Hence, in many early maps, subjective biases have entered the domain of mapping, but recently more objective mapping methodologies are developed and the objectivity is more enhanced.

The reliability of contour maps is directly dependent on the total density of the sampling points as well as on their uniform distribution (Chap. 3). However, in practice the uniformity of the sampling points is seldom encountered and the maps are prepared by avoiding this point. For instance, numerical, statistical, or probabilistic versions of weather prediction methods are based on data available from irregularly distributed sites within synoptic regions. The very success of such

prediction methods significantly depends on the production of regularly gridded data at nodes of a mesh laid over the region concerned. Hence, in any spatial prediction study, the following two steps are important:

1. That measured data at practically convenient but irregular sites must be transferred to regularly selected grid points.
2. The use of these grid point estimates in a convenient modeling scheme through digital computers.

5.2 Isotropy, Anisotropy, and Homogeneity

In operational regional interpolation systems, many simplifying assumptions are made about the nature of the correlation, and it is represented by an analytic function of the distance. This assumption implies that the ReV is statistically homogeneous and isotropic. Gustafsson (1981) has defined the statistical homogeneity properties and isotropy of meteorological fields in reference to the covariance. A ReV is homogeneous and second-order stationary, if the covariance is independent of a translation of the two positions. This means that the covariance depends only on the difference between the positional vectors, i.e., distance. Homogeneity in reference to the covariance implies that the variance is constant. A ReV is isotropic in reference to the covariance, if the covariance is independent of a rotation in the field around the center point on the line between two positions.

For any practical application of the spatial modeling, it is necessary to construct a model for the spatial correlation to be used in the scheme. To simplify the construction of these correlation models, it is an advantage, if the homogeneity and isotropy conditions are fulfilled. Much work has gone into proper choice of correlation functions of ReV. One simplifying assumption is that the correlations are assumed to be isotropic and homogeneous, which make them dependent only on distance as stated by Thiebaux and Pedder (1987) and Daley (1991). For instance, the real atmosphere is anything but homogeneous and isotropic, and hence it is assumed that the deviations from the first-guess (climatological mean) field are homogeneous and isotropic. These assumptions are made in most, if not all, interpolation analysis schemes (Bergman 1979; Lorenc 1981; Rutherford 1976; Schlatter 1975). Also homogeneity and isotropy were assumed by Gandin (1963), Eddy (1964), and Kruger (1969a, b) for most meteorological fields.

Natural phenomena physical processes have preferred orientations. For example, at the mouth of a river, the coarse material settles out fastest, while the finer material takes longer to settle. Thus, the closer one is to the shoreline, the coarser the sediments, while the further from the shoreline, the finer the sediments. When interpolating at a point, an observation 100 m away but in a direction parallel to the shoreline is more likely to be similar to the value at the interpolation point than is an equidistant observation in a direction perpendicular to the shoreline. Anisotropy takes these trends in the data into account during the gridding process.

Usually, points closer to the grid node are given more weight than points farther from the grid node. If, as in the example above, the points in one direction have more similarity than points in another direction, it is advantageous to give points in a specific direction more weight in determining the value of a grid node. The relative weighting is defined by the anisotropy ratio. The underlying physical process producing the data and the sample spacing of the data are important in the decision of whether or not to reset the default anisotropy settings.

Anisotropy is also useful when data sets have fundamentally different units along different dimensions. For example, consider plotting a flood profile along a river. The x coordinates are locations, measured in km along the river channel. The t coordinates are time, measured in days. The $Z(x, t)$ values are river depth as a function of location and time. Clearly in this case, the x and t coordinates would not be plotted on a common scale, because one is distance and the other is time (Fig. 5.1). One unit of x does not equal one unit of t. While the resulting map can be displayed with changes in scaling, it may be necessary to apply anisotropy as well.

Another example of anisotropy might be employed for an isotherm map (equal temperature lines, contour map) of average daily temperature over a region. Although the x and y coordinates (as Easting, say x, and Northing, say y) are measured using the same units, the temperature tends to be very similar. Along north-south lines (y lines), the temperature tends to change more quickly (getting colder as one heads toward the north) (see Fig. 5.2). In this case, in gridding the data, it would be advantageous to give more weights to data along the east-west axis than along the north-south axis. When interpolating a grid node, observations that lie in an east-west direction are given greater weight than observations lying an equivalent distance in the north-south direction.

In the most general case, anisotropy can be visualized as an ellipse. The ellipse is specified by the lengths of its two orthogonal axes (major and minor) and by an orientation angle, θ. The orientation angle is defined as the counterclockwise angle between the positive x, and, for instance, minor axis (see Fig. 5.3). Since the ellipse

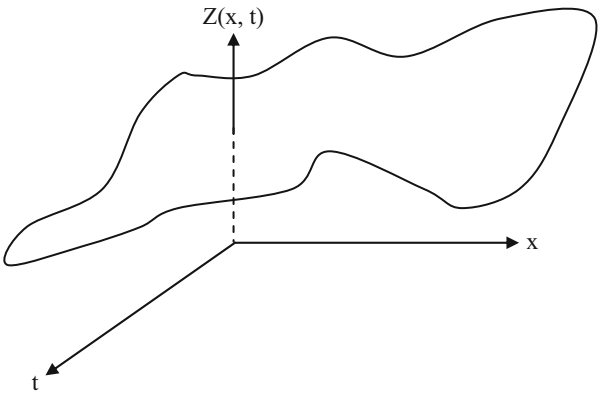


Fig. 5.1 Spatio-temporal depth variations

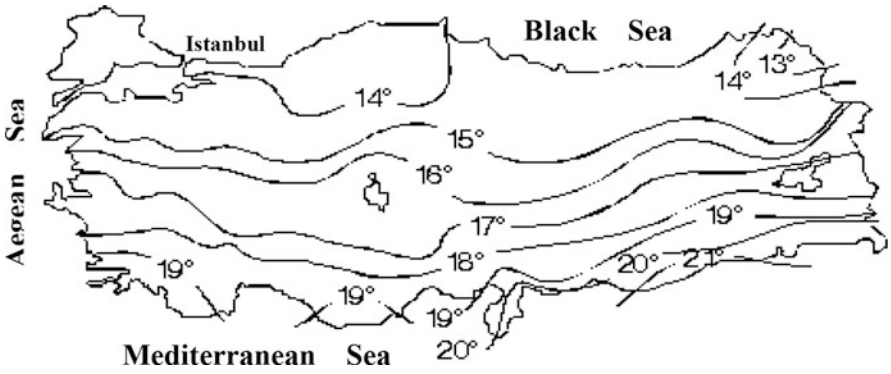


Fig. 5.2 Average annual temperature of Turkey

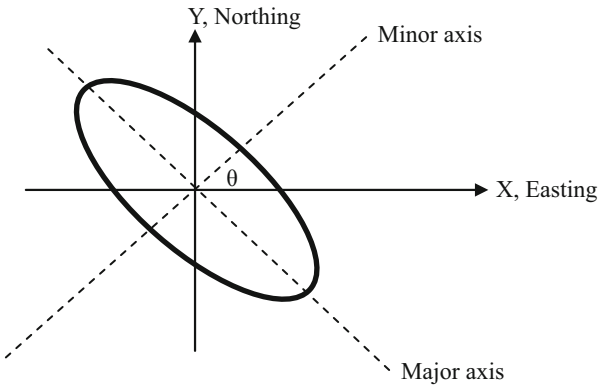


Fig. 5.3 Anisotropy ratio and angle

is defined in this manner, an ellipse can be defined with more than one set of parameters.

For most of the gridding methods, the relative lengths of the axes are more important than the actual length of the axes. The relative lengths are expressed as a ratio in the anisotropy group. The ratio is defined as major axis divided by minor axis. If it is equal to 1, then the ellipse takes the form of a circle. The angle is the counterclockwise angle between the positive x axes and minor axis. The small picture in the anisotropy group displays a graphic of the ellipse to help describe the ellipse. An anisotropy ratio less than 2 is considered mild, while an anisotropy ratio greater than 4 is considered severe. Typically, when the anisotropy ratio is greater than 3, its effect is clearly visible on grid-based maps. The angle is the preferred orientation (direction) of the major axis in degrees.

An example where an anisotropy ratio is appropriate is an oceanographic survey to determine water temperature at varying depths. Assume the data are collected every 1000 m along a survey line and temperatures are taken every 10 m in depth at each sample location. With this type of data set in mind, consider the problem of

creating a grid file. When computing the weights to assign to the data points, closer data points get greater weights than points farther away. A temperature at 10 m in depth at one location is similar to a sample at 10 m in depth at another location, although the sample locations are 1000s of meters apart. Temperatures might vary greatly with depth, but not as much between sample locations.

5.3 Spatial Dependence Function (SDF)

The first step is referred to as the objective analysis and the second one is the spatial modeling phase. For sure, a sound objective analysis is primary prerequisite of successful modeling. For instance, meteorologists strive for effective interpolation in order to enhance their mesoscale analysis and forecasts. Objective analysis studies of meteorological variables started with the work by Panofsky (1949). He attempted to produce contour lines of upper-wind movements by fitting third-order polynomials and employing least-squares method to the observations at irregular sites. The least-squares method leads to predicted field variables, which depend strongly on distribution of data points when a suitable polynomial is fitted to full grid. Optimum analysis procedures are introduced to meteorology by Eliassen (1954) and Gandin (1963). These techniques employ historical data about the structure of the atmosphere to determine the weights to be applied to the observations. Here, the implied assumption is that the observations are spatially correlated. Consequently, observations that are close to each other are highly correlated; hence, as the observations get farther apart, the spatial dependence decreases. It is a logical consequence to expect regional dependence function as in Fig. 5.4 assuming that at zero distance, the dependence is equal to 1, and then onward there is a continuous decrease or decreasing fluctuations depending on the ReV behavior.

In this figure, there are three spatial dependence functions (SDFs) as A, B, and C. Logically, A and B indicate rather homogeneous and isotropic regional behavior of ReV, whereas C has local differences at various distances. However, all of them decrease down to zero SDF value. The distance between the origin and the point where the SDF is almost equal to zero shows the radius of influence as R_1 or R_2. Provided that the ReV behavior is isotropic (independent of direction), then the radius of area can be calculated as a circle around each station with radius equal to

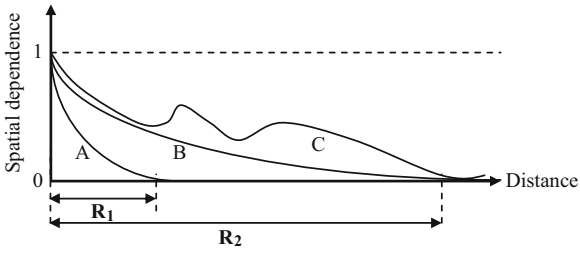


Fig. 5.4 Spatial dependence functions

the radius of influence. These are subjective and expert views about the spatial dependence structure of any ReV. Their objective counterparts can be obtained from a set of spatial data as will be explained later in this chapter. The spatial predictions are then made by considering a spatial model with a domain equal to the radius of area. For instance, Gilchrist and Cressman (1954) reduced the domain of polynomial fitting to small areas surrounding each node with a parabola. Bergthorsson and Döös (1955) proposed the basis of successive correction methods, which did not rely only on interpretation to obtain grid point values but also a preliminary guess field is initially specified at the grid points (Chap. 5). Cressman (1959) developed a number of further correction versions based on reported data falling within a specified distance R from each grid point. The value of R is decreased with successive scans (1500 km, 750 km, 500 km, etc.) and the resulting field of the latest scan is taken as the new approximation. Barnes (1964) summarized the development of a convergent weighted-averaging analysis scheme which can be used to obtain any desired amount of detail in the analysis of a set of randomly spaced data. The scheme is based on the supposition that the 2D distribution of a ReV can be represented by the summation of an infinite number of independent waves, i.e., Fourier integral representation. A comparison of existing objective methods up to 1979 for sparse data is provided by Goodin et al. (1979). Their study indicated that fitting a second-degree polynomial to each subregion triangular in the plane with each data point weighted according to its distance from the subregion provides a compromise between accuracy and computational cost. Koch et al. (1983) presented an extension of the Barnes method which is designed for an interactive computer scheme. Such a scheme allows real-time assessment both of the quality of the resulting analyses and of the impact of satellite-derived data upon various earth sciences data sets. However, all of the aforementioned objective methods have the following common drawbacks:

1. They are rather mechanical without any physical foundation but rely on the regional configuration of irregular sites. Any change in site configuration leads to different results although the same ReV is sampled.
2. They do not take into consideration the spatial covariance or correlation structure within the ReV concerned.
3. They have constant radius of influence without any directional variations.

Hence, spatial anisotropy of observed fields is ignored. Although some anisotropic distance function formulations have been proposed by Inman (1970) and Shenfield and Bayer (1974), all of them are developed with no explicit quantitative reference to the anisotropy of observed field structure of the ReV.

According to Thiebaux and Pedder (1987) assessment of the work done by Bergthorsson and Döös, "the most obvious disadvantage of simple inverse distance-weighing schemes is that they fail to take into account the spatial distribution of observations relative to each other." Two observations at equidistant from a grid point are given the same weight regardless of relative values at measurement sites. This may lead to large operational biases in grid point data when some observations are much closer together than others within the area of influence.

Especially after the 1980s, many researchers are concentrated on the spatial covariance and correlation structures of the ReV. Lorenc (1981) has developed a methodology whereby first of all the grid points in a subregion are analyzed simultaneously using the same set of observations and then subareas are combined to produce the whole study area analysis. Some papers are concerned with the determination of unknown parameters of the other covariance functions or SDFs which provide required weightings for ReV data assimilation. Along this line, the idea proposed by Bratseth (1986) depends on the interpretation of the ReV covariances into the objective analysis. His analysis caused a recent resurgence of the successive correction method in which the optimal analysis solution is approached. His method uses the correlation function for the forecast errors to derive weights that are reduced in regions of higher data density. Later, Sashegyi (1960) employed his methodology for the numerical analysis of data collected during the Genesis of Atlantic Lows Experiment (GALE). Practical conclusions of Bratseth's approach have been reported by Franke (1988) and Seaman (1988).

On the other hand, Buzzi et al. (1991) described a simple and economic method for reducing the errors that can result from the irregular distribution of data points during an objective analysis. They have demonstrated that a simple iterative method cannot improve only analysis accuracy but also results in an actual frequency response that approximates closely the predicted weight-generating function. They have shown that in the case of heterogeneous spatial sampling, a Barnes analysis could produce an unrealistic interpolation of the sampled field even when this is reasonably well resolved by error-free observations. Iteration of a single correction algorithm led to the method of successive correction (Daley 1991). The method of successive correction has been applied as a means to tune adaptively, the a posteriori weights. Objective analysis schemes are practical attempts to minimize the variance estimation (Thiebaux and Pedder 1987).

Pedder (1993) provided a suitable formation for successive correction scheme based on a multiple iteration using a constant influence scale that provides more effective approach to estimate ReV from scattered observations than the more conventional Barnes method which usually involves varying the influence scale between the iterations. Recently, Dee (1995) has presented a simple scheme for online estimation of covariance parameters in statistical data assimilation systems. The basis of the methodology is a maximum likelihood approach in which estimates are obtained through a single batch of simultaneous observations. Simple and adaptive Kalman filtering techniques are used for explicit calculation of forecast error covariance (Chap. 3). However, the computational cost of the scheme is rather high.

Field measurements of ReV such as ore grades, chemical constitutions in groundwater, fracture spacing, porosity, permeability, aquifer thickness, and dip and strike of a structure are dependent on the relative positions of measurement points within the study area. Measurements of a given variable at a set of points provide some insight into the spatial variability. This variability determines the ReV behavior as well as its predictability. In general, the larger the variability, the more heterogeneous is the ReV environment, and as a result, the number of

measurements required to model, to simulate, to estimate, and to predict the ReV is expected to be large. Large variability implies also that the degree of dependence might be rather small even for data whose locations are close to each other. A logical interpretation of such a situation may be that either the region was subjected to natural phenomena such as tectonics, volcanism, deposition, erosion, recharge, climate change, etc., or later to some other human activities as pollution, groundwater abstraction, mining, etc.

However, many types of ReV are known to be spatially related in that the closer their positions, the greater is their dependence. For instance, spatial dependence is especially pronounced in hydrogeological data due to groundwater flow as a result of the hydrological cycle, which homogenizes the distribution of chemical constituents within the heterogeneous mineral distribution in geological formations.

The factors of ReV are sampled at irregular measurement points within an area at regular or irregular time intervals. No doubt, these factors show continuous variations with respect to other variables such as temperature, distance, etc. Furthermore, temporal and spatial ReV evolution are controlled by temporal and spatial correlation structures within the ReV itself. As long as the factors are sampled at regular time intervals, the whole theory of time series is sufficient in their temporal modeling, simulation, and prediction. The problem is with their spatial constructions and the transfer of information available at irregular sites to regular grid nodes or to any desired point. Provided that the structure of spatial dependence of the ReV concerned is depicted effectively, then any future study, such as the numerical predictions based on these sites, will be successful. In order to achieve such a task, it is necessary and sufficient to derive the change of spatial correlation for the ReV data with distance.

In order to quantify the degree of variability within spatial data, variance techniques can be used in addition to classical autocorrelation methods (Box and Jenkins 1976). However, these methods are not helpful directly to account for the spatial dependence or for the variability in terms of sample positions. The drawbacks are due to either non-normal (asymmetric) distribution of data or irregularity of sampling positions. However, the semivariogram (SV) technique, developed by Matheron (1965, 1971) and used by many researchers (Clark 1979; Cooley 1979; David 1977; Myers et al. 1982; Journel 1985; Aboufirassi and Marino 1984; Hoeksema and Kitandis 1984; Carr et al. 1985) in diverse fields such as geology, mining, hydrology, earthquake prediction, groundwater, etc., can be used to characterize spatial variability and hence the SDF. The SV is a prerequisite for best linear unbiased prediction of ReV through the use of Kriging techniques (Krige 1982; Journel and Huijbregts 1978; David 1977).

5.4 Spatial Correlation Function (SCF)

By definition, SCF, ρ_{ij} between i and j, takes values between -1 and $+1$ and can be calculated from available historical data as

$$\rho_{ij} = \frac{\overline{\left(Z_i^o - \overline{Z}_i\right)\left(Z_j^o - \overline{Z}_j\right)}}{\sqrt{\left(\overline{Z_i^o - \overline{Z}_i}\right)^2 \left(\overline{Z_j^o - \overline{Z}_j}\right)^2}} \tag{5.1}$$

where over bars indicate time averages over a long sequence of past observations, Z_i^o and Z_j^o represent observed precipitation amounts at these stations, and, finally, $\overline{Z}_i$ and $\overline{Z}_j$ are the climatological mean of precipitations. Furthermore, ρ_{ij} is thought as attached with the horizontal distance $D_{i,j}$ between stations and j. Consequently, if there are n stations, then there will be $m = n(n-1)/2$ pairs of distances and corresponding correlation coefficients. Their plot results in a scatter diagram which indicates the SCF pattern for the regional rainfall amounts considered as a random field. Figure 5.5 presents such scatter diagrams of empirical SCFs concerning monthly rainfall amounts (Şen and Habib 2001). At a first glance, it is obvious from this figure that there are great scatters at any given distance in the correlation coefficients, and unfortunately, one cannot identify easily a functional trend. The scatter can be averaged out by computing mean correlation coefficient over relatively short distance intervals (Thiebaux and Pedder 1987). The following significant points can be deduced from these SCFs:

1. Each monthly average SCF shows a monotonically decreasing trend.
2. Due to averaging procedure within the first 15 km interval, it may appear in Figs. 5.5 and 5.6 that the correlation coefficient at lag zero is not equal to +1 as expected. Monthly average SCFs for data considered are given in Fig. 5.6. Herein, averaging is taken over successive 30 km intervals. This discrepancy is entirely due to the averaging scheme rather than a physical reality. Hence, this is not a physically plausible result but unavoidable consequence of the averaging procedure.
3. The more the averaging correlation coefficient within the first 30 km, the more strongly related spatial correlation appears between the measurement sites.

5.4.1 Correlation Coefficient Drawback

Although the cross-correlation function definition can give a direct indication of the dependence of variations from the mean at any two sites, it suffers from the following drawbacks:

1. Autocorrelation and cross-correlation formulations require symmetrical (normal, Gaussian) PDF of data for reliable calculations. It is well established in the literature that most of the earth sciences data PDFs accord rarely with normal (Gaussian) PDF but better with Weibull, gamma, or logarithmic PDFs (Benjamin and Cornell 1970; Şen 2002).

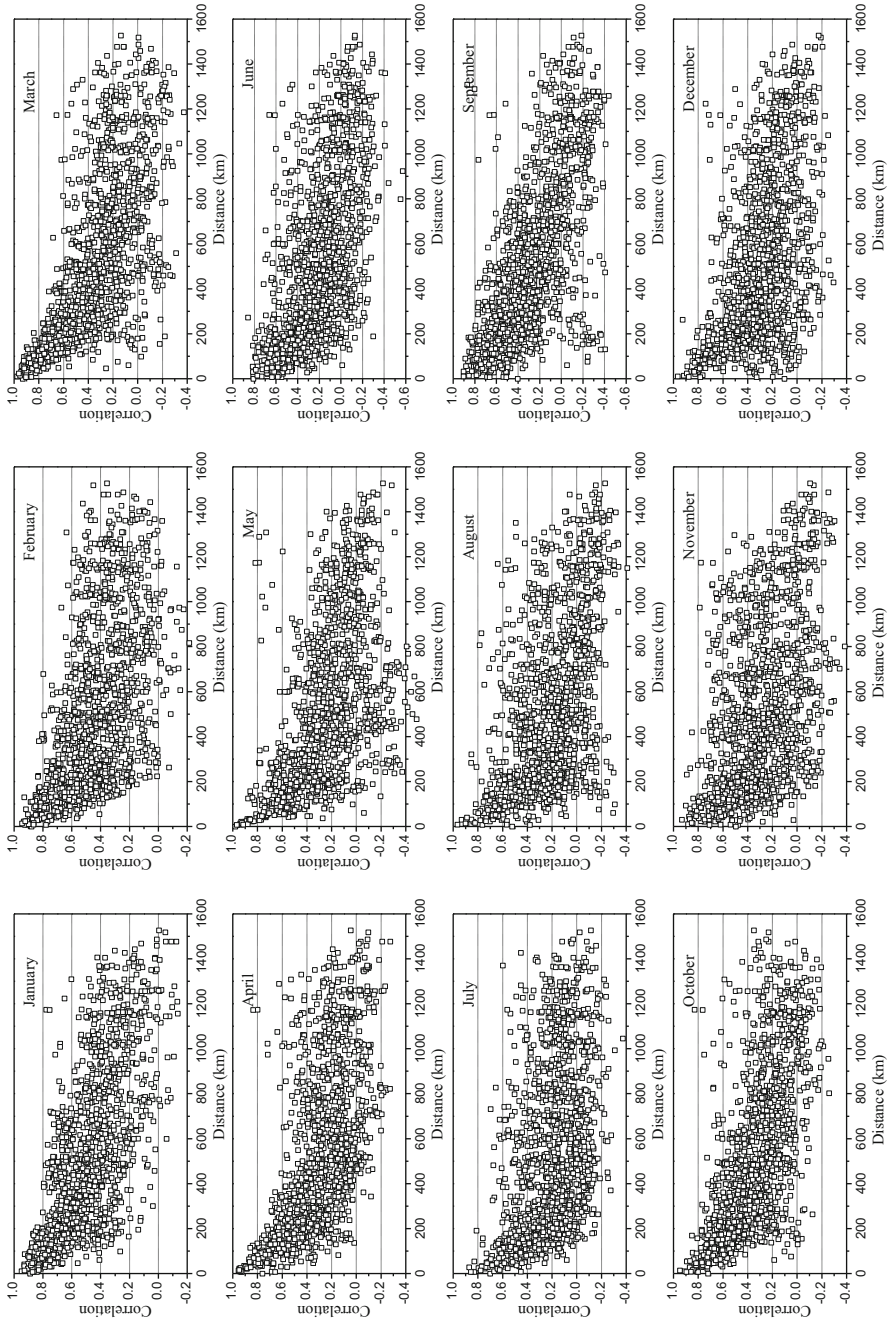


Fig. 5.5 Empirical spatial correlation functions

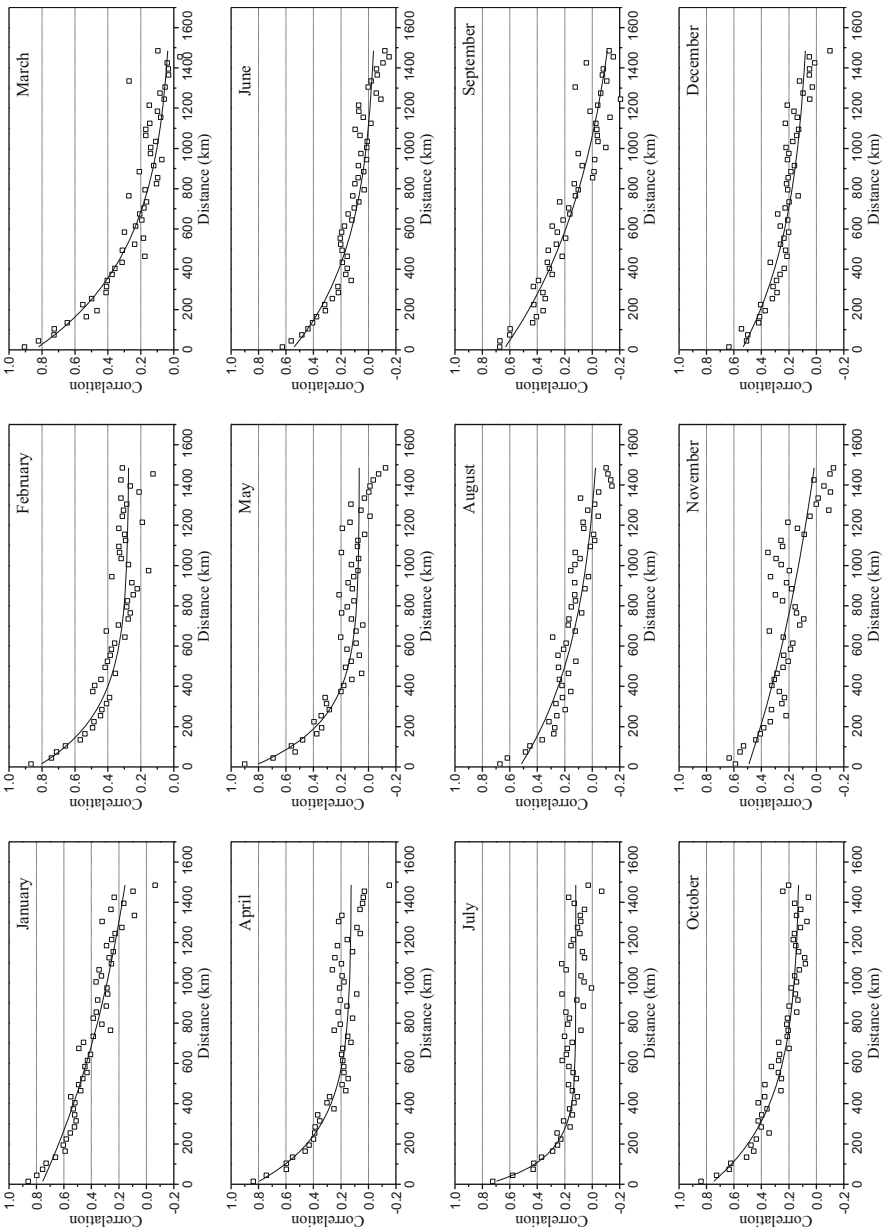


Fig. 5.6 Average and theoretical spatial correlation coefficient

2. Since the cross correlation (as the autocorrelation) is valid for symmetrically distributed ReV (and random variables, RVs), the available data must be transformed into normal PDF prior to application of these methods.
3. In the spatial calculation of the cross correlation, it is necessary to have a sequence of measurements with time at each site, which is not the case in many earth sciences problems where there are only one measurement at each site.
4. The correlation function measures the variation around the arithmetic average values of the measurements at individual sites. However, in the spatial variability calculations, a measure of relative variability between two sites is necessary.
5. Especially for the last two points, the SV (Matheron 1965) or cumulative SV (CSV) (Sen 1989) concepts are developed, and their modifications as the point CSV (PCSV) are presented and used for the regional assessment of earth sciences data.

For instance, Barros and Estevan (1983) presented a method for evaluating wind power potential from a 3-month long wind record at a site and data from a regional network of wind systems. Their key assumption was that "wind speed has some degree of spatial correlation" which is a logical conclusion, but they failed to present an effective method for the objective calculation of the spatial variability except by employing cross- and autocorrelation techniques. Their statement does not provide an objective measure of spatial correlation. Skibin (1984) raised the following questions:

1. What is "a reasonable spatial correlation"? Are the correlation coefficients between the weekly averages of wind speed a good measure of it? Answers to these questions are necessary by any objective method. For instance, PCSV technique can be employed for this purpose.
2. Do the weekly averages represent the actual ones?
3. How applicable to the siting of wind generators are the results obtained by the use of spatial correlation coefficients?

In deciding about the effectiveness of the wind speed measurement at a site around its vicinity, the topographic and climatic conditions must be taken into consideration. The smaller the area of influence, the more homogeneous orographic, weather, and climatologic features are, and, consequently, the simplest is the model. However, large areas more than 1000 km in radius around any site may contain different climates with different troughs and ridges and high- and low-pressure areas with varying intensities. Furthermore, in heterogeneous regions with varying surface properties (such as land-sea-lake-river interfaces) and variable roughness parameters, the local wind profile and wind potential can be affected significantly. The wind energy potential and the groundwater availability are highly sensitive to height variations of hills, valleys, and plains (Şen 1995). The reasons for wind speed variations are not only of orographical origin but also of different flow regimes (i.e., anabatic-katabatic influences compared with hilltop conditions, upwind compared with leeside sites, flow separation effects). All these effects will

lose their influence further away from the siting point. It can be expected that a smaller distance from the site corresponds to a larger correlation. Here again it is obvious that the spatial dependence decreases with distance as in Fig. 5.5 (correlation property). Barros and Estevan (1983) noticed that a small region had higher correlation coefficients between the sites. On the contrary, the spatial independence increases with the distance (SV property).

Barchet and Davis (1983) have stated that better estimates are obtained when the radius of influence is about 200 km from the site. However, this information is region dependent, and there is a need to develop an objective technique whereby the radius of influence can be estimated from a given set of sites.

5.5 Semivariogram (SV) Regional Dependence Measure

The regional correlation coefficient calculation requires a set of assumptions which are not taken into consideration in practical applications (Şen 2008). First of all, calculation of spatial correlation coefficient requires time series records at each site. This is not possible in many earth sciences studies such as in ore grading, soil properties, hydrogeological parameters, etc. Rather than a time series availability at each site, there is only one record, say ore grade record at a set of sites, and therefore it is not possible to calculate spatial correlation function. However, the only way to depict the spatial correlation from a set of single records at a set of locations is through the SV methodology.

5.5.1 SV Philosophy

The very basic definition of the SV says that it is the half-square difference variation of the ReV by distance. ReV theory does not use the autocorrelation, but instead uses a related property called the SV to express the degree of relationship between measurement points in a region. The SV is defined simply as half-square (variance) of the differences between all possible point pairs spaced a constant distance, d, apart. The SV at a distance $d = 0$ should be zero, because there are no differences (variance) between points that are compared to themselves. The magnitude of the SV between points depends on the distance between the points. The smaller the distance, the smaller is the SV and at larger distances SV value is larger. The SV is a practical measure of average spatial changes. The underlying principle is that, on the average, two observations closer together are more similar than two observations farther apart. This is a general statement where the directional changes are not considered. The plot of the SV values as a function of distance from a point is referred to as a SV. However, as points are compared to increasingly distant points, the SV increases.

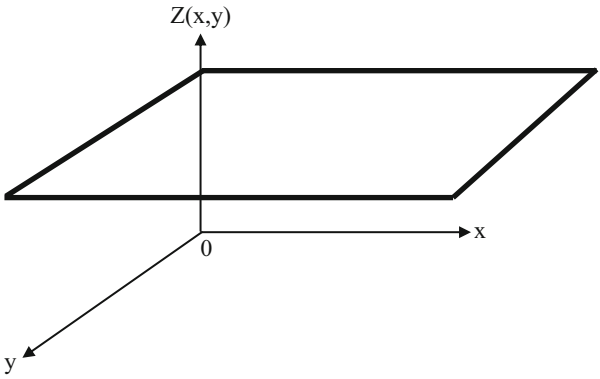


Fig. 5.7 Homogeneous, isotropic, and uniform ReV

Fig. 5.8 Perfectly homogeneous and isotropic ReV SV

The simplest and most common form of ReV is a triplet, and therefore it is illuminating first to consider the surface in 3D, and then according to the SV definition, it is possible to infer its shape intuitively by mental experiment:

1. Continuously deterministic uniform spatial data: If the ReV is a deterministic horizontal surface of homogeneous, isotropic, and uniform data as in Fig. 5.7, then the average half-square difference of such data is zero at every distance as in Fig. 5.8.
2. Discontinuously deterministic partially uniform spatial data: The continuity in Fig. 5.7 is disrupted by a discontinuous feature (cliff, fault, facies change, boundary, etc.) as in Fig. 5.9.

   The average square difference at various distances leads to an SV with a discontinuity at the origin (see Fig. 5.10), the amount of which is equal to the square difference between higher, $Z_H(x, y)$, and lower, $Z_L(x, y)$, data values as

$$\gamma(d) = [Z_H(x, y) - Z_L(x, y)]^2 \tag{5.2}$$

   The resulting SV is expected to take the shape as in Fig. 5.10, where there is a nonzero value at the origin. Such a jump at the origin indicates discontinuity embeddings in the spatial event and it is referred to as "sill" in geostatistical literature.

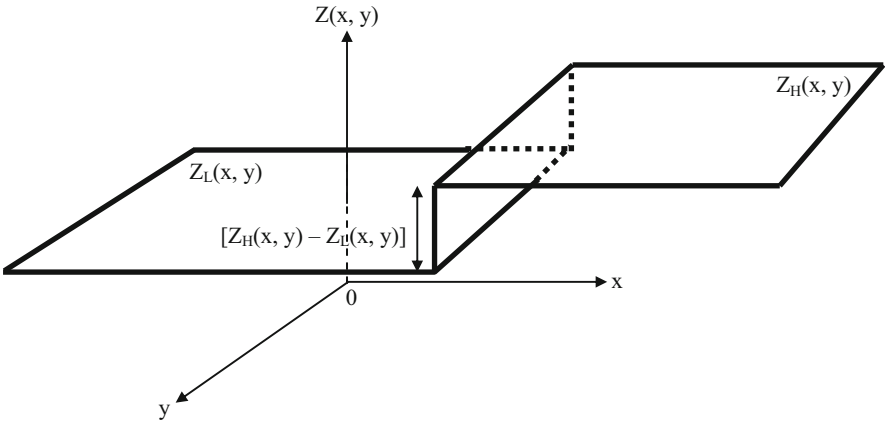
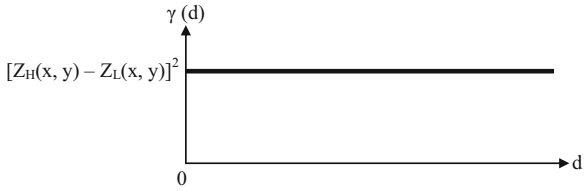


Fig. 5.9 Discontinuous surface

Fig. 5.10 Completely random ReV SV



3. Continuously deterministic spatially linear trend data: If the ReV is a linear surface along the *x* axis as in Fig. 5.11, then the SV along the *x* axis by definition has a quadratic form without any decrease (Fig. 5.12).
   This SV does not have any horizontal portion, and at large distances, the slope increases in an extreme manner.
4. Discontinuously deterministic spatially linear trend data: If the trend surface in Fig. 5.11 has a discontinuity (Fig. 5.13), then the SV shape appears as in Fig. 5.14, where there is a jump at the origin, which is referred to as nugget effect in SV terminology.
5. Completely independent spatial data: If the ReV is completely random with no spatial correlation as in Fig. 5.15, then the SV will be equal to the variance, σ^2, of the ReV at all distances as in Fig. 5.16. A decision can be made about the continuity (or discontinuity) and smoothness of the ReV by visual inspection from the sample SV. If at small distances the sample SV does not indicate passage from the origin (nugget effect), then the ReV includes discontinuities, where there is no regional dependence in the ReV at all. Its SV appears as a horizontal straight line similar to SV in Fig. 5.10.

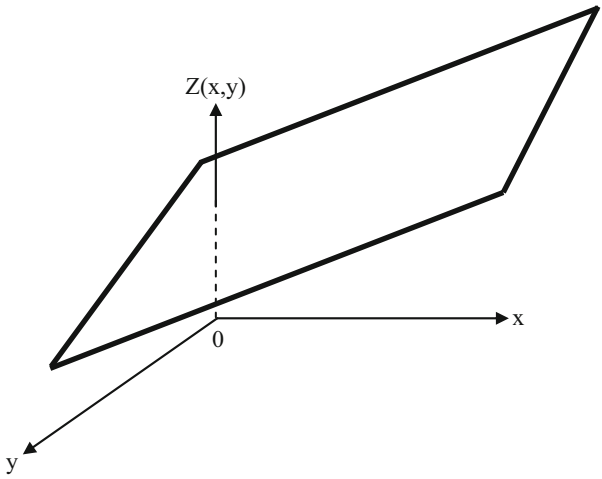


Fig. 5.11 Continuous linear trend

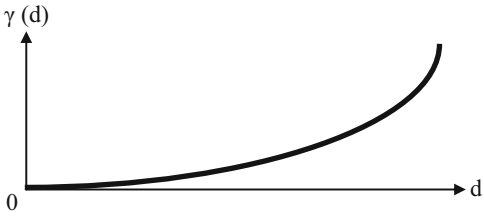


Fig. 5.12 Linear trend surface SV in x direction

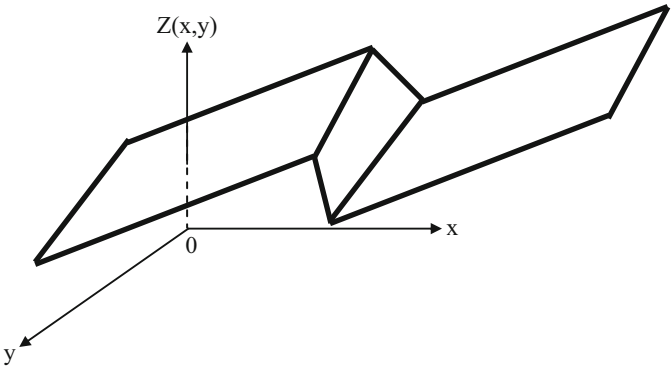


Fig. 5.13 Discontinuous trend surface

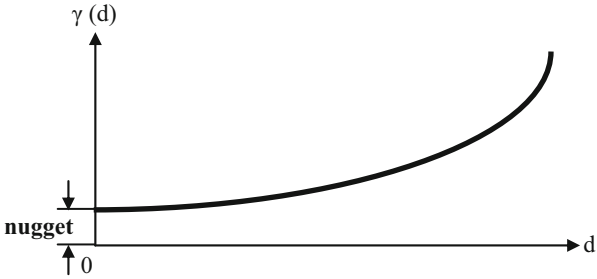


Fig. 5.14 Discontinuous trend surface SV in x direction

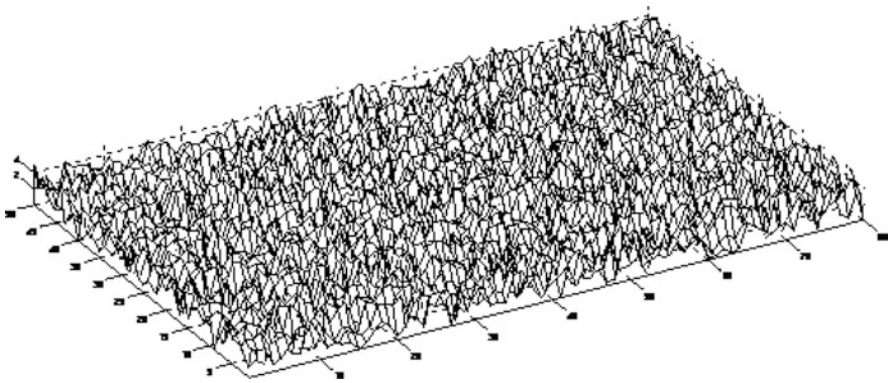


Fig. 5.15 Independent spatial data

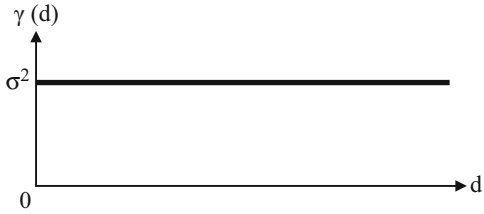


Fig. 5.16 Completely random ReV SV

The SV in this spatial random event case is equivalent to the expectation of Eq. 5.2, which after expansion and expectation E(.) operation application on both sides leads to

$$E[\gamma(d)] = E\left[Z_H^2(x,\ y)\right] - 2E[Z_H(x,\ y)Z_L(x,\ y)] + E\left[Z_L^2(x,\ y)\right]$$

Since the ReV is assumed as spatially independent with zero mean (expectation), the second term of this expression is equal to zero and the other terms are equal to the variance, σ^2, of the spatial event. Finally, this last expression yields $E[\gamma(d)] = 2\sigma^2$. In order to have the SV expectation equal to the variance in practical applications, it is defined as the half-square difference instead of square difference

as in Eq. 5.2. Consequently, the SV of an independent ReV appears as having a sill value similar to Fig. 5.10 but this time the sill value is equal to the spatial variance of the ReV.

5.5.2 SV Definition

The SV is the basic geostatistical tool for visualizing, interpreting, modeling, and exploiting the regional dependence in a ReV. It is well known that even though the measurement sites are irregularly distributed, one can find central statistical parameters such as mean, median, mode, variance, skewness, etc., but they do not yield any detailed information about the phenomenon concerned. The greater the variance the greater is the variability, but unfortunately this is a global interpretation without detailed useful information. The structural variability in any phenomenon within an area can best be measured by comparing the relative change between two sites. For instance, if any two sites, distant d apart, have measured concentration values Z_i and Z_{i+d}, then the relative variability can simply be written as $(Z_i - Z_{i+d})$. However, similar to Taylor (1915) theory concerning turbulence, the square difference, $(Z_i - Z_{i+d})^2$, represents this relative change in the best possible way. This square difference has appeared first in the Russian literature as the "structure function" of ReV. It subsumes the assumption that the smaller the distance, d, the smaller will be the structure function. Large variability implies that the degree of dependence among earth sciences records might be rather small even for sites close to each other.

In order to quantify the degree of spatial variability, variance and correlation techniques have been frequently used in the literature. However, these methods cannot account correctly for the spatial dependence due to either non-normal PDFs and/or irregularity of sampling positions.

The classical SV technique has been proposed by Matheron (1965) to eliminate the aforementioned drawbacks. Mathematically, it is defined as a version of Eq. 5.26 by considering all of the available sites within the study area as (Matheron 1965; Clark 1979)

$$\gamma(d) = \frac{1}{2n_d}\sum_{k=1}^{n_d}(Z_i - Z_{i+d})^2 \tag{5.3}$$

where k is the counter of the distance which can be expanded by considering the regional arithmetic average, $\overline{Z}$, of the ReV as follows:

$$\begin{aligned}\gamma(d) &= \frac{1}{2}\sum_{k=1}^{n_d}\left[(Z_i - \overline{Z}) - (Z_{i+d} - \overline{Z})\right]^2 \\ &= \left[(Z_i - \overline{Z})^2 - 2(Z_i - \overline{Z})(Z_{i+d} - \overline{Z}) + (Z_{i+d} - \overline{Z})^2\right]\end{aligned}$$

The elegancy of this formulation is that the ReV PDF is not important in obtaining the SV, and furthermore, it is effective for regular data points. It is to be recalled, herein, that the classical variogram, autocorrelation, and autorun techniques (Şen 1978) all require equally spaced data values. Due to the irregularly spaced point sources, the use of classical techniques is highly questionable, except that these techniques might provide biased approximate results only. The SV technique, although suitable for irregularly spaced data, has practical difficulties as summarized by Sen (1989). Among such difficulties is the grouping of distance data into classes of equal or variable lengths for SV construction, but the result appears in an inconsistent pattern and does not have a nondecreasing form as expected in theory. As the name implies a SV, $\gamma(d)$, is a measure of spatial dependence of a ReV.

Due to independence any cross multiplication of Z_i and Z_j will be equal to zero on the average, and hence this is equivalent to regional variance, σ^2, as explained in the previous section. Figure 5.16 shows this mental experiment SV as a horizontal straight line. Hence, at every distance, the SV is dominated by sill value only. Expert reasoning of SV models in the previous figures helps to elaborate some fundamental and further points as follows:

1. If the ReV is continuous without any discontinuity, then the SV should start from the origin, which means that at zero distance, SV is also zero (Figs. 5.8 and 5.12).
2. If there is any discontinuity within the ReV, then at zero distance, a nonzero value of the SV appears as in Figs. 5.10, 5.14, and 5.16.
3. If there is an extensive spatial dependence, then the SV has increasing values at large distances (Figs. 5.12 and 5.14).
4. When the spatial dependence is not existent, then the SV has a constant nonzero value equal to the regional variance of the ReV at all distances as in Fig. 5.16.
5. Under the light of all what have been explained so far, it is logically and rationally obvious that in the case of spatial dependence structure in ReV, the SV should start from zero at zero distance and then will reach the regional variance value as a constant at large distances. The SV increases as the distance increases until at a certain distance away from a point, it equals the variance around the average value of the ReV and will therefore no longer increase, causing a flat (stabilization) region to occur on the SV, which is called as a sill (Fig. 5.17). The horizontal stabilization level of sample SV is referred to as its sill. The distance at which the horizontal SV portion starts is named as the range, R, radius of influence or dependence length after which there is no spatial (regional) dependence between data points. Only within this range, locations are related to each other, and hence all measurement locations in this region are the nearest neighbors that must be considered in the estimation process. This implies that the ReV has a limited areal extend over which the spatial dependence decreases or independence increases in the SV sense as in Fig. 5.17.

   The classical SV is used to quantify and model spatial correlations. It reflects the idea that closer points have more regional dependence than distant points. In general, spatial prediction is a methodology that embeds the spatial dependence in the model structure,

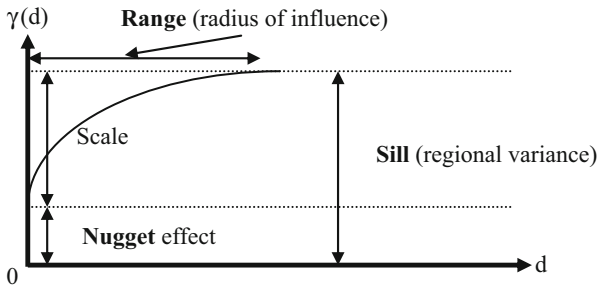


Fig. 5.17 Classical global SV and elements

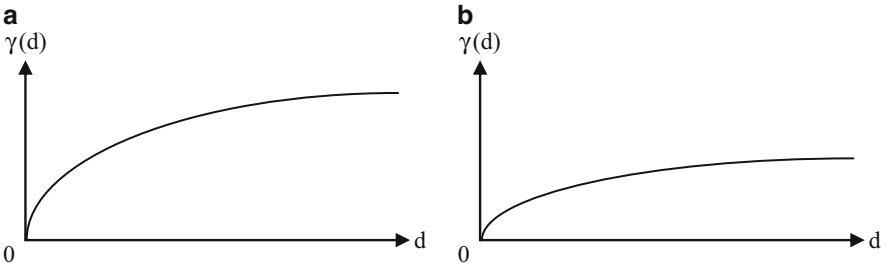


Fig. 5.18 Classical directional SV, (**a**) major axis, (**b**) minor axis

6. At some distance, called the range, the SV will become approximately equal to the variance of the ReV itself (see Fig. 5.17). This is the greatest distance over which the value at a point on the surface is related to the value at another point. The range defines the maximum neighborhood over which control points should be selected to estimate a grid node, to take advantage of the statistical correlation among the observations. In the circumstance where the grid node and the observations are spaced so that all distances exceed the range, Kriging produces the same estimate as classical statistics, which is equal to the mean value.
7. However, most often natural data may have preferred orientations, and as a result, ReV values may change more along the same distance in one direction than another (Fig. 5.3). Hence, in addition to distance, the SV becomes a function of direction (Fig. 5.18).
   It is possible to view the general of SV as a 3D function as the change of SV value, $\gamma(\theta,d)$, with respect to direction, θ, and separation distance, d. Of course, θ and d are the independent variables. In general, specification of any SV requires the following information:

   (a) Sill (regional variance)
   (b) Range (radius of influence)
   (c) Nugget (zero distance jump)
   (d) Directional values of these parameters

The last point is helpful for the identification of regional isotropy or anisotropy. For the Kriging application, the convenient composition of these parameters must be identified through a theoretical SV. Whether a given sample SV is stationary or not can be decided from its behavior at large distances. If the large distance portion of the SV approaches a horizontal line, then it is stationary, which means intuitively that there are rather small fluctuations with almost the same variance at every corner of the region.

If the SV is generated from paired points selected just based on distance (with no directional component), then it is called isotropic (iso means the same; tropic refers to direction) or omnidirectional. In this case, the lag-distance measure is a scalar and the SV represents the average of all pairs of data without regard to their orientation or direction. A standardized SV is created by dividing each SV value by the overall sample variance, which allows SVs from different data sets on the same entity for facilitating the mutual comparison.

On the other hand, SVs from points that are paired based on direction and distance are called anisotropic (meaning not isotropic). In this case, the lag measure is a vector. The SVs in this case are calculated for data that are in a particular direction as explained in Sect. 4.3. The regularity and continuity of the ReV of a natural phenomenon are represented by the behavior of SV near the origin. In SV models with sill (Fig. 5.17), the horizontal distance between the origin and the end of SV reflects the zone where the spatial dependence and the influence of one value on the other occur, and beyond this distance, the ReV $Z(x)$ and $Z(x + d)$ are independent from each other. Furthermore, SVs, which increase at least as rapidly as d^2 for large distances d, indicate the presence of drift (trend), i.e., nonstationary mathematical expectation. Plot of SV graphs for different directions gives valuable information about continuity and homogeneity. If SV depends on distance d only, it is said to be isotropic, but if it depends on distance as well as direction, it is said to be anisotropic. A properly fitted theoretical SV model allows linear estimation calculations that reflect the spatial extent and orientation of spatial dependence in the ReV to be mapped. Details on these points can be found in standard textbooks on geostatistics (Davis 1986; Clark 1979).

There are also indicator SVs which are calculated from data that have been transformed to a binary form (1 or 0), indicating the presence or absence of some variable or values that are above some threshold. In the calculation of sample SVs, the following rules of thumb must be considered:

1. Each distance lag (d) class must be represented by at least 30–50 pairs of points.
2. The SV should only be plotted out to about half the width of the sampling space in any direction.

Characterizing spatial correlation across the site through experimental SV can often be the most time-consuming step in a geostatistical analysis. This is particularly true if the data are heterogeneous or limited in number. Without a rationale for identifying the major direction of anisotropy, the following steps might be useful in narrowing the focus of the exercise:

1. Begin with an omnidirectional SV with a bandwidth large enough to encompass all data points on the site. In practice, maximum lag distance can be taken as one third of the maximum distance between the data points.
2. Select the number of lags and lag distances sufficient to span a significant portion of the entire site, and choose the lag tolerance to be very close in value to the lag distance itself.
3. Calculate the SV. In most cases, data become less correlated as the distance between them increases. Under these circumstances, the SV values should produce a monotonic increasing function, which approaches a maximal value called the sill. In practice, this may not be the case with SV values that may begin high or jump around as distance increases.
4. Adjust the number of lags and lag tolerances until, generally, a monotonic increasing trend is seen in the SV values. If this cannot be achieved, it may be that a geostatistical approach is not viable or that more complicated trends are occurring than can be modeled. If a visual inspection of the data or knowledge about the dispersion of contamination indicates a direction of correlation, it may be more appropriate to first test this direction.
5. Assuming the omnidirectional SV is reasonable, add another direction to the plot with a smaller tolerance. You may have to adjust the bandwidth and angle tolerance to produce a reasonable SV plot.
6. If the second direction rises slower to the sill or rises to a lower sill, then this is the major direction of anisotropy.
7. If neither direction produces significantly lower spatial correlation, it may be reasonable to assume an isotropic correlation structure.
8. Add a cone structure with direction equal to the major direction plus 90°, and model the SV results in this direction.
9. If the data are isotropic, choose the omnidirectional SV as the major direction.

5.5.3 SV Limitations

The SV model mathematically specifies the spatial variability of the data set, and after its identification, the spatial interpolation weights, which are applied to data points during the grid node calculations, are direct functions of the Kriging model (Chap. 5). In order to determine the estimation value, all measurements within the SV range are assigned weights depending on the distance of neighboring point using the SV. These weights and measurements are then used to calculate the estimation value through Kriging modeling. Useful and definite discussions on the practicalities and limitations of the classicaltheoretical function, which is called SV have been given by Sen (1989) as follows:

1. The classical SV, $\gamma(d)$, for any distance, d, is defined as the half-square difference of two measurements separated by this distance. As d varies from zero to the maximum possible distance within the study area, the relationship of the half-square difference to the separation distance emerges as a theoretical

function, which is called the SV. The sample SV is an estimate of this theoretical function calculated from a finite number, n, of samples. The sample SV can be estimated reliably for small distances when the distribution of sampling points within the region is regular. As the distance increases, the number of data pairs for calculation of SV decreases, which implies less reliable estimation at large distances.

2. In various disciplines of the earth sciences, the sampling positions are irregularly distributed in the region, and therefore, an unbiased estimate of SV is not possible. Some distances occur more frequently than others and accordingly their SV estimates are more reliable than others. Hence, a heterogeneous reliability dominates the sample SV. Consequently, the sample SV may have ups and downs even at small distances. Such a situation gives rise to inconsistencies and/or experimental fluctuations with the classical SV models which are, by definition, nondecreasing functions, i.e., a continuous increase with distance is their main property. In order to give a consistent form to the sample SV, different researchers have used different subjective procedures:

   (a) Journel and Huijbregts (1978) advised grouping of data into distance classes of equal length in order to construct a sample SV. However, the grouping of data pairs into classes causes a smoothing of the sample SV relative to the underlying theoretical SV. If a number of distances fall within a certain class, then the average of half-square differences within this class is taken as the representative half-square difference for the mid-class point. The effect of outliers is partially damped, but not completely smoothed out by the averaging operation.
   (b) To reduce the variability in the sample SV, Myers et al. (1982) grouped the observed distances between samples into variable length classes. The class size is determined such that a constant number of sample pairs fall in each class. The mean values of distances and half-square differences are used for the classes as a representative point of sample SV. Even this procedure resulted in an inconsistent pattern of sample SV (Myers et al. 1982) for some choices of the number, m, of pairs falling within each class. However, it was observed by Myers et al. that choosing $m = 1000$ gave a discernible shape. The choice of constant number of pairs is subjective, and in addition, the averaging procedures smooth out the variability within the experimental SV. As a result the sample SV provides a distorted view of the variable in that it does not provide, for instance, higher-frequency (short wavelength) variations. However, such short wavelength variations, if they exist, are so small that they can be safely ignored.

The above procedures have two basic common properties, namely, predetermination of a constant number of pairs or distinctive class lengths and the arithmetic averaging procedure for half-square differences as well as the distances. The former needs a decision, which in most cases is subjective, whereas the latter can lead to unrepresentative SV values. In classical statistics, only in the case of symmetrically distributed data, the mean value is the best estimation; otherwise, the median

becomes superior. Moreover, the mean value is sensitive to outliers. The following points are important in the interpretation of any sample SV:

1. The SV has the lowest value at the smallest lag distances (d) and increases with distance, leveling off at the sill, which is equivalent to the overall regional variance of the available sample data. It is the total vertical scale of the SV (nugget effect + sum of all component scales). However, linear, logarithmic, and power SVs do not have a sill.
2. The range is the average distance (lag) within which the samples remain spatially dependent and it corresponds to the distance at which the SV values level off. Some SV models do not have a length parameter; e.g., the linear model has a slope instead,
3. The nugget is the SV value at which the model appears to intercept the ordinate. It quantifies the sampling and assaying errors and the short-scale variability (i.e., spatial variation that occurs at distance closer than the sample spacing). It represents two often co-occurring sources of variability:
   (a) All unaccounted for spatial variability at distances smaller than the smallest sampling distance.
   (b) Experimental error is often referred to as human nugget. According to Liebhold et al. (1993), interpretations made from SVs depend on the size of the nugget because the difference between the nugget and the sill (if there is one) represents the proportion of the total sample variance that can be modeled as spatial variability.

5.6 Sample SV

In practice, one is unlikely to get SVs that look like the one shown in Fig. 5.17. Instead, patterns such as those in Fig. 5.19 are more common.

Important practical information in the interpretation and application of any sample SV is to consider only about $d/3$ of the horizontal distance axis values from the origin as reliable.

A digression is taken in this book as for the calculation of sample SVs. Instead of easting- and northing-based SVs, it is also possible to construct SVs based on triple variables. In the following, different triple values are assessed for the SV shapes and interpretations. For instance, in Fig. 5.20, the chloride change with respect to calcium and sodium is shown in 3D and various sample SVs along different directions are presented in Fig. 5.21.

This figure indicates that the change of chloride data with respective independent variables (magnesium and calcium) is of clumped type without leveling effect. It is possible to consider Fig. 5.20 as having two parts, namely, an almost linear trend and fluctuations (drift) around it. In such a case, a neighborhood definition and weight assignments become impossible. Therefore, the ReV is divided into two parts, the residual and the drift. The drift is the weighted average of points within

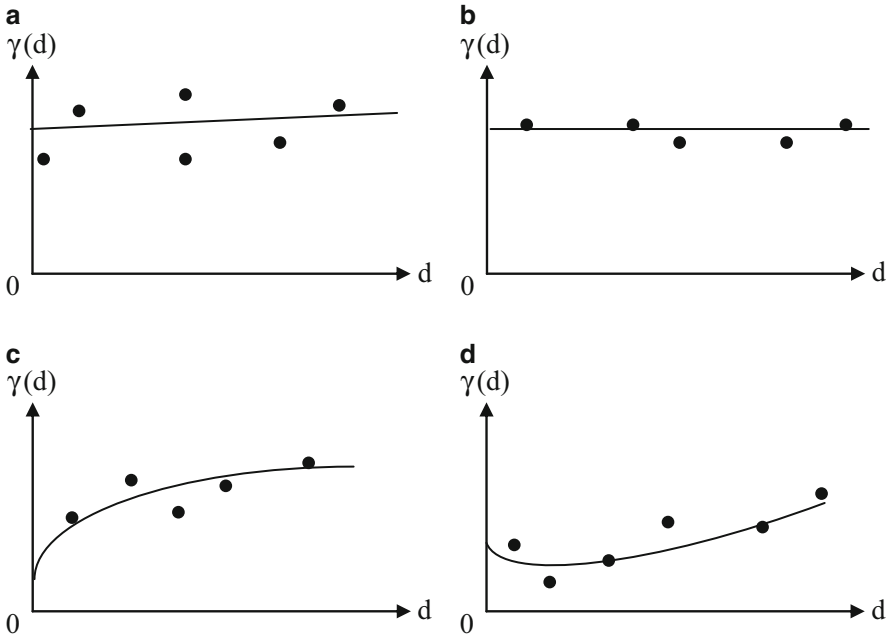


Fig. 5.19 Common forms of sample SVs, (**a**) random, (**b**) uniform, (**c**) clumped with leveling, (**d**) clumped without leveling

the neighborhood around the estimation value. The residual is the difference between the ReV and the drift.

The residual is a stationary ReV in itself and hence allows construction of a SV. However, once again the problem of not being able to define a neighborhood arises. Therefore, an arbitrary neighborhood is chosen from which a drift can be calculated. The calculation includes the points within the assumed neighborhood and a corresponding coefficient for each point, which will be explained in more detail in the Kriging section. The only variable left in the calculation is the SV; however, no SVs exist from which to obtain the SV. Therefore, a reasonable SV is assumed and compared to the resultant residual SV. If the two are the same, then the assumptions made about the neighborhood and SV are correct, and regional estimation can be made. If they differ, then another SV must be used until they become the same. It is possible to identify from a SV the optimum distance after which regional dependence is zero or constant. By definition the SV is half the variance of the difference between all possible points at a constant distance apart. The existence of underlying trend implies the expectation of sample SV similar to ideal case as in Fig. 5.12.

Zero and 30° directional SVs in Fig. 5.21a, b have such a trend, whereas other directional sample SVs along 60° and 90° have more or less random type provided that only $d/3$ of the distance axis variables (0–15) are considered.

It is obvious that practically, there is no nugget effect in these sample SVs, which is rather obvious from Fig. 5.20 where the trend surface does not have discontinuity

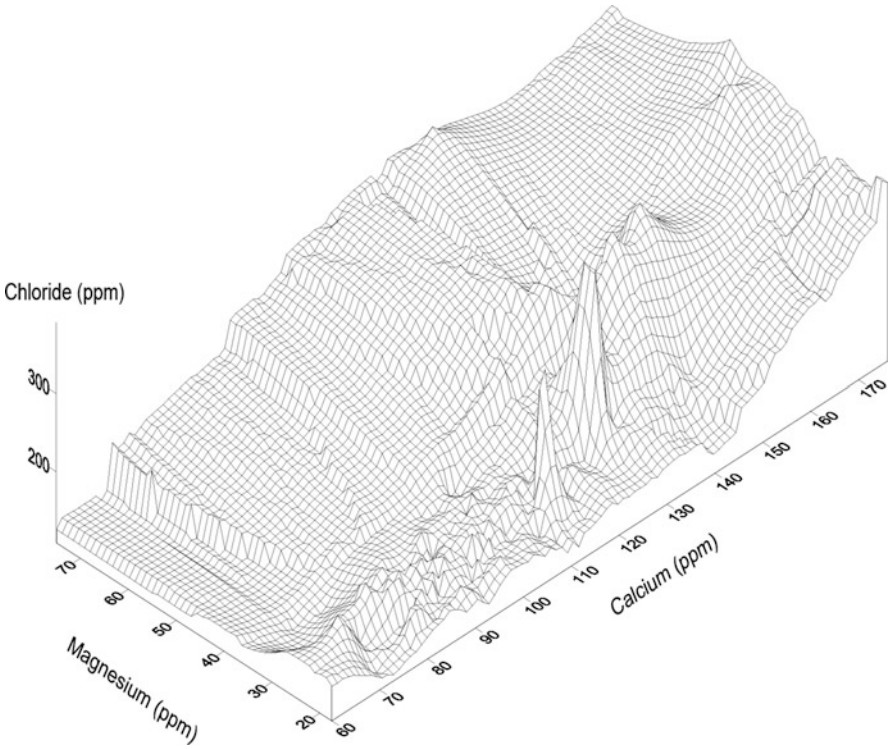


Fig. 5.20 Triple surface of chloride

as in Fig. 5.12. Similarly, 3D representation of bicarbonate variation with calcium and magnesium is presented in Fig. 5.22 with a global (without different directions) sample SV in Fig. 5.23.

Visual inspection of Fig. 5.22 indicates that a trend surface is embedded in the ReV. This trend surface is almost horizontal with a very small angle and accordingly its global sample SV in Fig. 5.23 has almost horizontal sector within $d/3$.

Total dissolved solids (TDS) measurement is an indicator of the water quality variation with respect to calcium and magnesium and its variation is presented in Fig. 5.24 with corresponding global SV in Fig. 5.25.

5.7 Theoretical SV

A useful discussion on the computation of the SV in one or two dimensions has been given by Clark (1979) and Ludwig and Reynolds (1988). The SV as computed from the data will tend to be rather lumpy, and the more irregular the data, the less regular it will appear. Whatever the extend of lumpiness, the graph of $\gamma(d)$ may

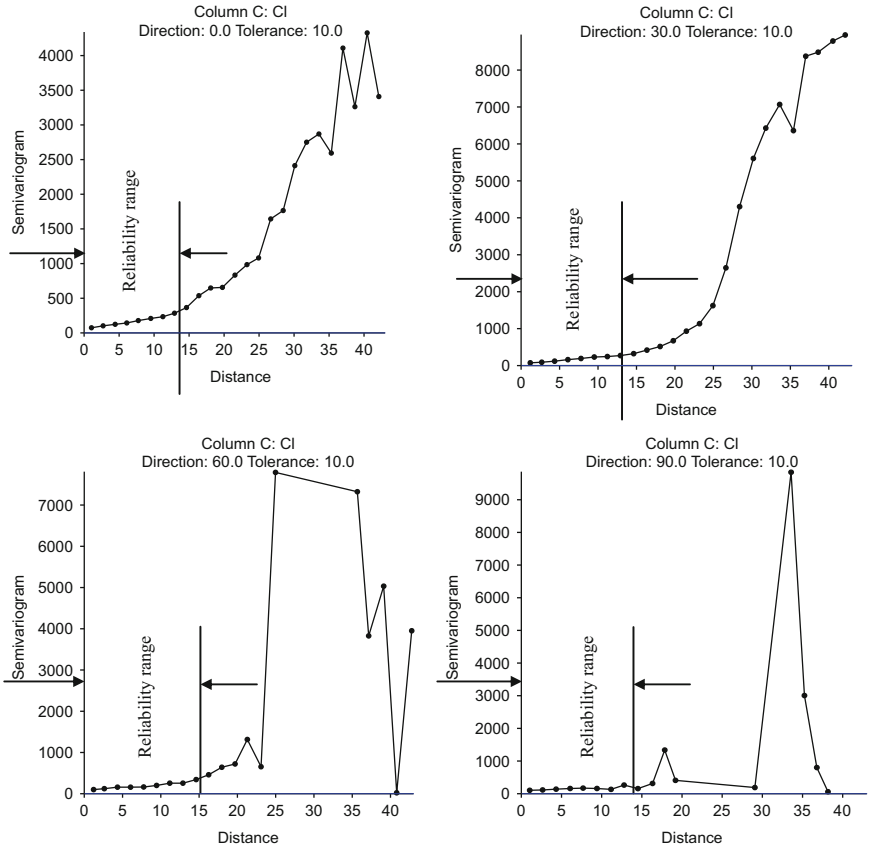


Fig. 5.21 Chloride sample SVs

often be linked to one or other of a small number of theoretical (ideal) and simple curves that relate $\gamma(d)$ to d, which are referred to as the theoretical SVs. These theoretical curves are models of SV which have been defined on theoretical and sample basis. The fitting of the theoretical SV curve, from a set of functions to the experimentally available one derived from real data, has been developed into the art of "structural analysis" discussed in detail by a number of researchers (Journel and Huijbregts 1978; Myers et al. 1982; Ludwig and Reynolds 1988). The main theoretical SV types are linear, spherical, exponential, Gaussian, or cubic types as will be explained later in detail. These functions express rather well qualitatively characteristics of ReV and act as a quantified summary of the structural information, which is then channeled into the estimation procedures of the natural (geological, hydrological, meteorological, atmospheric, etc.) phenomena.

In order to apply Kriging modeling to a ReV, the first step is to obtain sample SV from the available data and then to match this sample SV to a theoretically suitable mathematical function. It describes the relationship between the difference of

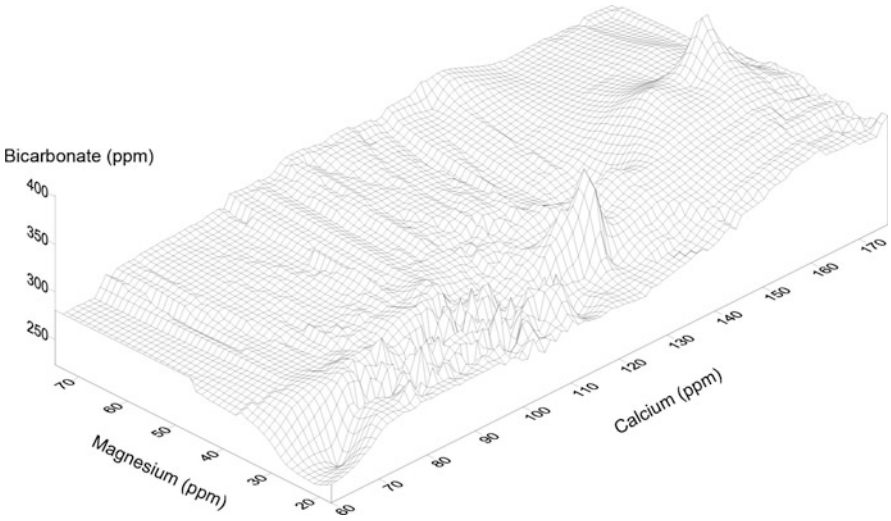


Fig. 5.22 Triple surface of chloride

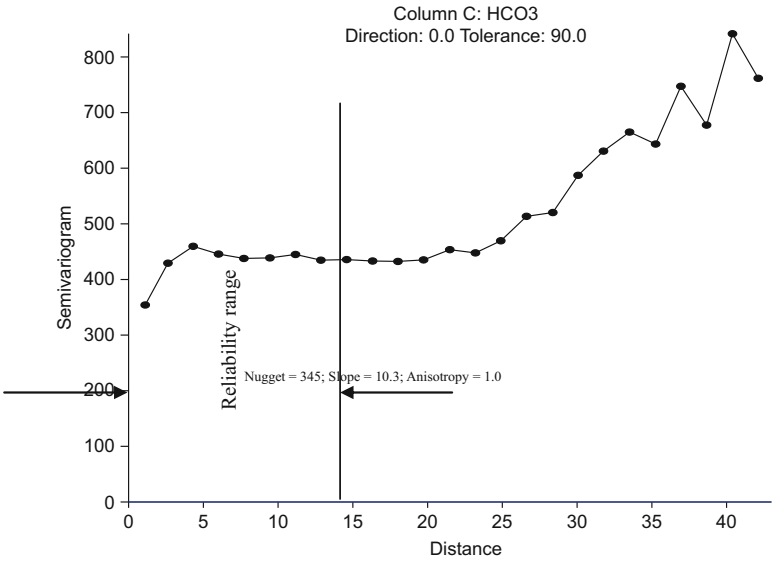


Fig. 5.23 Bicarbonate sample SV

values and distance with a mathematical function. Several different types of functions can be used, each with a different form to the distance function. In addition to its shape or form, the SV model is described by three parameters, namely, nugget, sill, and range. Nugget shows how much variance is observed at a distance of zero. It shows up because there may be variation at distances shorter

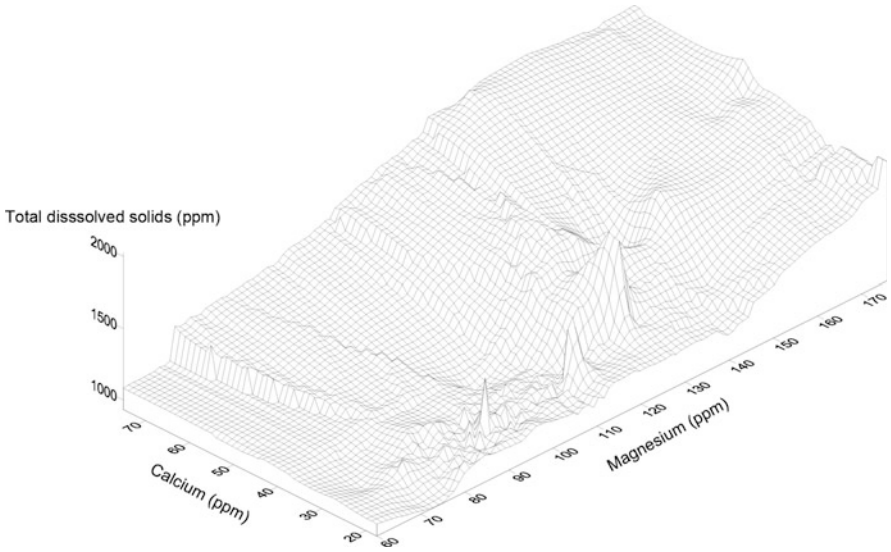


Fig. 5.24 Triple surface of total dissolved solids

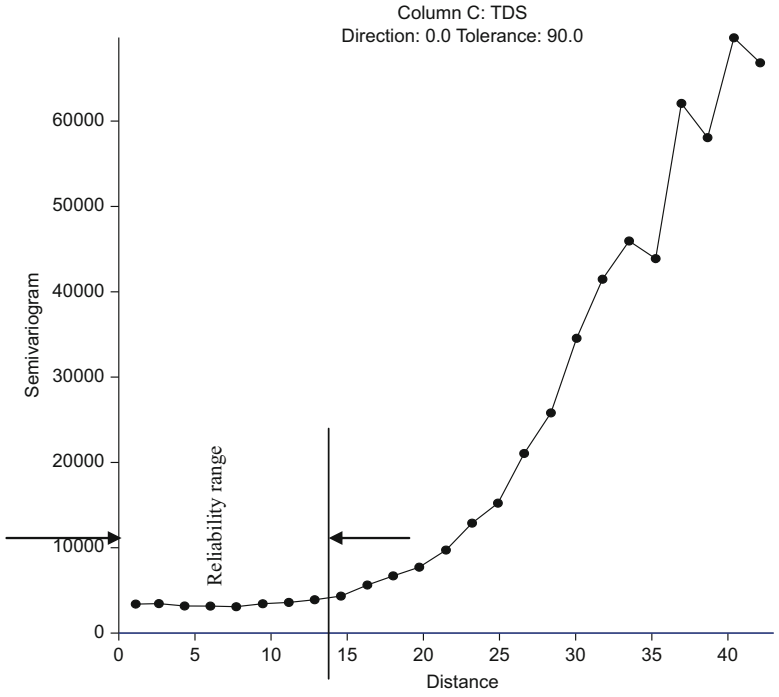


Fig. 5.25 Total dissolved solids SV

than the sample spacing or because there are errors in the measurements. It corresponds to discontinuity feature existence in the ReV. The sill shows how much variation is observed when the SV levels off at large distances. Most SVs become constant at large distances provided that the ReV has spatial dependence without any systematic trend component. Once it is far enough away, there is no relationship in the ReV between the distances of two points. The range shows how far one has to cover distance before the SV levels off to the sill. At distances less than the range, a ReV will be said to be spatially dependent, and beyond the range distance, the ReV has no effect and there is no spatial dependence. Finally, if the directional SVs are very different, one may need to specify an anisotropy parameter. For instance, the NS SV may be different from the EW SV (there is a different range, sill, nugget value, and rather different shape).

The development of an appropriate SV model for a data set requires the understanding and application of advanced statistical concepts and tools, which is the science of SV modeling. In addition, the development of an appropriate SV model for a data set requires knowledge of the tricks, traps, pitfalls, and approximations inherent in fitting a theoretical SV model to real-world data, which is the art of SV modeling. Skills with the science and the art are both necessary for success.

The development of an appropriate SV model requires numerous correct decisions. These decisions can only be properly addressed with an intimate knowledge of the data at hand and a competent understanding of the data genesis (i.e., the underlying processes from which the data are drawn).

Several SV models have been developed to describe the various underlying spatial patterns in data. Examples of isotropic models including spherical, exponential, linear, power, and Gaussian models are used as input in the Kriging ReV estimation process (Chap. 5).

5.7.1 Simple Nugget SV

This corresponds to a random field with no regional dependence structure at all. In such a case, the ReV is represented by its variance, σ^2, only. Its mathematical form is given as

$$\gamma(d) = \begin{cases} \sigma^2 & \text{for} \quad d > 1 \\ 0 & \text{for} \quad d = 0 \end{cases} \tag{5.4}$$

The random field that this SV represents is not continuous (Fig. 5.15). No matter what the distance is (small or large), each ReV is completely independent and different from others. In this case, the spatial analysis methodology is the probability principles only (Fig. 5.16). The special case of this nugget SV occurs when there is not any spatial variability but a uniform ReV value at each point (see Figs. 5.7 and 5.8). The mathematical model is then $\gamma(d) = 0$ for all d values.

5.7.2 *Linear SV*

If the ReV does not have any discontinuity, then its simplest model is given as

$$\gamma(d) = \beta d \tag{5.5}$$

where β is the only model parameter with its meaning of the slope of the straight line (see Fig. 5.26).

A more general form of theoretical linear SV is the mixture of nugget and the linear SVs. It postulates a linear relationship between the cumulative half-square difference and the distance as

$$\gamma_c(d) = \alpha + \beta d \tag{5.6}$$

in which α corresponds to nugget effect and β is the slope of the theoretical SV as in Fig. 5.27.

The unknowns α and β can be solved from a sample SV either by substitution of two most characteristic points that should be preserved in the theoretical SV or through a least-squares regression line fitting. The first approach is rather subjective but may represent expert in a better way. On the other hand, according to theory, the expected value of the sample variance is the average value of the SV between all possible pairs of sample locations, which yields one equation (Barnes 1991). The

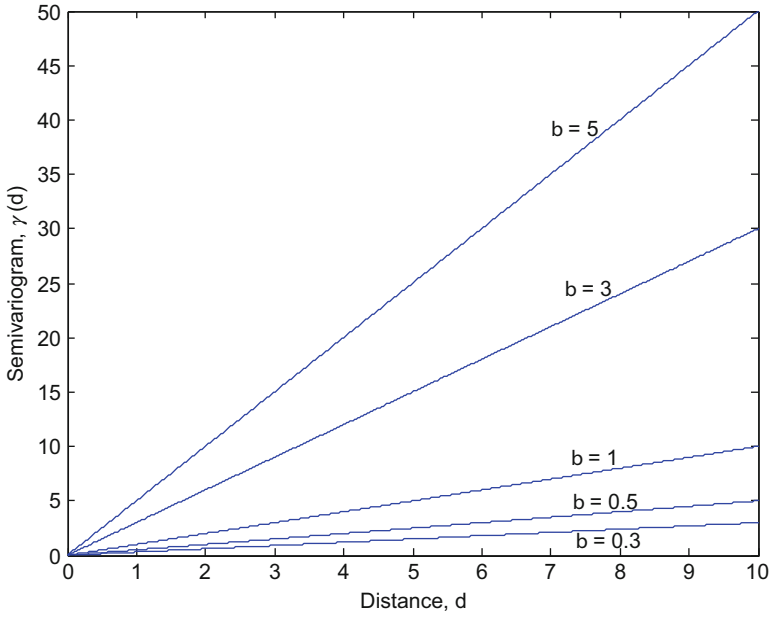


Fig. 5.26 Linear SV model

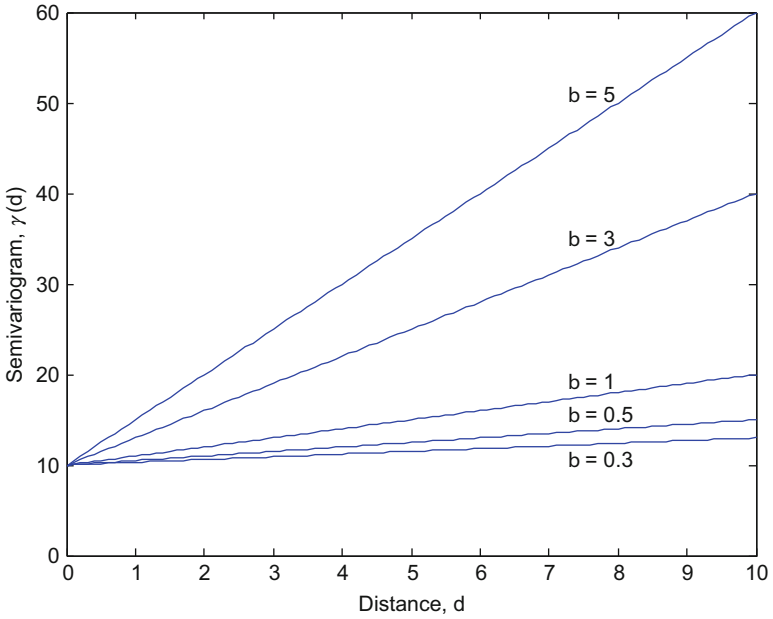


Fig. 5.27 General linear SV

second equation is generated by equating the experimental SV for nearest neighbors to the modeled SV. Thus, it is possible to write

$$\gamma_s(d) = \alpha + \beta\overline{d}$$
$$\gamma_n(d) = \alpha + \beta\overline{d}_n$$

where $\overline{d}_n$ = average distance to the nearest neighbor, $\overline{d}$ = average inter-sample separation distance, $\gamma_n(d)$ = one half the average square difference between nearest neighbors, and $\gamma_s(d)$ = sample variance. By solving the two equations for the two unknown parameters and checking for unreasonable values, one can get the final formulae used in the Kriging technology:

$$\alpha = \max\left[\frac{\overline{d}\gamma_n(d) - \overline{d}_n\gamma_s(d)}{\overline{d} - \overline{d}_n},\ 0\right]$$

and

$$\beta = \max\left[\frac{\gamma_s(d) - \gamma_n(d)}{\overline{d} - \overline{d}_n},\ 0\right]$$

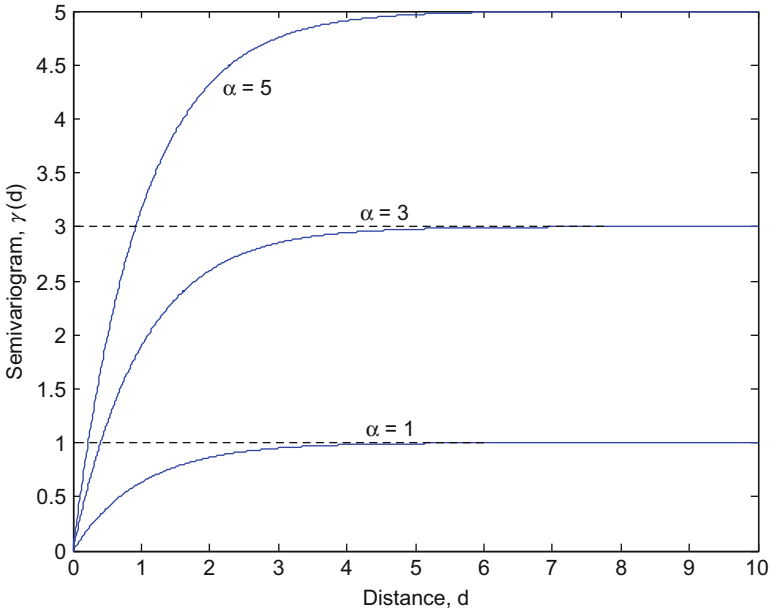


Fig. 5.28 Exponential SV model

5.7.3 *Exponential SV*

The mathematical expression of theoretical exponential SV model is given as (Cressie 1993)

$$\gamma(d) = \alpha\left(1 - e^{-\beta d}\right) \tag{5.7}$$

where the two model parameters are $\alpha > 0$ and $\beta > 0$. This model is used commonly in hydrological studies. Figure 5.28 shows exponential models with different variance values and $\beta = 1$. It does not have a nugget value, hence represent continuous ReV, which are stationary, because at large distances, exponential SV approaches a horizontal asymptote.

5.7.4 *Gaussian SV*

Its general expression is given in the following mathematical form (Pannatier 1996):

$$\gamma(d) = \alpha\left(1 - e^{-\beta d^2}\right) \tag{5.8}$$

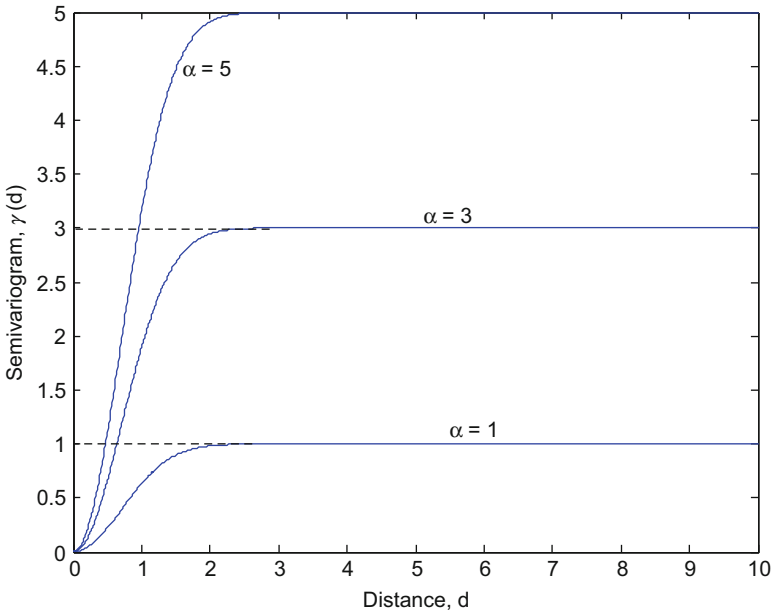


Fig. 5.29 Gaussian SV model

where $\alpha > 0$ and $\beta > 0$ are the two parameters, and according to their values, it takes different forms. Figure 5.29 shows a set of Gaussian SV with different α values and $\beta = 1$. Since at large distances there is a horizontal portion, the representative ReV is also stationary.

For the initial portion, i.e., at small distances, the Gaussian model appears as proportional to d^2, which implies that the ReV is smooth enough to be differentiable, i.e., the slope of SV tends to a well-defined limit as the distance between the two points.

5.7.5 *Quadratic SV*

Its mathematical expression is given by Alfaro (1980) as

$$\gamma(d) = \begin{cases} \alpha d(2-d) & d < 1 \\ \alpha & d \geq 1 \end{cases} \tag{5.9}$$

where the single model parameter $\alpha > 0$. Its shape is given in Fig. 5.30 for different α values. As it increases the sill value decreases, which means that the fluctuations become smaller. It represents stationary ReV with non-differentiable properties.

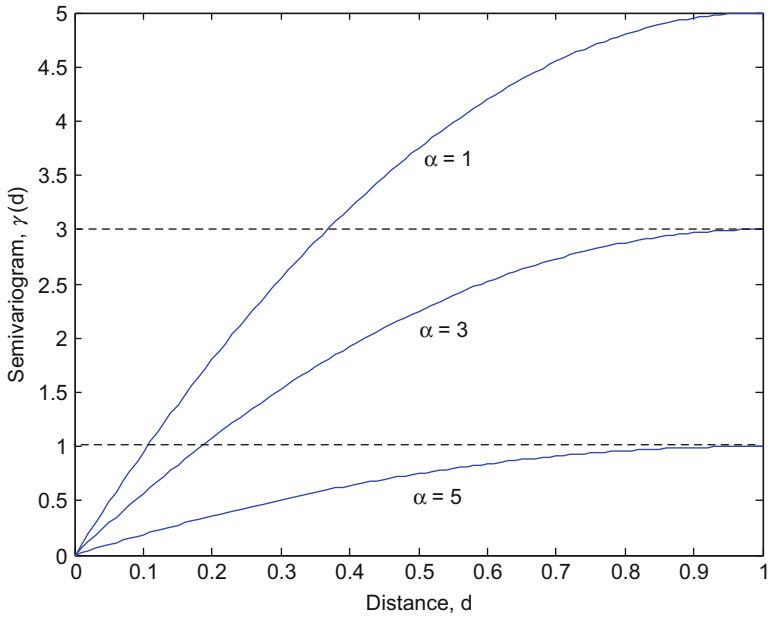


Fig. 5.30 Quadratic SV model

5.7.6 *Rational Quadratic SV*

Cressie (1991) provided this SV for use with the following expression:

$$\gamma(d) = \alpha\left(\frac{d^2}{1+d^2}\right) \tag{5.10}$$

where $\alpha > 0$ is the model parameter. It has similar behavior to Gaussian model at small distances which implies that the ReV is differentiable and rather continuous. An increase in α parameter value causes increase in the sill level, which further shows that the fluctuations in the ReV from its average level become bigger (see Fig. 5.31).

5.7.7 *Power SV*

Its general form is presented by Pannatier (1996) as

$$\gamma(d) = \alpha|d^m| \tag{5.11}$$

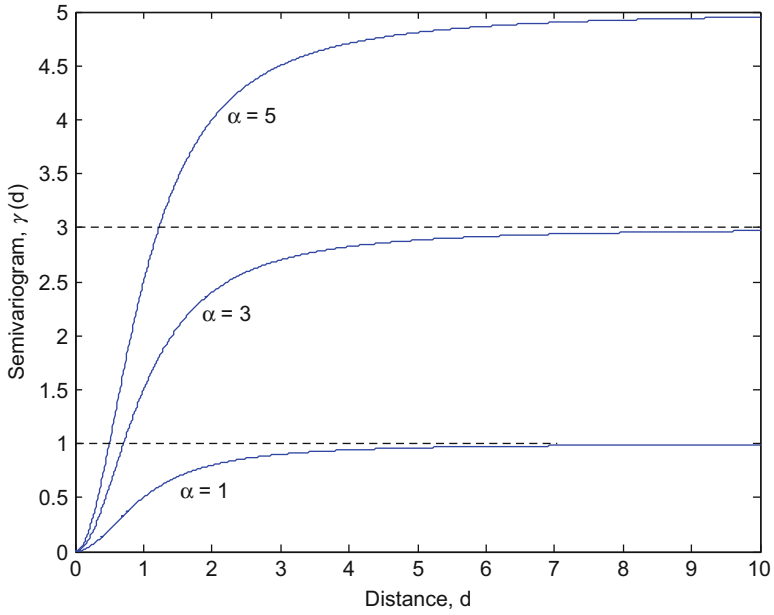


Fig. 5.31 Rational quadratic SV model

where $\alpha > 0$ and $0 < m < 2$ are model parameters. Convex and concave forms result, respectively, in $0 < m < 1$ and $1 < m < 2$ as in Fig. 5.32. Besides, it reduces to a linear model (Eq. 5.5) for $m = 1$. Depending on whether $m > 1$ ($m < 1$), the SV represents a differentiable (non-differentiable) ReV. Additionally, none of the theoretical SVs approach to a sill value, and therefore, the represented ReV does not have stationary property. These SVs describe the same appearance of realizations at every scale, and therefore, they are referred to as self-similar ReV. This is because they appear as a straight line on a double-logarithmic paper with different slopes.

The main difference of this model from the others is that it has a nonzero value for zero distance, i.e., it has a nugget effect. The forms of different cumulative SVs resulting from Eq. 5.11 are shown in Fig. 5.32.

5.7.8 Wave (Hole Effect) SV

Cressie (1991) gave the general mathematical expression of this model as

$$\gamma(d) = \alpha\left(1 - \frac{\sin d}{d}\right) \tag{5.12}$$

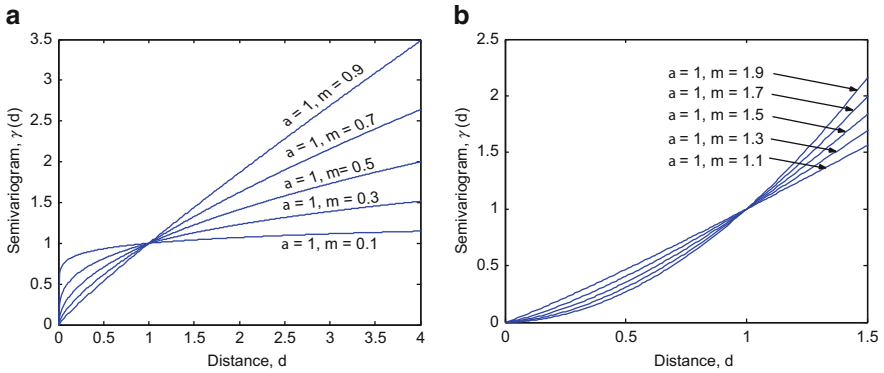


Fig. 5.32 Power SV model, **a** concave, **b** convex

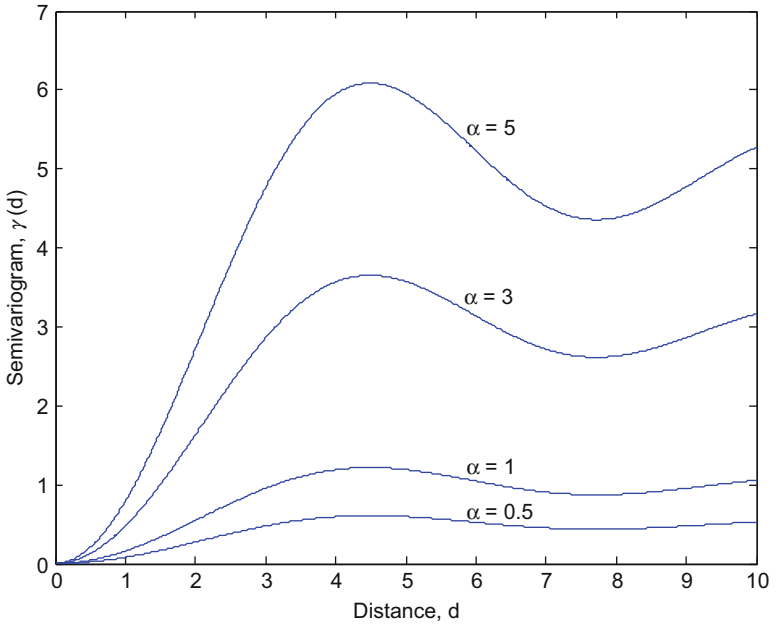


Fig. 5.33 Wave (hole effect) SV model

where again $\alpha > 0$ is the model parameter (see Fig. 5.33). This model describes ReV with excursions above the mean which tend to be compensated by excursions below the mean. It exhibits linear behavior at small distances which implies that the corresponding ReV are continuous but not differentiable. They are less smooth than the realization of a random field.

5.7.9 Spherical SV

Pannatier (1996) presented the general expression as

$$\gamma(d) = \begin{Bmatrix} \alpha d\left(1.5 - 0.5d^2\right) & d < 1 \\ \alpha & d \geq 1 \end{Bmatrix} \tag{5.13a}$$

with model parameter $\alpha > 0$. At small distances it is proportional with distance and therefore its representative ReV is non-differentiable but continuous. At large distance the existence of horizontal level (sill) implies that the ReV is stationary. Its shapes are presented in Fig. 5.34.

5.7.10 Logarithmic SV

It is a single parameter model with mathematical expression as

$$\gamma(d) = \alpha \ln d \quad d > 0 \tag{5.13b}$$

It appears as a straight line on a semilogarithmic paper (logarithmic distances) and α represents the slope of this line. Various logarithmic SVs are presented in

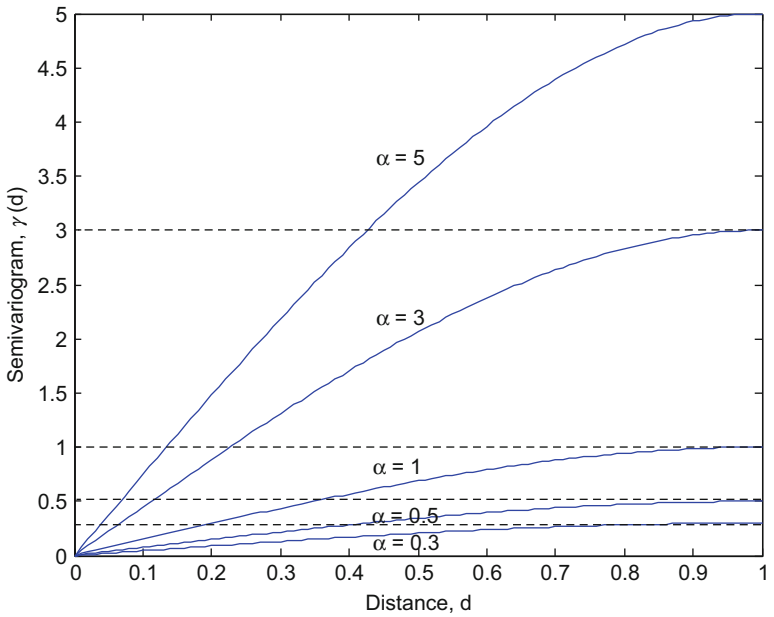


Fig. 5.34 Spherical SV model

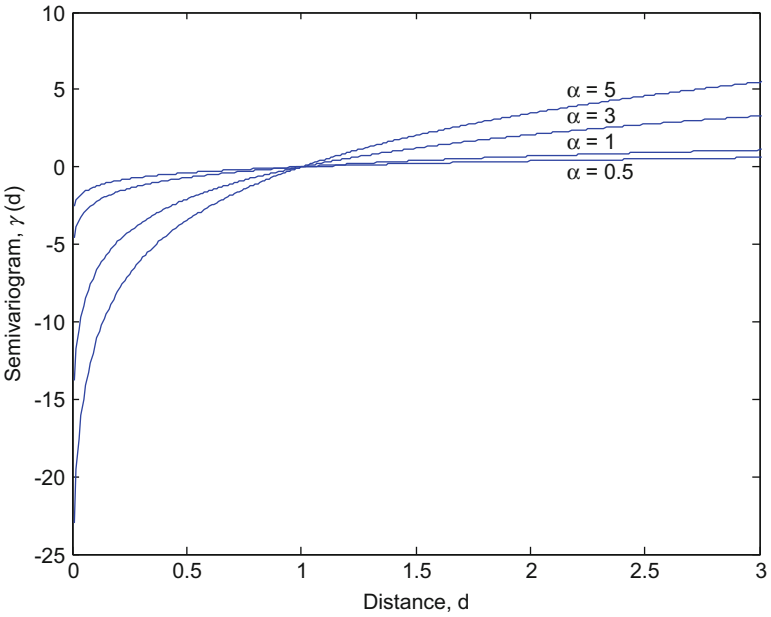


Fig. 5.35 Logarithmic SV model

Fig. 5.35. Since there is no sill, the corresponding ReV is nonstationary. It can be used only for integrals over finite volumes and cannot be used directly with point data values of the ReV (Kitanidis 1997).

5.8 Cumulative Semivariogram (CSV)

The CSV method proposed by Sen (1989) as an alternative to the classical SV technique of Matheron (1965) has various advantages over any conventional procedure in depicting the regional variability, and hence spatial dependence structure CSV is defined similar to the SV; the only difference is that successive cumulative summations are adopted. CSV has the same advantages of the SV and it provides an objective way in construction of spatial variable regional dependence behavior. Furthermore, standardization of CSV provides a basis in identifying regional stationary stochastic models (Şen 1992). The CSV is a graph that shows the variation of successive half-square difference summations with distance. Hence, a nondecreasing CSV function is obtained which exhibits various significant clues about the regional behavior of the ReV. The CSV provides a measure of spatial dependence. The CSV can be obtained from a given set of ReV data by executing the following steps:

1. Calculate the distance $d_{i,j}$, ($i \neq j = 1, 2, \ldots, m$) between every possible pair of sparse measurement sites. For instance, if the number of sample sites is n, then there are $m = n(n-1)/2$ distance values.
2. For each distance, $d_{i,j}$, calculate the corresponding half-square differences, $D_{i,j}$, of the ReV data. For instance, if the ReV has values of Z_i and Z_j at two distinct sites at distance $d_{i,j}$ apart, then the half-square difference is

$$D_{i,j} = \frac{1}{2}\left(Z_i - Z_j\right)^2 \tag{5.14a}$$

3. Take the successive summation of the half-square differences starting from the smallest distance to the largest in order. This procedure will yield a nondecreasing function as

$$\gamma\left(d_{i,j}\right) = \sum_{i=1}^{m}\sum_{i=1}^{m} D_{i,j} \tag{5.14b}$$

where $\gamma(d_{i,j})$ represents CSV value at distance $d_{i,j}$.
4. Plot $\gamma(d_{i,j})$ values versus the corresponding distance $d_{i,j}$. The result will appear similar to the representative CSV functions as in Fig. 5.36. The sample CSV functions are free of subjectivity because no a priori selection of distance classes is involved in contrast to the analysis as suggested by Perrie and Toulany (1989) in which the distance axis is divided into subjective intervals, and subsequently, averages are taken within individual intervals which are regarded as the representative value for this interval.

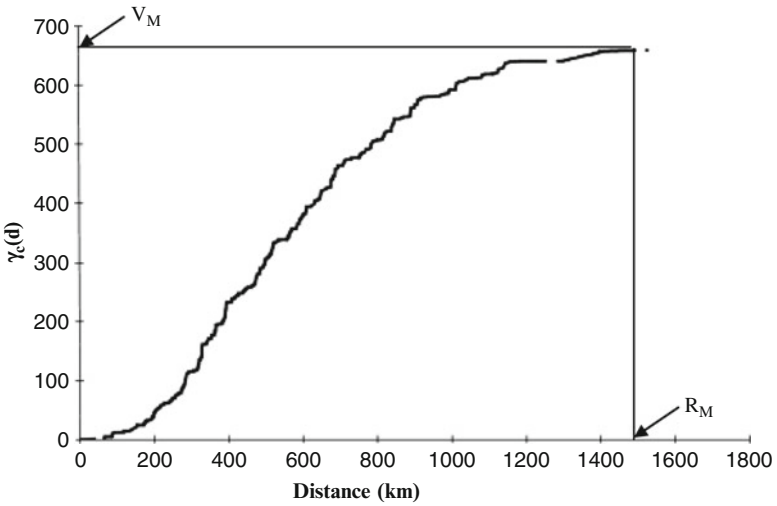


Fig. 5.36 Representative CSV

In his discussion on the practical difficulties of classical SV, Şen (1989) has noticed the reliability of SV estimations at small distances with regular distribution of sampling points within the region. However, in various disciplines of geological sciences, the sampling positions are irregularly distributed in the region (such as in the earthquake epicenters) where some distances occur more frequently than others; thus, a heterogeneous reliability dominates the sample SV. In order to overcome these shortcomings, he developed CSV technique on the basis of the ReV theory. This technique possesses all of the objective properties of the classical SVs and in addition it helps to identify the hidden and local dependencies within a region. It is defined as the successive summation of half-square differences, which are ranked according to the ascending order of distances extracted from all possible pairs of sample locations within a region. Mathematically, CSV is expressed as

$$\gamma_c(d_k) = \sum_{i=1}^{k} d\left[d^i\right] (k = 1, 2, \ldots, m) \tag{5.15}$$

where $\gamma_c(d_k)$ is the value of the kth ordered distance CSV value and superscript i indicates the rank. The procedure of sample CSV calculations and its model form and equations have been given by Şen (1989) as in Fig. II.12. These models are counterpart to those of classical SV models, but with different interpretations of the model parameters. The attributes and advantages of CSV can be summarized as follows:

1. The CSV is a nondecreasing function; however, there may be local flat portions, implying constancy of the regionalized variables at certain distance, i.e., the same values have been observed at two locations h apart.
2. The slope of the theoretical CSV at any distance is an indicator of the dependence between pairs of regionalized variables separated by that distance.
3. The sample CSV reflects even smaller dependencies between data pairs, which are not possible to detect with classical SV due to the averaging procedure.
4. The sample CSV is straightforward in applications and free of subjectivity because there is no need for a priori selection of distance classes. In fact, the real distances are employed in the construction of the sample CSV rather than class midpoint distance.
5. The CSV model may be used for irregularly distributed sample positions within the study region.
6. The underlying model for any regionalized variable can be detected by plotting the cumulative half-square differences versus distances on arithmetic-semilogarithmic or double logarithmic. Appearance of sample CSV points on any one of these papers as a straight line confirms the type of model. Such an opportunity is missing in the samples of classical SV.
7. Model parameter estimates are obtained from slope and intercept values of the straight line.

8. Any classical SV model has a theoretical CSV counterpart which can be obtained through an integration operation.
9. These characteristics and advantages make the CSV attractive for practical applications.

5.8.1 Sample CSV

The CSV proposed in the previous section is applied to the transmissivity, total dissolved solids, and piezometric level records in the Wasia sandstone aquifer in the eastern part of the Kingdom of Saudi Arabia (Şen 1989). A complete hydrogeological study of this area has been performed recently by Subyani (1987).

GAMA3 software developed for computing the classical SV by Journel and Huijbregts (1978) has been applied to groundwater variables such as transmissivity, piezometric level, and total dissolved solids from the Wasia sandstone aquifer. The resulting sample SV and sample CSV plots are presented in Figs. 5.37, 5.38, and 5.39. It is clear from these figures that the half-square difference points are scattered in such a way that it is not possible to distinguish a clear pattern in the sample SVs, which suffer from fluctuations even at small distances. Comparisons of the sample SVs in these figures with the sample CSVs indicate that the latter are more orderly and have distinctive nondecreasing patterns.

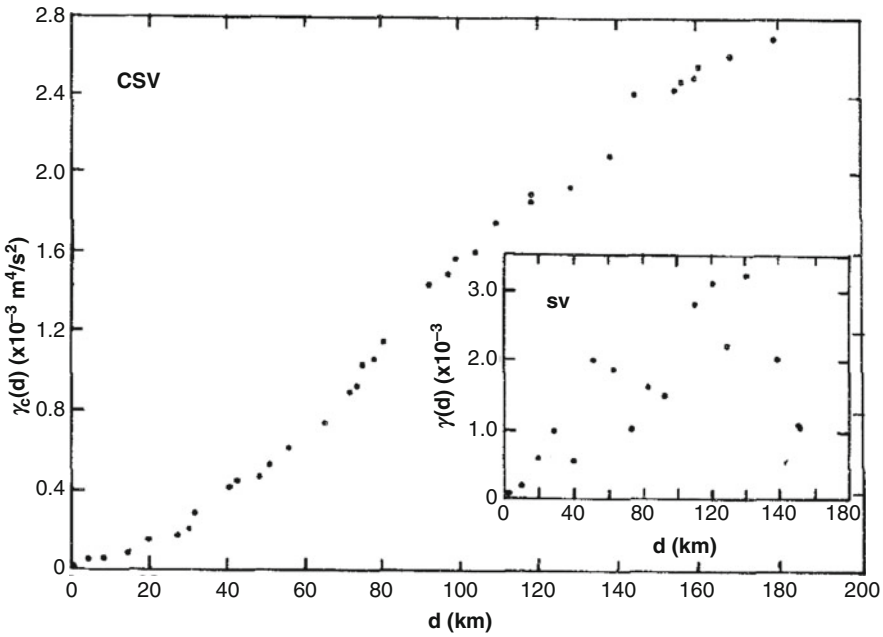


Fig. 5.37 Sample CSV for Wasia sandstone transmissivity

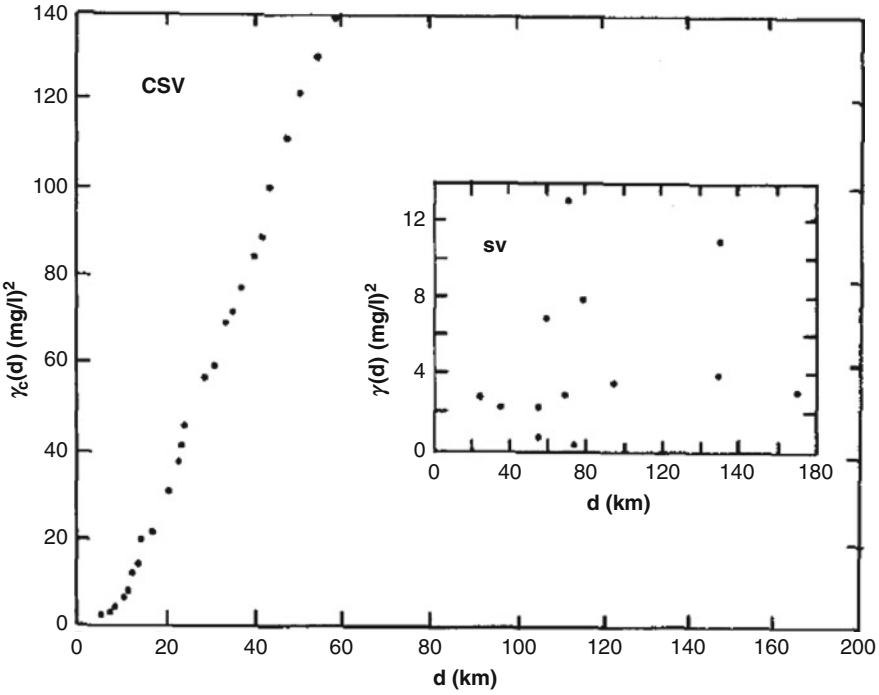


Fig. 5.38 Sample CSV for Wasia sandstone dissolved solids

A sample CSV often yields more or less a straight line for large distances, which corresponds to the sill concept in the classical SV. Furthermore, the sample CSV starts as a curve with different curvatures over some distance domain before it becomes almost a straight line. The length of the distance domain over which the sample CSV occurs as a curve is a counterpart of the range in the classical SV. Hence, it is straightforward to determine the range from the sample CSV. The piezometric level sample CSV in Fig. 5.39 shows an initial range portion, which has zero half-square differences for about 10 km. Such a portion implies physically that the piezometric level does not change significantly within distances less than 60 km. In fact, the Wasia aquifer has remained free of any tectonic movements, it is extensive, and the recharge is negligible, but it is discharged by local well groups which are situated at large distances from each other (Powers et al. 1966).

5.8.2 Theoretical CSV Models

In order to be able to apply Kriging estimation techniques to a ReV, a functional relationship must be established between the distance and the measure of regional dependence which, herein, is the CSV. These models must be nondecreasing

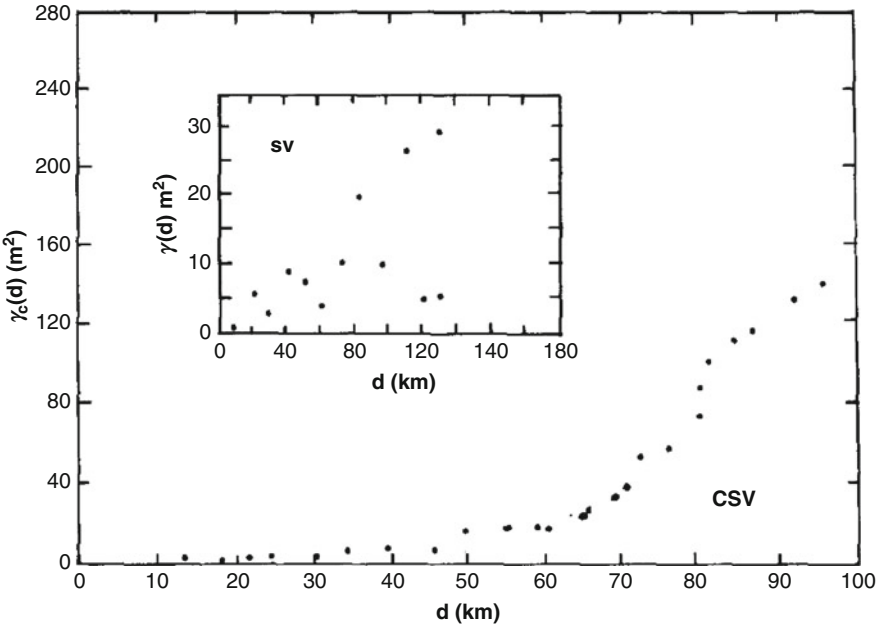


Fig. 5.39 Sample CSV for Wasia sandstone piezometric level

functions. Although numerous functions have this property, in practice restricting them to a few simple ones is desirable. By considering the basic definitions of both the classical and CSVs, they may be related through an integration as

$$\gamma_c(d) = \int_0^d \gamma(u)du \tag{5.16}$$

or through differentiation as

$$\gamma(d) = \left.\frac{d\gamma_c(u)}{du}\right|_{u=d} \tag{5.17}$$

Therefore, a CSV counterpart may be found for any given classical SV using Eq. 5.16. Furthermore, Eq. 5.17 indicates that the theoretical classical SV value at any distance is equal to the slope of the theoretical CSV at the same distance. In the following, models which have been used previously for SVs by many researchers will be assessed from the CSV point of view.

5.8.2.1 Linear Model

The linear CSV model signifies that the phenomenon has regional independence and it evolves in accords with an independent (white noise) process. Hence, in an

objective analysis of such a phenomenon, one needs to know only the mean and standard deviation, and consequently, the prediction of any grid point value can be withdrawn from an independent process because the radius of influence is equal to zero. This is the most extreme case, which rarely happens in earth sciences domains. This model postulates a linear relationship between the cumulative half-square difference and the distance as

$$\gamma_c(d) = \alpha + \beta d \tag{5.18}$$

in which α and β are the model parameters (Fig. 5.40a). The sample CSV of the regionalized variable that abides by this model will appear as a straight line on arithmetic paper.

In fact, α is the intercept on the CSV axis and β is the slope of this straight line. This slope corresponds to the sill value in the classical SV which represents a pure nugget effect (Sen 1989). Furthermore, β represents exactly the variance of the underlying random field. Hence, the smaller the slope of the straight line, the smaller the random fluctuation in the ReV. If the slope is equal to zero, theoretically, this indicates a complete deterministic uniform variation in the ReV. The sample CSV scatter diagram and the fitted regression line to pH values based on the measurements at 71 sample locations gave the form in Fig. 5.41. The study area is Umm Er Radhuma Limestone aquifer in the eastern province of the Kingdom of Saudi Arabia.

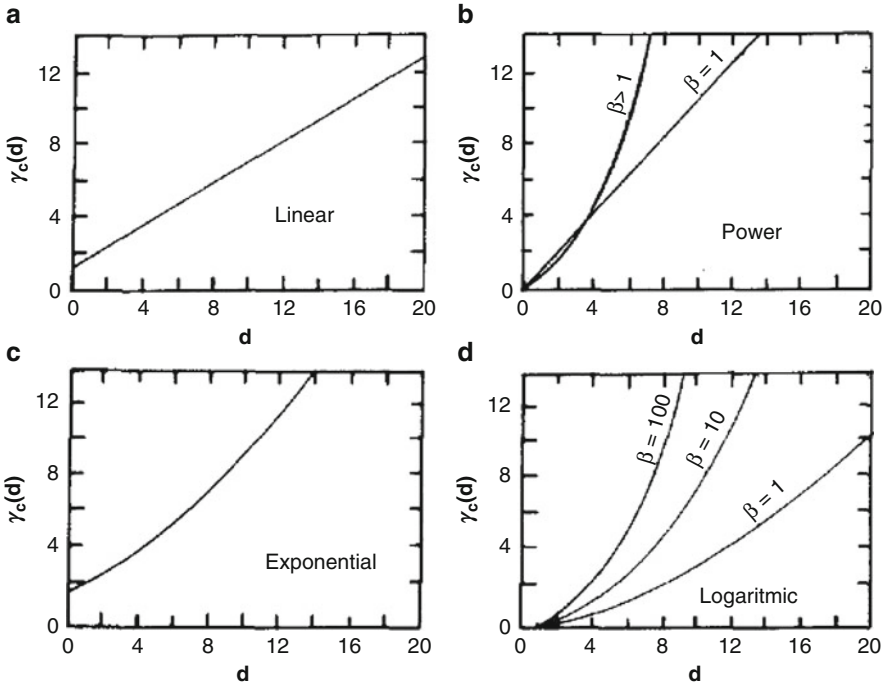


Fig. 5.40 Theoretical CSVs

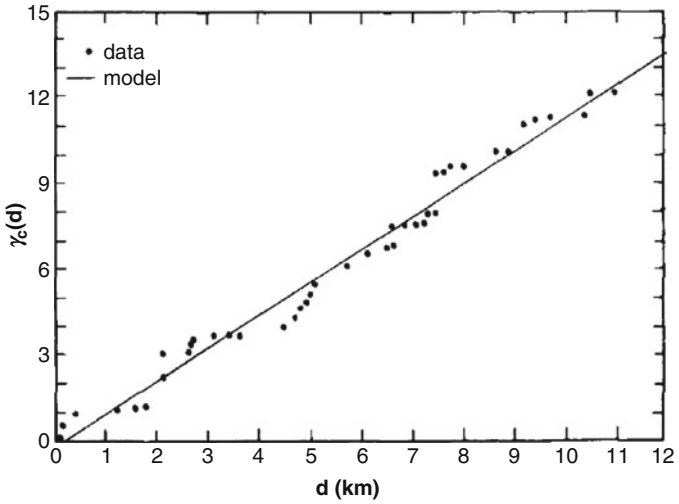


Fig. 5.41 Sample CSV for pH in Umm Er Radhuma aquifer

$$\gamma_c(d) = -0.213 + 1.144d$$

from which the parameter estimates are $\alpha = -0.213$ and $\beta = 1.144$. The hydrochemical data were presented by Sen and Al-Dakheel (1985) for major anions and cations.

5.8.2.2 Power Model

This is a two-parameter model which yields a set of different shapes for the theoretical CSV (Fig. 5.40b). The mathematical expression for this model is

$$\gamma_c(d) = \alpha d^{\beta} \tag{5.19}$$

in which α is the scale parameter and β is the shape parameter. Because $0 < \beta < 2$ for a theoretical SV from a power family (Journel and Huijbregts,1978, p. 165), parameter β for the theoretical CSV in Eq. 5.19 is restricted to the range $1 < \beta < 3$. The derivative of Eq. 5.19 yields also a power form for the classical SV. Obviously, the use of a double-logarithmic paper facilitates parameter estimation. Sulfate concentrations in the Umm Er Radhuma aquifer groundwater show on a double-logarithmic paper a more or less straight-line pattern (Fig. 5.42).

The mathematical expression of this straight line by the regression technique can be found as

$$\log\gamma_c(d) = 0.46 + 0.841d$$

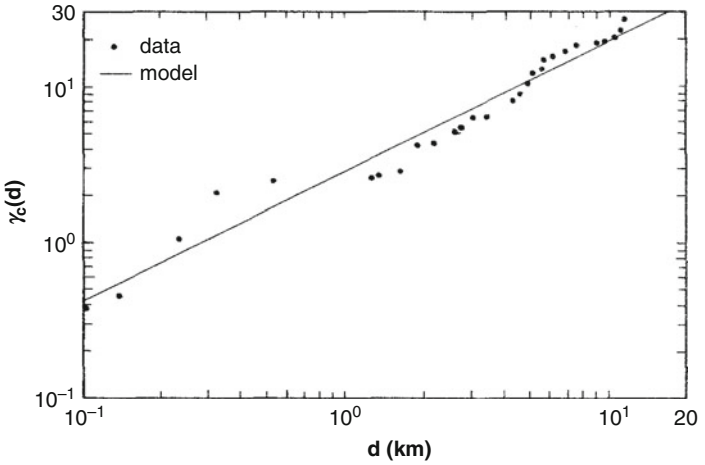


Fig. 5.42 Sample CSV for sulfate in Umm Er Radhuma aquifer

Hence, parameter estimates are $\log \alpha = 0.46$ or $\alpha = -2.88$ and $\beta = 0.84$. The original form of this model prior to transformation can be written as $\gamma_c(d) = 2.88\mathrm{d}^{0.84}$.

5.8.2.3 Exponential CSV

The general form of this CSV can be written as

$$\gamma_c(d) = \alpha e^{\beta d} \tag{5.20}$$

where α and β are scale and shape parameters, respectively. The main difference of this model from the others is that it has a nonzero value for zero distance, i.e., it has a nugget effect. Forms of different CSVs resulting from Eq. 5.20 are shown (Fig. 5.40c). The sample CSV can be checked for concordance with this model by plotting $\log\gamma_c(d)$ vs. d on semilogarithmic paper. If the sample points appear as a straight line, the exponential model is the generating mechanism of the regional variability within the regionalized variable. The slope of this line directly yields an estimate of β, whereas the intercept on the $\gamma_c(d)$ axis leads to an estimate of β. This model does not have a unique classical SV which has appeared in the geostatistical literature. The sample CSV for bicarbonate concentrations in the Umm Er Radhuma aquifer appears as a straight line on semilogarithmic paper (Fig. 5.43). The appearance of this straight line implies that the convenient model for bicarbonate concentrations for this aquifer is of exponential type. The regression line of this scatter diagram is

$$\log\gamma_c(d) = -0.86 + 0.079d,$$

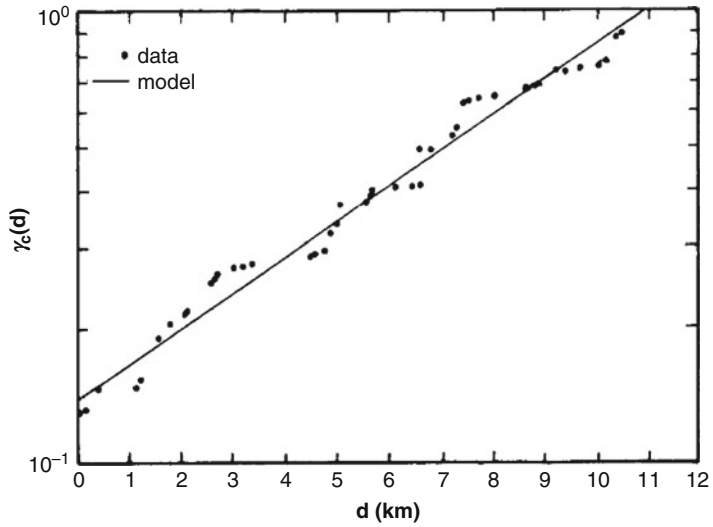


Fig. 5.43 Sample CSV for bicarbonate in Umm Er Radhuma aquifer

and, correspondingly, the model parameter estimates are $\log\gamma_c(d) = -0.86$ or $\alpha = 0.14$ and $\beta = 0.079$. Hence the original form of the model can be written as $\gamma_c(d) = 0.14e^{0.079d}$.

5.8.2.4 Logarithmic CSV

The mathematical expression of this model can be written as

$$\gamma_c(d) = \begin{cases} \alpha + \beta \log d & \text{for } d > 1 \\ 0 & \text{for } d < 1 \end{cases} \tag{5.21}$$

in which α and β are two model parameters. This model differs from the exponential one in that it has an intercept on the distance axis similar to the sample CSV for piezometric level (Fig. 5.39). Different forms of the logarithmic models are presented (Fig. 5.40d). The model can be depicted from a sample CSV plotted on semilogarithmic paper as $\gamma_c(d)$ vs. log d. If the sample points appear as a straight line, the validity of the logarithmic model is confirmed. The slope of this straight line is equal to β, and the cumulative half-square difference corresponding to $d = 1$ yields the estimate of a. Such a model is similar to what is referred to in the classical SV terminology as the De Wijsian model (DeWijs 1972).

Other models for the CSV can be constructed from classical CSV models through Eq. 5.16. For instance, the exponential model of the classical SV, which is

$$\gamma(d) = \alpha\left[1 - e^{-\beta d}\right], \tag{5.22}$$

corresponds to a CSV model, which is

$$\gamma_c(d) = \alpha\left[d - \frac{1}{\beta}e^{-\beta d}\right] \tag{5.23}$$

in which α and β are model parameters. A close inspection of Eq. 5.23 indicates that for large distances, $(1/\beta)\exp(-\beta d) \approx 0$; consequently, at large distances this model appears as a straight line (on arithmetic paper) whose slope is an estimate of α. In addition, this model has an intercept value, $\gamma_c(0)$, which is equal to α/β. Provided that α is known from the slope at large distances, this ratio yields the estimate of β. These α and β values are the parameters of the classical exponential SV model. This last example shows that the CSV method may help to estimate the parameters of the classical SV by simple graphical procedures.

5.8.2.5 Gaussian CSV

The Gaussian classical SV corresponds to the CSV model:

$$\gamma_c(d) = \alpha\left[d - \sqrt{\frac{2\pi}{\beta}}\phi(d,\beta)\right] \tag{5.24}$$

where $\phi(d, \beta)$ is the area under the normal probability density function (with zero mean and variance $1/\beta$) from 0 to d. Obviously, α can be estimated as the slope of this straight line. That is,

$$\phi(d,\beta) = \sqrt{\frac{\beta}{2\pi}\int_0^d \exp\left(-\frac{d^2\beta}{2}\right)dd} \tag{5.25}$$

Last but not the least, it is also possible to have different types of CSV models along different directions for the same earth sciences phenomenon. In such a situation, there is structural heterogeneity within the phenomenon.

Various theoretical CSV models are fitted to the sample CSV by simple least-squares technique. A weighted or generalized least-squares approach would probably be preferable because the sample CSV values are correlated and do not have equal variance. Future researches should be directed toward how to implement a weighted or generalized least-squares approach in particular, what should the weights be, and how strong are the correlations between neighboring CSV values?

5.9 Point Cumulative Semivariogram (PCSV)

The PCSV function is proposed by Şen and Habib (1998) as the CSV calculated for a single point (site). The PCSV identifies the spatial variability of earth sciences ReV around a single site rather than the whole region. It presents the regional effect of all the sites within the study area on a particular site. Consequently, the number of PCSVs is equal to the number of available sites. Each PCSV provides a basis for nearby station variability interpretations, and their mutual comparisons at different sites lead to invaluable information for describing the heterogeneity of the ReV in an area. The treatment of available spatial data at n sites according to the following steps leads to a sample PCSV for a particular site:

1. Calculate the arithmetic average, $\overline{Z}$, and standard deviation, Z_s, from the available data. Standardize the data according to the following formulation:

$$z_i = \frac{Z_i - \overline{Z}}{Z_S} \tag{5.26}$$

2. Calculate distances between the desired site and the remaining sites. If there are m sites, the number of distances is $n-1$, d_i ($i = 1, 2, . ., n–1$).
3. For each pair calculate the square differences as $(z_c–z_i)^2$ where z_c and z_i are the ReV at the concerned and ith sites, respectively. Consequently, there are $(n-1)$ square differences.
4. Rank the distances in ascending order and plot distances d_i versus corresponding successive cumulative sums of half-square differences. Hence, a nondecreasing function is obtained similar to Fig. 5.36, which is named as the sample PCSV for the desired site.

   All these steps imply that the PCSV, $\gamma(d_c)$, for site C can be expressed as

$$\gamma(d_c) = \frac{1}{2}\sum_{i=1}^{m-1}(z_D - z_i)^2 \tag{5.27}$$

5. Application of these steps in turn for each site leads to n sample PCSVs of ReV.

These sample PCSVs are potential information sources in describing the ReV characteristics around each site. Among these characteristics are the radius of influence, spatial dependence, and structural behavior of the regionalized variable near the site such as the nugget (sudden changes) and sill effects and heterogeneity as will be explained in the following section.

Example 5.1
The PCSV proposed in the previous section is applied to seismic data from Turkey (Erdik et al. 1985). Conventional probabilistic procedures are used to construct the seismic hazard maps for Turkey. Seismic hazard is defined as the probability occurrence of ground motion due to an earthquake of a particular site capable of

causing significant loss of value through damage or destruction at a given site within a definite time period. The PCSV technique is applied to all sites; some of the sample PCSVs are presented in Fig. 5.44, and after grouping, the major sample trends of PCSV appeared as shown in Fig. 5.45. Consideration of these figures individually or comparatively leads to the following significant interpretations which can be implemented in any future seismic hazard mapping models of Turkey (Şen 1997):

1. Individual seismic PCSVs have a rather different appearance from each other which indicates that the regional seismic distribution in Turkey is heterogeneous. It is a necessary and sufficient requirement for areal heterogeneity that the sample PCSV should exhibit different patterns within the limits of sampling error at different sites. A first glance through the whole sample PCSVs gave the impression that in general, there are ten distinctive categories of them within Turkey. These categories are labeled alphabetically from A to J as shown in Fig. 5.45. Each of these categories has different features, which reflect the seismic record behavior around the station concerned. For instance, those sample PCSVs, which are in category F (see Fig. 5.45), have an initial part with convex curvature (at small distances), and then either single or multiple broken straight lines follow at large distances. On the other hand, in category D sample PCSVs do not have any curvature but many broken straight lines and have an intercept on the horizontal distance axis. An abundance of broken straight lines indicates the heterogeneity involved around the station concerned at different distances.

In category B, the sample PCSVs expose a single straight line indicating that there are rather homogeneous areas of influence around these stations.

The stations within each category can be regarded as homogeneous collectively and hence it is clear that the whole study area has about ten distinct homogeneity regions of seismic variation. Further useful interpretations about the sample PCSVs can be listed as follows:

2. Some of the sample PCSVs do not pass through the origin. This is tantamount to saying that the seismic occurrences at these sites cannot be considered as regionally smooth processes, but rather the seismic event is under the control of some local and/or regional geological factors. This further implies that in the spatial seismic occurrences, uniform conditions do not prevail but rather complex combination of multitude tectonic events. Last but not the least, if the sample PCSV has an intercept on the vertical axis, it implies the existence of a nugget effect in the regional variability at the site concerned (Fig. 5.45J).
3. Some of the sample PCSVs have intercepts, R_0, on the horizontal (distance) axis (see categories C, F, and I). This means that at distances less than R_0, the sample PCSV value is equal to zero; hence, from Eq. 5.27, $Z_c \cong Z_i$ implying structural control within the regional seismic event. Furthermore, in general, big seismic values follow big seismic values and small ones follow small

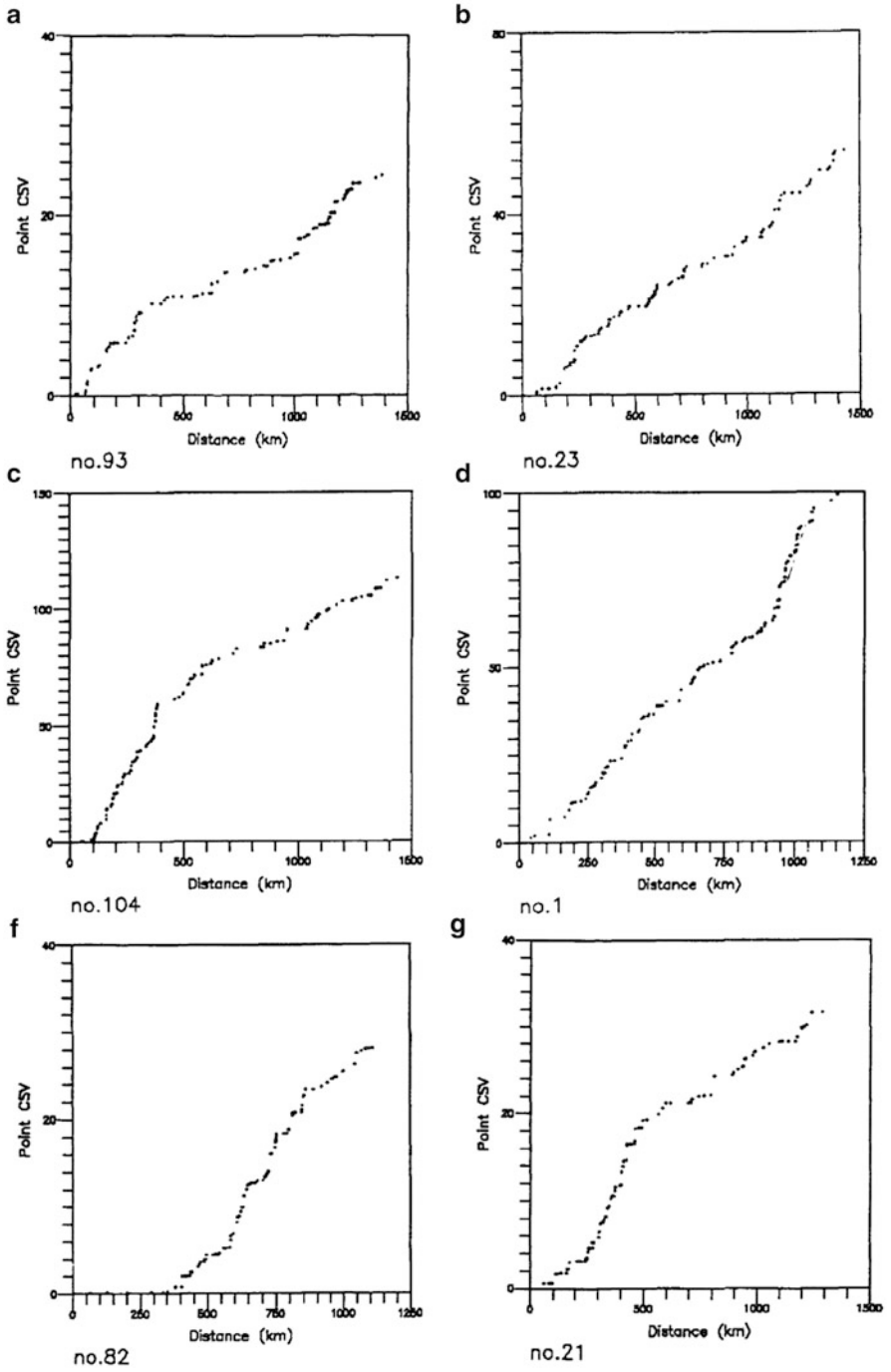


Fig. 5.44 PCSVs of seismic events in Turkey

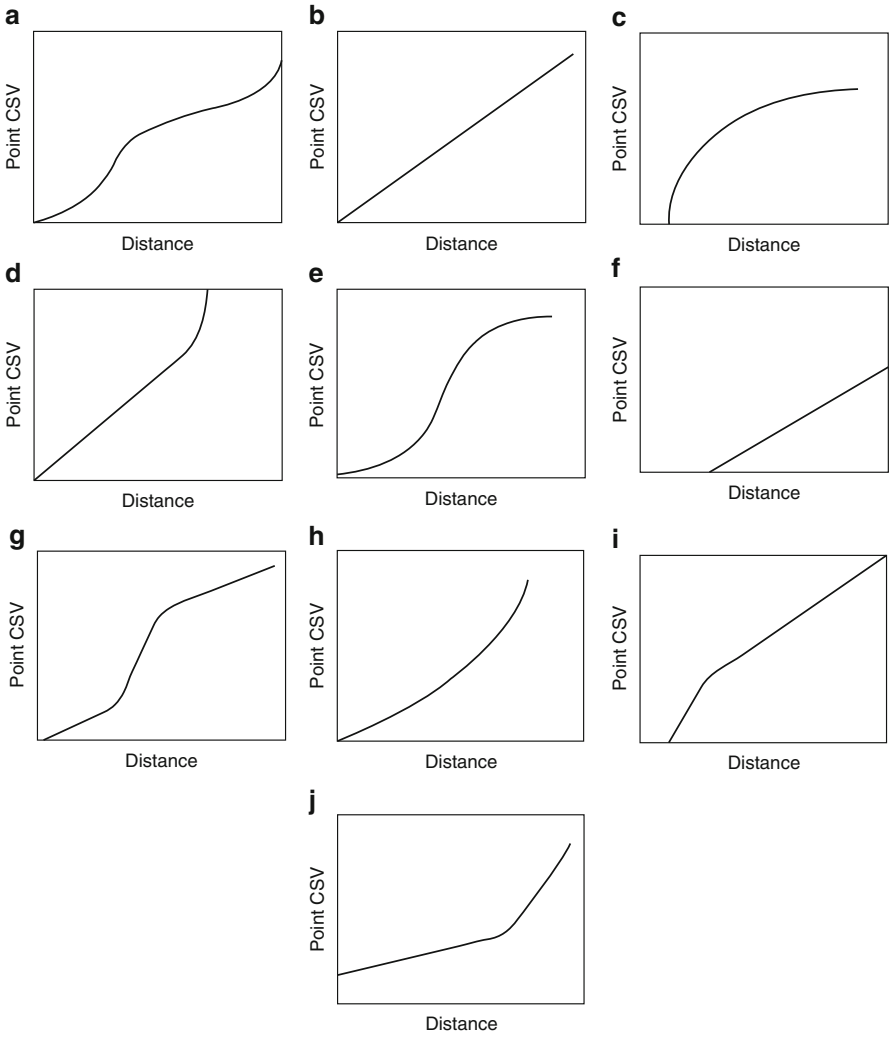


Fig. 5.45 Seismic PCSV categories in Turkey

seismic values, i.e., there are isolated islands of high or low seismic locations around the site.

4. That each sample PCSV fluctuates about a straight line for large distances. The existence of straight-line portions in the PCSV implies that seismic activities at large distances are independent from each other. This is equivalent to the regional sill property in the classical SV but PCSV provides information about the sill at individual sites (see Fig. 5.45J). These portions correspond to horizontal segments at large distances in classical SV as defined by

Matheron (1965). Furthermore, this is the only range where the classical formulations keep their validity.

5. A group of sample PCSVs passes through the origin (see category in figures A, B, D, E, G, and H). Such a property on the PCSV diagram implies the continuity of seismic effects from the site outward. Continuity means that there are no nugget effects or discontinuities within the regional seismic variable at the site concerned.
6. Some of the sample PCSVs has "curvature portions" for a moderate range of distances (categories E and G). In fact, such a range corresponds to the distance scale as defined in the turbulent flow by Taylor (1915). After this range, the PCSVs converge to straight lines (Sen 1989). The initial curvature implies that the seismic at these sites has regional dependencies, which weaken toward the end of curvature distance range (Sen 1989). Since curvatures are convex, there is positive regional structural dependence (Şen 1992). Furthermore, the curvature implies that the seismic has areal structural dependence.
7. In some of the sample PCSVs, there is no curvature part at all (categories B and F). Such a situation is valid in cases where the regional seismic distributions arise predominantly due to the activities of external factors only. Furthermore, there is no structural correlation, i.e., seismic phenomena evolve randomly over the region concerned.
8. As suggested by Şen (1992), the sample CSVs help to identify the underlying generating mechanism of regional phenomenon. Likewise, the sample PCSVs provide clues about the seismicity-generating mechanism around the site concerned. For instance, if the sample PCSV passes through the origin and has a straight-line portions only, then the regional phenomenon concerned complies with independent (white noise) process with no regional dependence at all. However, when the sample PCSV is in the form of straight line but does not pass through the origin, then a moving average process is the underlying generating mechanism of the regional variability (Şen 1992). The PCSVs in category F have such a property, and therefore, it is possible to conclude that moving average mechanisms are dominant at these sites.
9. In the case of a single straight line following a curved initial portion, the slope of long-distance straight-line portion is related to the regional standard deviation of the underlying precipitation-generating mechanism.
10. As mentioned above the PCSV is an indicator of cumulative similarity of the seismic variation at a station with other stations. Practically, if the two PCSV at different sites follow the same pattern within the limits of sampling errors, then they are said to be similar. Such a similarity implies heterogeneity to exist between two sites. They may be partially similar to each other at some distances if the PCSV values at the same distances are close to each other again within the limits of an acceptable sampling error. In practice, in order to appreciate the distances over which the seismic records measured at different sites are similar to each other, it is helpful to look at the regional map for fixed PCSV levels. Herein, for the sake of discussion, two PCSV levels at 10 and 20 are considered for construction of such a regional similarity map. Table 5.1

Table 5.1 Similarity measures at fixed level of PCSV

	Fixed PCSV levels	
Station number	10	20
1	180	305
2	300	1110
3	440	940
4	405	1295
5	300	1110
6	295	550
7	500	800
8	540	760
9	595	1120
12	370	600
13	180	435
14	280	435
20	470	1310
21	325	520
22	460	1065
23	220	500
24	200	280
30	580	1460
31	250	500
32	400	1070
33	305	880
34	350	1035
40	360	975
41	95	115
42	450	735
43	405	900
50	370	1165
51	220	470
52	430	650
53	280	420
60	115	295
61	300	420
62	320	1025
63	400	1205
70	705	1050
71	210	395
72	300	750
73	275	700
80	300	205
81	405	1110
82	620	800

(continued)

Table 5.1 (continued)

	Fixed PCSV levels	
Station number	10	20
83	215	500
91	200	645
92	200	500
93	320	1135
100	200	280
101	325	580
102	220	620
103	230	680
104	320	195
105	110	290
106	500	645

shows the distances obtained from the PCSV graphs at 10 and 20 PCSV values. Obviously, it is possible to prepare similar maps for any desired study area.

The relevant similarity maps are shown in Figs. 5.46 and 5.47. For a fixed PCSV level, the smaller the distance, the more the seismic activity effect at the site, and the smaller the regional dependence, i.e., the location of the site has relatively intense seismic occurrences than the other sites or the regions. For instance, in Fig. 5.46, the map shows intense seismic occurrences in the eastern part of Turkey surrounded by the 120 contour lines. The next most intensive seismic variations are observed in the central, northeast, and western portions of the country where contour lines of 280 occur. However, the least sensitive locations are in the southern parts with similarity contours of about 600.

On the other hand, at the higher level of similarity as presented in Fig. 5.47, the study area seems more heterogeneous but major distinctive regional zones as in Fig. 5.46 remain the same.

Principles of PCSV have been explained and applied to seismic data in Turkey. This type of PCSVs provides detailed information about a regional variable at and near the measurement sites as well as among the sites. The main purpose of the CSV technique is to check the heterogeneity of the regionalized variable. If the empirical point CSVs at different sites have similar patterns within a certain error band such as 5–10 %, then the regionalized variable is homogeneous and otherwise heterogeneity exists. The point CS concept brings an additional new concept which provides an opportunity to make spatial variability interpretations at each site rather than regionally. Interpretations of relevant PCSV at any site provide useful information concerning the smoothness, structural control, regional dependence, continuity, and radius of influence. The PCSV methodology proposed in this paper is applied to the scattered seismic variation over Turkey. Finally, a similarity map is obtained which provides a basis for the regional heterogeneity assessments.

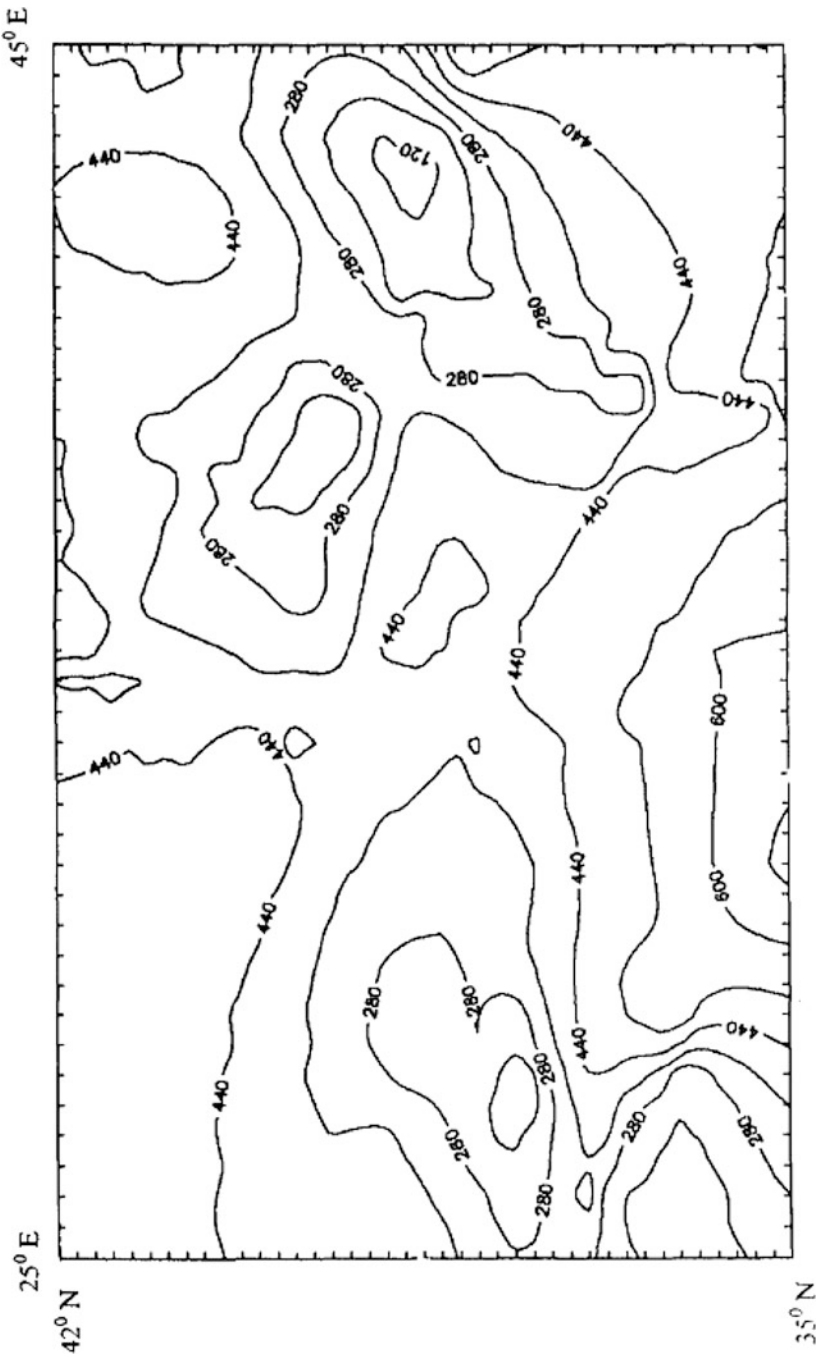


Fig. 5.46 Similarity map at ten PCSV value

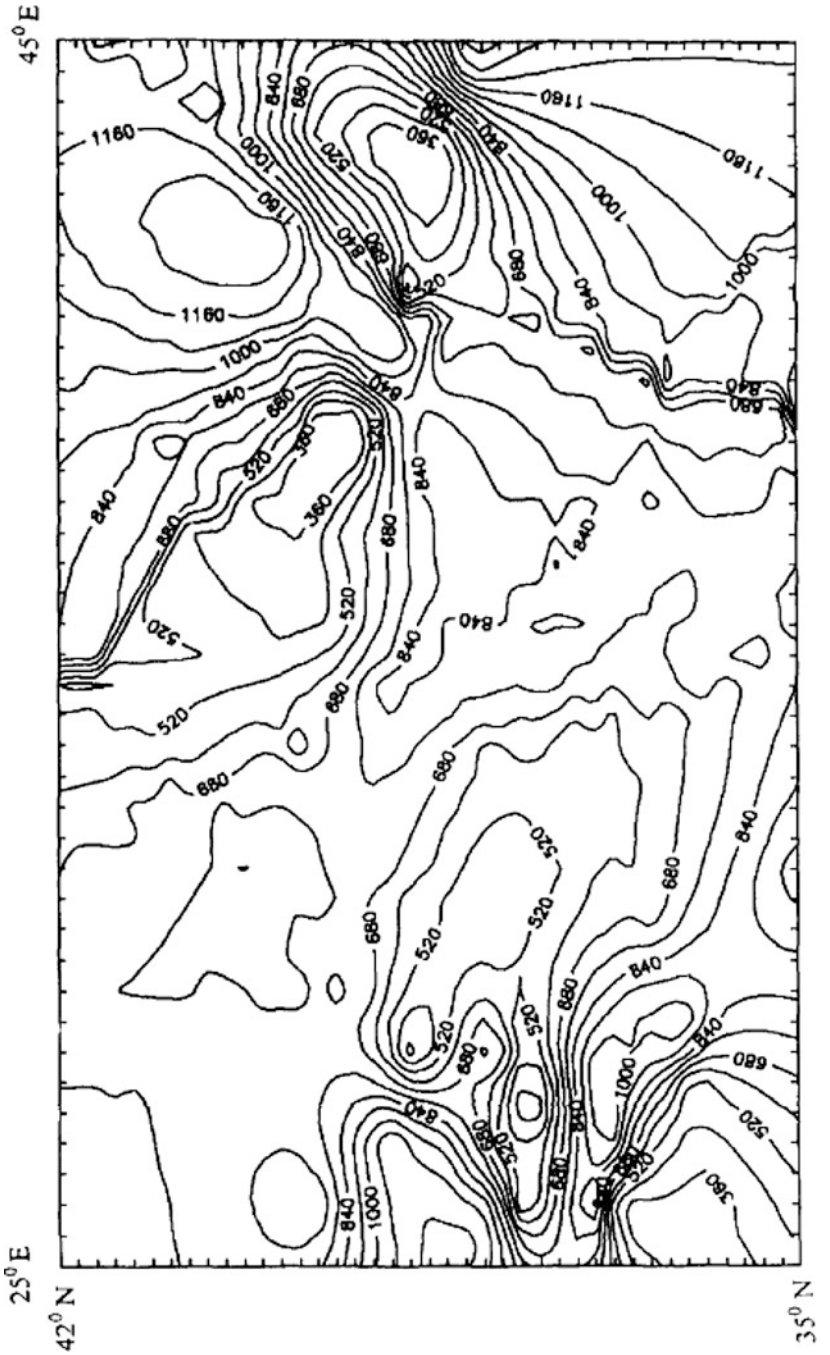


Fig. 5.47 Similarity map at 20 PCSV value

5.10 Spatial Dependence Function (SDF)

As already mentioned calculation of a sample CSV (or PCSV) leads to a nondecreasing function with distance. It is said in Sect. 2.2 that all the classical weighting functions appear as a nonincreasing function with distance. It is, therefore, logical to execute the following steps in order to obtain a valid and standard weighting function from the sample CSV similar to the classical weighting functions:

1. Depict on the sample CSV the maximum distance, R_M, and corresponding sample CSV value, V_M. R_M corresponds to the distance between the two farthest station locations in any study area (see Fig. 5.36).
2. Divide all the distances (CSV values) by R_M (by V_M), and the result appears as a scaled form of the sample CSV within limits of zero and one on both axes. This shows the change of dimensionless CSV with dimensionless distance,
3. Subtraction of the dimensionless CSV values from the maximum value of one appears as a nonincreasing function of the dimensionless distance as shown in Fig. 5.48 which has a similar pattern to all the classical weighting functions as explained in the previous section. This function is referred to as the standard dependence function (SDF).

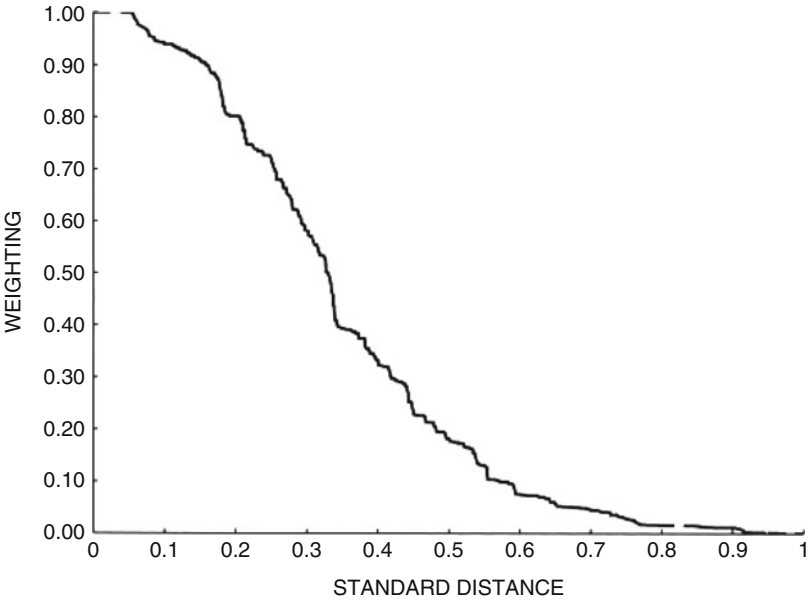


Fig. 5.48 A representative SDF

Example 5.2

In order to explain the experimental CSV and thereof derived SDFs, the monthly rainfall amounts at a set of stations are considered, each with at least 30 years of records in the northwestern part of Turkey (Şen 1997). Some of the CSVs are presented in Fig. 5.49 with corresponding classical and calculated SDF values in Fig. 5.50.

It is possible to conclude from Fig. 5.49 that:

1. They all reach almost horizontal CSV value at large distances which means that after a certain distance, there is no regional effect of one station on other stations' rainfall amount. This distance corresponds to R in Eqs. 3.24–3.26.
2. Initially, all the CSVs have an intercept on the horizontal distance axis at about 5 km. This corresponds to almost the smallest distance between the stations.

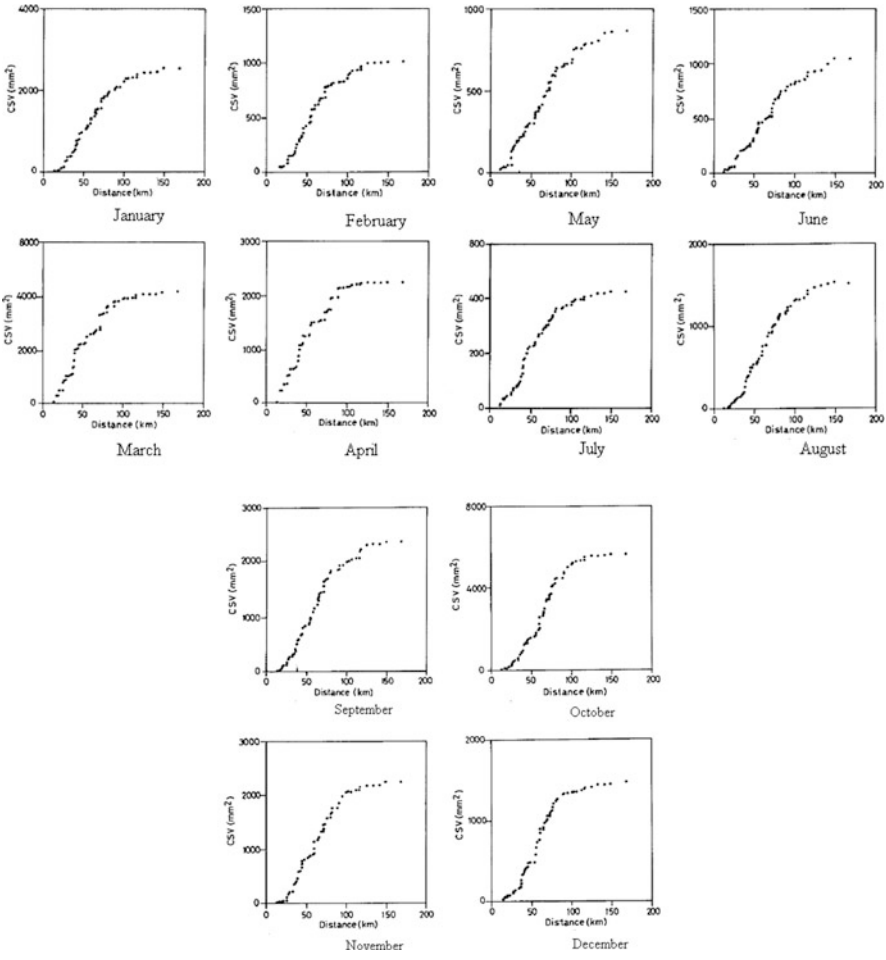


Fig. 5.49 Monthly CSVs

3. The smallest and the greatest CSV values at large distances occur during July and October, respectively, which are the transition months in this region from the Black Sea to the Mediterranean Sea climate in July and vice versa in October.

On the other hand, Fig. 5.50 includes the geometric weighting functions already given in Fig. 3.31 for the sake of comparison. The following interpretations and conclusions can be drawn from these figures for each month:

1. In January the experimental CSV weighting function does not conform by any of the classical models. Initially, at small distances, it is above all the models and then becomes closer to exponential model but only up to about 0.30 dimensionless distances, deviating from it thereafter.

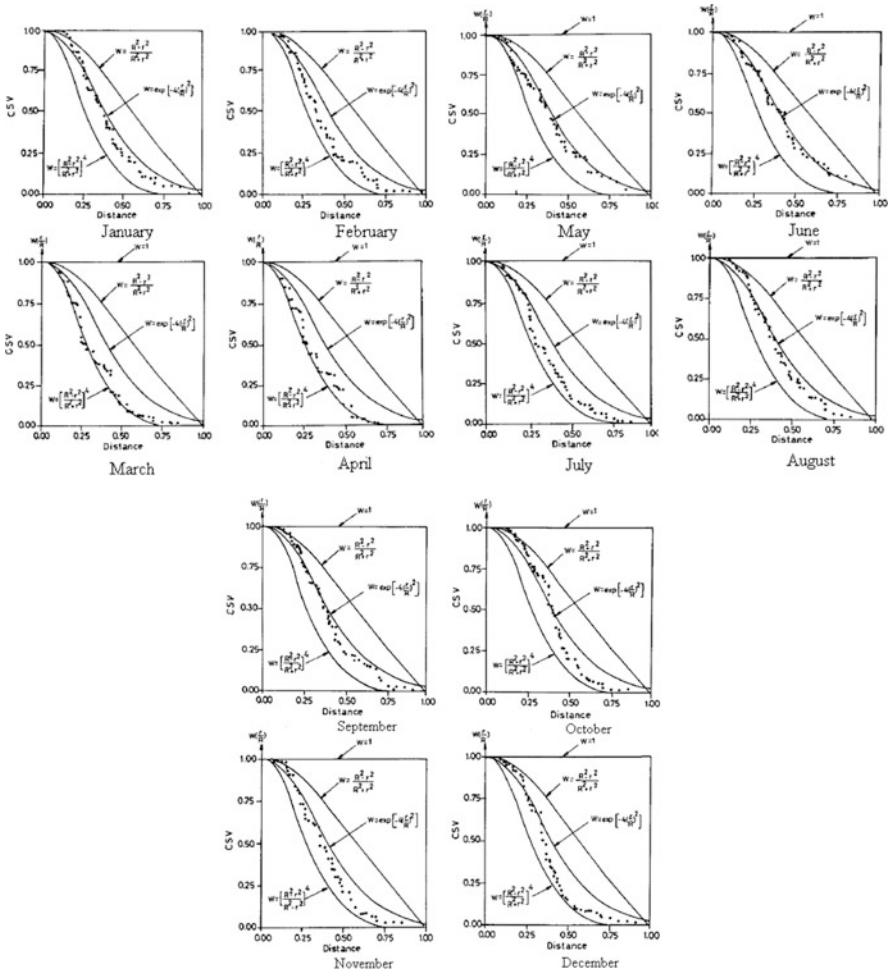


Fig. 5.50 Monthly weighting functions

2. In February perhaps initial portion confirms with the exponential model but later becomes closer to the power model.
3. March experimental CSV follows the power law of the weighting functions at almost every distance.
4. Similar pattern to March repeats itself in April, but in May although initially there is a portion abiding with the power model, then it converts to exponential model. In June, the experimental CSV weighting function comes even closer to the exponential weighting function.
5. During September very small and big distances come closer to the ratio and power models, respectively, but for moderate distance, it has an exponential form. So the mixture of three models appears as a hybrid model in representing the regional variability. Such a phenomenon cannot be estimated only by considering one of the classically available models. In the remaining months, similar interpretations are valid.

These discussions indicate that classical geometric weighting models do not have full justification for the whole of the meteorological phenomenon but they are good first approximations. They cannot be valid for the whole regional variability in any study area.

Example 5.3 For the implementation of the SDF and spatial estimation methodology, the iron (Fe) percentage (%) concentration data set is used by Clark (1979, page 95, Table 5) and is adopted. Although there are 50 sampling points, in order to make comparison with fuzzy clustering results by Pham (1997), only 21 sites are given in Table 5.2, because he considered only 21 sampling points from Clark's table. Figure 5.51 shows the locations of the sites within the study area.

It is obvious that there is a rather uniform distribution of these sites in a representative manner over the whole study area.

The CSV is given in Fig. 5.52. The maximum CSV corresponding to 350 m is 4250 and accordingly the SDF is obtained and the results are presented in Fig. 5.53.

Similar to all the regional estimation procedures, weighted average formulation as in Eq. 2.27 is used together with the weights obtained from SDF in Fig. 5.53.

In order to assess the validity of the proposed weighted average procedure, a cross-validation technique is used. According to this a data value at one site is supposed to be unknown and is removed from the data set. This removed value is then estimated with the remaining set of data and by using the SDF together with Eq. 2.27. This procedure is repeated for all the sites knowing that a data removed for estimation at its location is put again in the set for the estimation at another location.

There are two procedures in the estimation of the site Fe % concentrations. In the first one, all the other sites are considered for their simultaneously contributions, and, therefore in the estimation of any site's Fe % concentration, all the distances from this site to others ($n-1$ sites) are measured from the map in Fig. 5.51. Subsequently, these distances are entered into Fig. 5.53 on the horizontal axis and the corresponding RDF weights are found from the vertical axis for each distance. In this manner, all the sites are treated equally by the same procedure, and hence,

Table 5.2 Measured and estimated Fe % concentrations and errors

Sample no.	Measured	Estimate	Relative error (%)
(1)	(2)	(3)	(4)
1	35.5	32.85	7.45
2	29.4	32.78	11.51
3	36.8	33.72	8.36
4	33.7	34.15	1.32
5	35.3	33.33	5.58
6	32.5	32.56	0.19
7	30.6	33.88	10.73
8	30.1	33.52	11.37
9	40.1	33.76	15.81
10	31.6	33.84	7.10
11	34.8	33.32	4.24
12	28.6	32.76	14.55
13	41.5	33.48	19.32
14	33.2	33.51	0.92
15	34.3	33.56	2.31
16	31.0	33.82	9.09
17	29.6	32.50	9.78
18	40.4	33.70	16.59
19	28.5	34.08	19.59
20	24.4	34.55	41.62
21	39.5	32.91	16.69
Average	33.40	33.46	11.15

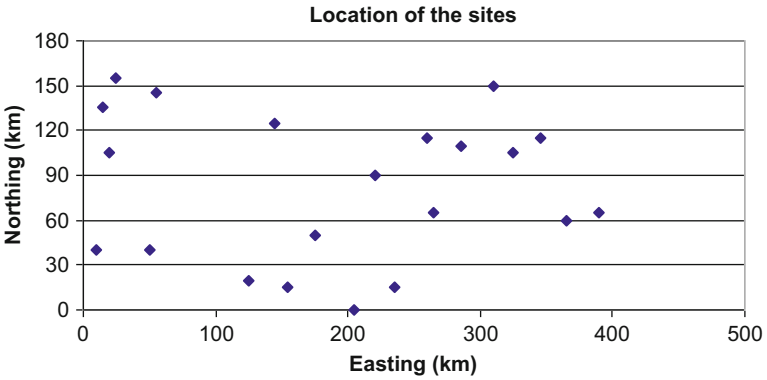


Fig. 5.51 Fe % concentration locations

instead of measured values, their estimations through the SDF and the cross-validation procedure are calculated. Column (3) in Table 5.2 shows estimated Fe % concentrations, and their corresponding relative errors are calculated as the

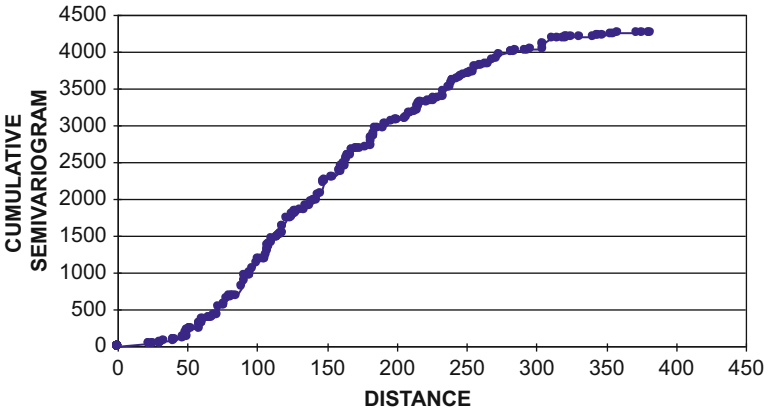


Fig. 5.52 Fe % concentration sample CSV

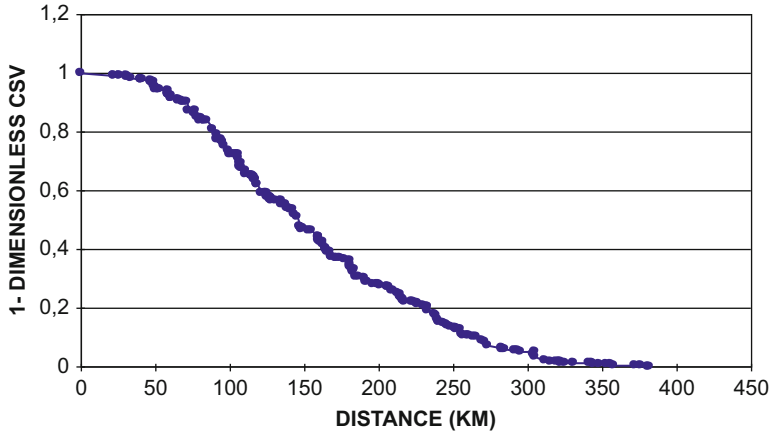


Fig. 5.53 Fe % concentration, (**a**) dimensionless CSV, (**b**) SDF

ratio of the absolute difference between the measured and estimated values divided by the measured value multiplied by 100, which is shown in column (4):

$$\text{Relative Error} = 100^{*}\frac{|\text{measured} - \text{estimated}|}{\text{measured}}$$

It is obvious that for almost half of the sites, the relative error is more than 10 %, which indicates the unsuitability of the so far proposed procedure. For extreme Fe % concentration sites, the relative errors are very high. However, the averages of measured and estimated values are very close together with 2 % relative error. On this basis, it may be concluded that the proposed procedure yields reasonable values on the average but fails to estimate individual site values. This procedure takes into

account the contribution of all the sites in the estimation and disregards the concept of the radius of influence.

In order to assess visually the correspondence between the measured and estimated values, Fe % concentrations are presented in Fig. 5.54a against the site number sequence along the horizontal axis. Unfortunately, consideration of all the sites in this regional estimation procedure by using Eq. 2.27 and SDF function for weight calculation yields to average estimations which do not represent the high and low points satisfactorily as obvious from Fig. 5.54a. On the other hand, Table 5.3 shows also the results by the inverse distance square and four fuzzy clustral procedures as suggested by Pham (1997). The comparisons of average

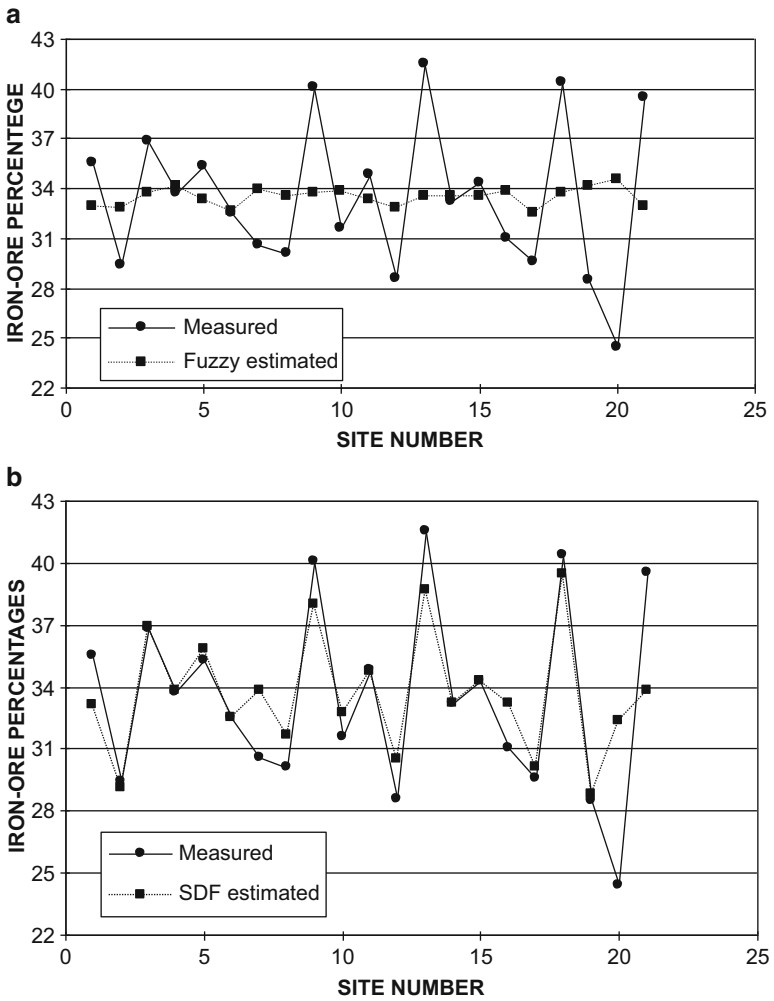


Fig. 5.54 Measured and estimated Fe % concentration, (**a**) whole points, (**b**) adaptive method

Table 5.3 Fe % concentration estimations by different techniques

Sample no.	RDF	Inverse square distance	Fuzzy clustering			
			(4)			
(1)	(2)	(3)	Three	Four	Five	Six
1	32.85	33.44	34.16	34.19	34.83	33.10
2	32.78	31.92	33.28	32.39	31.76	33.20
3	33.72	30.87	34.04	30.50	31.28	31.00
4	34.15	30.77	33.96	33.60	30.23	31.15
5	33.33	36.67	33.11	33.53	35.75	35.23
6	32.56	35.61	33.46	33.11	34.10	34.69
7	33.88	33.24	33.97	33.60	33.00	33.22
8	33.52	35.47	33.63	33.22	33.95	37.46
9	33.76	32.62	33.88	32.83	34.23	34.44
10	33.84	33.24	33.38	33.28	33.12	32.38
11	33.32	36.66	33.84	33.04	34.02	34.29
12	32.76	30.50	33.61	32.19	31.63	31.69
13	33.48	29.85	33.93	33.93	31.32	30.17
14	33.51	34.60	34.11	33.82	30.95	32.20
15	33.51	37.35	33.23	32.96	33.54	34.68
16	33.82	33.24	34.34	33.78	33.78	34.06
17	32.50	34.67	33.05	33.25	38.10	35.17
18	33.70	34.04	34.93	34.37	34.47	35.73
19	34.08	34.12	33.78	33.21	34.68	34.98
20	34.55	32.29	31.87	32.06	32.48	32.13
21	32.91	34.06	33.16	33.18	32.74	31.98
Averages	33.45	33.58	33.65	33.14	33.33	33.49
Average relative errors (%)	**11.15**	**12.66**	**12.55**	**12.65**	**12.01**	**12.61**

relative errors in the last row of this table indicate that the SDF method has the least relative error percentage. It is even better than the five-cluster case which has been stated as the best solution by Pham (1997).

In order to improve the representativeness of the Fe % concentration regionalized variable estimations at sites, herein, an adaptive new technique is suggested, which does not only estimate the regional value at a site but also provides the number of the nearest sites that should be considered in the best possible regional estimation. Accordingly, the radius of influence is defined as the distance between the estimation site and the far distant site within the adjacent sites that are considered in the regional estimation procedure. The following steps are necessary for the application of this adaptive procedure:

(a) Take any site for cross-validation and apply Eq. 2.27 by considering the nearest site only. Such a selection is redundant and corresponds to the assumption that,

if the nearest site measurement is considered only, then the regional estimation will be equal to the same value. This means that in such a calculation, the radius of influence is the minimum and equal to the distance between the estimation and the nearest sites.

(b) Consider now two of the nearest sites to the estimation site and apply the RDF weighting method according to Eq. 2.27. Consideration of two sites will increase the radius of influence as the distance between the estimation and the next nearest site and the estimation value will assume the weighted value of the two nearest sites. Since the weights and measurements are positive numbers, the estimated value will be between the measurements at the two nearest sites. There will be a squared estimation error as the square of the difference between measured and estimated values.

(c) Repeat the same procedure now with three nearest stations and calculate the square error likewise. Subsequently, it is possible to continue with 4, 5,. . ., $(n-1)$ nearest sites consequently and for each case to calculate the corresponding square error. The first one with the least square error yields the number of nearest sites for the best regional Fe % concentration estimation. The distance of the farthest site in such a situation corresponds to the radius of influence. As an example, herein, only site 14 calculations are presented in detail in Table 5.4. It is obvious that when Eq. 2.27 is applied by considering 11 nearest sites, the estimation error square becomes the least with the radius of influence equal to 127.47 m.

Table 5.4 Detailed calculation steps for site 14 regional estimation

Number of nearest sites	Estimation	Square error
(1)	(2)	(3)
2	31.94	1.58
3	29.46	13.96
4	30.75	6.00
5	32.42	0.60
6	32.21	0.97
7	32.38	0.67
8	32.82	0.14
9	32.59	0.37
10	32.72	0.23
11	**33.21**	**0.00**
12	33.12	0.01
13	33.48	0.08
14	33.80	0.36
15	33.69	0.24
16	33.63	0.18
17	33.57	0.13
18	33.55	0.12
19	33.48	0.08
20	33.51	0.09

Application of the above adaptive procedure to Fe % concentrations results in the estimation values, the number of the nearest sites with the minimum squared error, and the radius of influence, which are presented in Table 5.5. Notice that this table includes the complete list of samples as given by Clark (1979, page 95, Table 5.5). This table yields the following points:

1. The adaptive estimation procedure gives average iron-ore concentration value similar to average measurements with 2.31 % error. Hence, similar to all the previous methods adaptive estimation also gives reasonable average values,
2. Comparison of average relative error in Table 5.4 with average relative errors in Table 5.4 shows clearly that the adaptive method with 4.74 % error is the best among all approaches and the reduction in the relative error implies that deviations from average level are taken into account effectively. Figure 5.49b presents the adaptive estimations together with the measured values for Pham (1997) data set. If Figs. 5.49a and b are compared, it is then obvious that deviations are better accounted by the adaptive method,
3. The adaptive approach provides the radius of influence for each station as shown in the last column of Table 5.5. The average radius of influence is about 88.5 m with maximum and minimum values of about 223 m and 13 m, respectively.

By making use of the radius of influence from Table 5.5, it is possible to construct equal radii regional map as shown in Fig. 5.55.

From this map, one can know the relevant radius of influence for any desired point within the study area. Once this radius is determined, then a circle with the center at the prediction point is drawn. The measurement sites within this circle are taken into consideration in the application of Eq. 2.27 for regional estimation through the SDF weights.

After having completed the cross-validation procedure and the map of the radius of influence, it is now time to present spatial interpolation procedure with RDF usage as follows:

1. Select any number, say 15, of spatially scattered points within the study area as shown in Fig. 5.51. These sites are locations without measurements. For the sake of argument, they are selected rather arbitrarily with easting and northing coordinates as shown in Table 5.6.
2. The radius of influence of each site is determined from the map in Fig. 5.55 and written in the fourth column of Table 5.6.
3. Consideration of the radius of influence for each site defines the number of measurement sites within this radius that are basis for the Fe % concentration estimation through Eq. 2.27 at this site. Hence, the measurement sites that will be considered in the spatial interpretation of the Fe % concentration at the site are identified.
4. Subsequently, distances between the interpolation site and the effective measurement sites are calculated.
5. The entrance of these distances on the horizontal axis in the SDF (Fig. 5.49b) yields the weights on the vertical axis.

Table 5.5 Estimations and radius of influence

Easting	Northing	% Fe	Estimation	Relative error (%)	Number of nearest sites	Radius of influence (m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0	170	34.30	32.15	6.27	3	13.25
10	40	35.50	33.10	6.76	10	199.06
15	135	28.60	30.50	6.64	3	41.23
55	145	29.40	29.10	1.02	2	41.23
125	20	41.50	38.65	6.87	2	58.31
175	50	36.80	36.91	0.30	5	69.46
120	180	33.40	31.52	5.63	6	73.16
160	175	36.00	33.27	7.58	5	97.84
240	184	30.20	31.30	3.64	4	86.17
260	115	33.20	33.21	0.03	11	127.47
235	15	33.70	33.84	0.42	3	69.46
365	60	34.30	34.26	0.12	4	94.34
285	110	35.30	35.83	1.50	3	47.17
345	115	31.00	33.17	7.00	8	127.47
335	170	27.40	28.12	2.63	6	65.13
325	195	33.90	34.75	2.51	6	123.10
350	235	37.60	38.02	1.12	5	73.10
290	230	39.90	41.15	3.13	4	68.17
10	390	27.20	29.12	7.06	3	55.64
85	380	34.20	32.75	4.24	3	112.51
50	270	30.20	31.60	4.64	6	78.32
200	280	30.40	31.06	2.17	4	85.76
400	355	39.90	41.73	4.59	5	67.23
360	335	40.00	42.16	5.40	5	72.15
335	310	40.60	41.18	1.43	4	83.72
5	195	33.90	34.73	2.45	3	198.12
20	105	32.50	32.50	0.00	9	164.50
25	155	29.60	30.14	1.82	3	50.25
50	40	30.60	33.82	10.52	14	222.99
155	15	40.40	39.50	2.23	3	52.50
145	125	30.10	31.63	5.08	3	92.19
130	185	35.30	33.18	6.01	3	187.15
175	185	41.40	39.17	5.39	3	59.37
220	90	28.50	28.82	1.12	2	51.48
205	0	40.10	37.96	5.34	4	82.46
265	65	24.40	32.34	32.54	3	51.48
390	65	31.60	32.73	3.58	2	67.29
325	105	39.50	33.89	14.20	4	60.21
310	150	34.80	34.77	0.09	4	61.03
385	165	29.90	28.16	5.82	5	97.25
325	220	37.80	36.08	4.55	6	75.34

(continued)

Table 5.5 (continued)

Easting	Northing	% Fe	Estimation	Relative error (%)	Number of nearest sites	Radius of influence (m)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
375	215	29.80	26.18	12.15	5	56.76
200	230	37.40	35.83	4.20	4	82.63
55	375	27.40	28.17	2.81	3	62.75
395	245	36.50	35.08	3.89	5	192.50
165	355	40.80	39.15	4.04	5	55.25
270	285	32.90	34.75	5.62	4	155.50
365	340	40.00	40.80	2.00	5	86.13
330	320	44.10	45.83	3.92	5	69.12
330	290	41.40	37.16	10.24	4	60.73
Averages		**34.50**	**34.42**	**4.74**	**5**	**88.51**

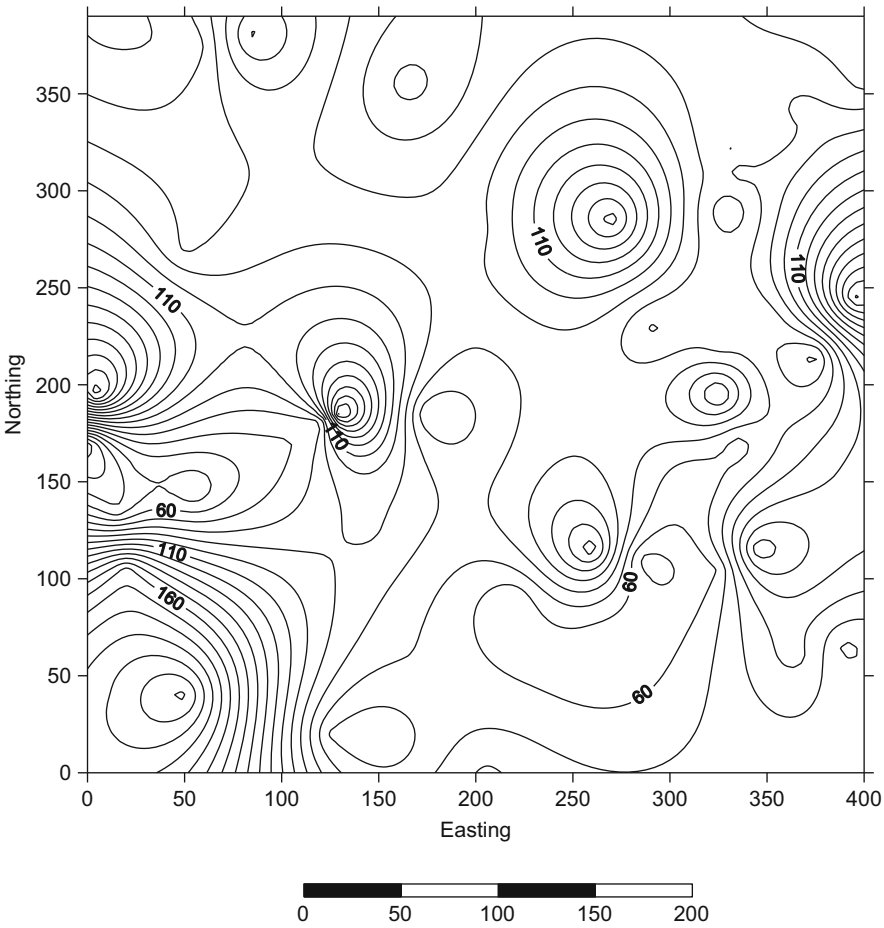


Fig. 5.55 Iron ore radius of influence map (m)

6. Substitution of all the relevant values into Eq. 2.27 provides the Fe % concentration value estimations at each site, which are shown in the last column of Table 5.6.
7. In order to check the reliability of the estimations, the question is now whether these spatial estimations will yield almost the same SDF or not. For this purpose, the SDF calculation steps are applied to data in Table 5.6.
8. Figure 5.56 indicates the resulting SDF for the measured and estimated Fe % concentrations. The maximum relative difference between these two SDFs is

Table 5.6 Regional interpolation

Site no.	Easting	Northing	Radius of influence (m)	Estimation Fe (%)
(1)	(2)	(3)	(4)	(4)
1	36	136	63	30.02
2	27	90	172	32.71
3	86	116	97	30.13
4	45	29	200	33.81
5	95	47	118	34.88
6	186	23	68	38.49
7	218	58	50	32.63
8	222	129	95	31.77
9	268	98	80	32.35
10	327	134	93	33.47
11	272	43	58	28.99
12	340	40	80	32.30
13	331	65	77	32.78
14	327	87	70	33.08
15	368	96	103	34.37

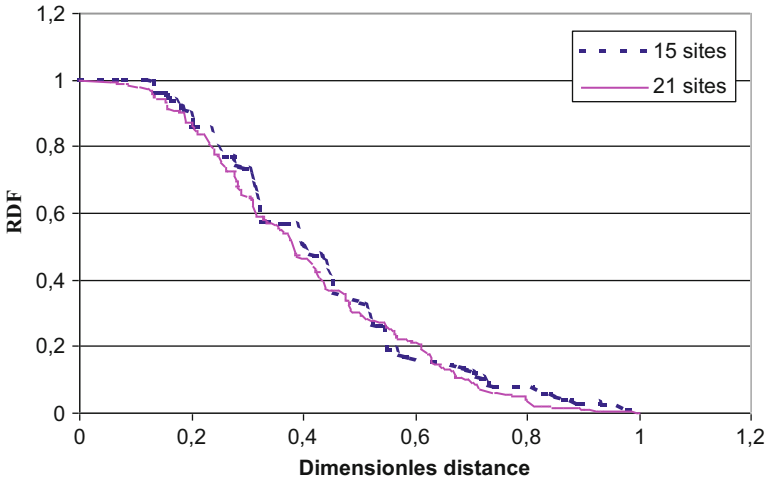


Fig. 5.56 Measured and estimated SDFs

less than 5 % which confirms the practical validity of the SDF adaptive estimation procedure methodology as proposed in this paper.

It is now time to compare all the methods with the measured data on a Cartesian coordinate system with the measured values on the horizontal axes and the model outputs on the vertical axes as in Figs. 5.57 and 5.58.

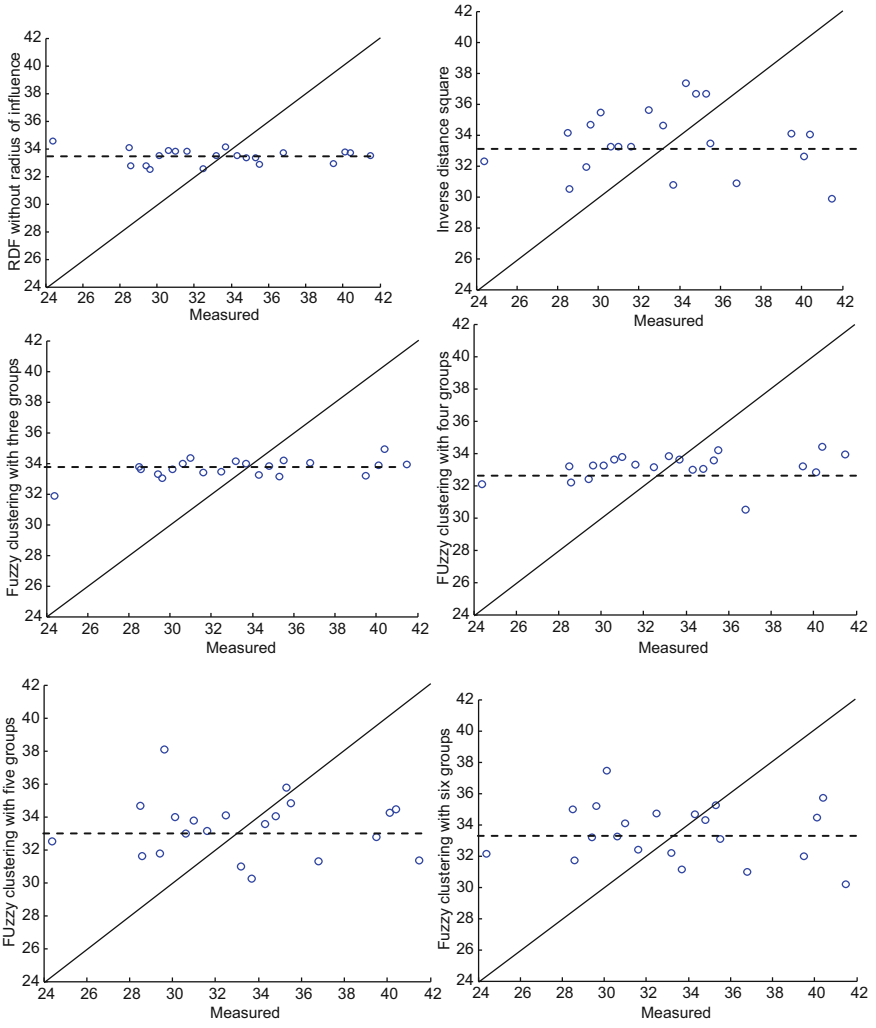


Fig. 5.57 Comparative model verifications

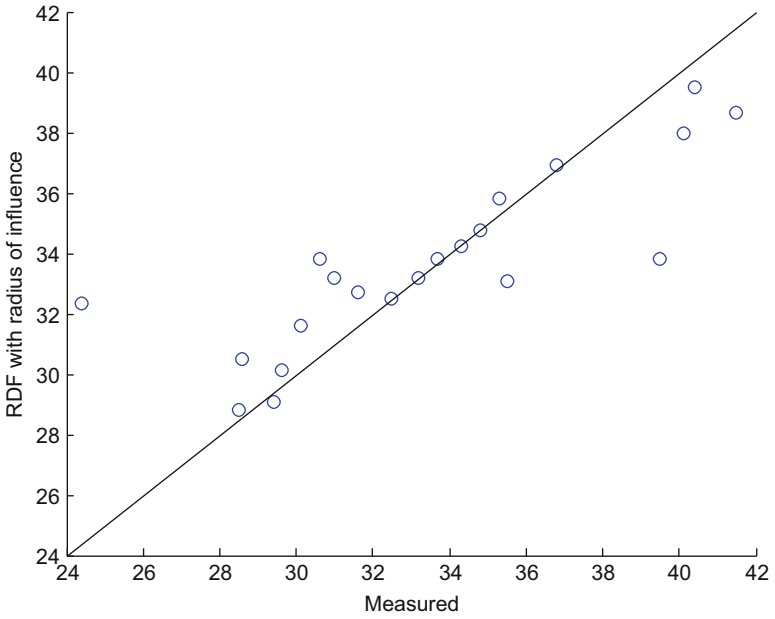


Fig. 5.58 SDF model verification

References

Aboufirassi M, Marino MA (1984) A geostatistically based approach to the identification of aquifer transmissivities in Yolo Basin, California. Math Geol 16(26):125–137

Barnes SL (1964) A technique for maximizing details in numerical weather map analysis. J Appl Meteorol 3:396–409

Barnes RJ (1991) The variogram sill and the sample variance. Math Geol 23(4):673–678

Barros VR, Estevan EA (1983) On the evaluation of wind power from short wind records. J Clim Appl Meteorol 22:1116–1123

Benjamin JR, Cornell CA (1970) Probability statistics and decision making in civil engineering. McGraw-Hill Book, New York

Bergman K (1979) Multivariate analysis of temperatures and winds using optimum interpolation. Mon Weather Rev 107:1432–1444

Bergthorsson P, Döös BR (1955) Numerical weather map analysis. Tellus 7:329–340

Box GFD, Jenkins GM (1976) Time series analysis, control and forecasting. Holden Day, San Francisco

Bratseth AM (1986) Statistical interpretation by means of successive corrections. Tellus 38A:438–447

Buzzi A, Gomis D, Pedder MA, Alonsa S (1991) A method to reduce the adverse impact that inhomogeneous station distributions have on spatial interpolations. Mon Weather Rev 119:2465–2491

Carr JR, Bailey RE, Deng ED (1985) Use of indicator variograms for enhanced spatial analysis. Math Geol 17(8):797–812

Clark I (1979) The semivariogram – part 1. Eng Min J 180(7):90–94

Cooley RL (1979) A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 2, applications of statistical analysis. Water Resour Res 15:603–617
Cressie NAC (1993) Statistics for spatial data, revisedth edn. Wiley, New York, p 900
Cressman GP (1959) An operational objective analysis system. Mon Weather Rev 87:367–374
Daley R (1991) Atmospheric data analysis. Cambridge University Press, Cambridge, 475 pp
David M (1977) Geostatistical ore reserve estimation. Elsevier, New York, p 340
Davis J (1986) Statistic and data analysis in geology. Wiley, New York, 560 p
Dee PD (1995) On-line estimation of error covariance parameters for atmospheric data assimilation. Mon Weather Rev 123:1128–1145
DeWijs HJ (1972) Method of successive differences applied to mine sampling. Trans Inst Min Metal Sect A Min Ind 81:78–81
Eddy A (1964) The objective analysis of horizontal wind divergence fields. Q J Roy Meteorol Soc 90:424–440
Eliassen A (1954) Provisional report on calculation of spatial covariance and autocorrelation of the pressure field, report no. 5. Videnskaps-Akademiet Institut for Vaer og Klimaforskning, Oslo
Erdik M, Doyuran V, Akkaş N, Gülkan P (1985) A probabilistic assessment of the seismic hazard in Turkey. Tectonophysics 117(3/4):295–330
Franke R (1988) Statistical interpretation by iteration. Mon Weather Rev 116:961–963
Gandin LS (1963) Objective analysis of meteorological fields. Hydromet Press, 242 pp
Gilchrist B, Cressman GP (1954) An experiment in objective analysis. Tellus 6:309–318
Goodin WR, McRea GJ, Seinfeld JH (1979) A comparison of interpolation methods for sparse data: application to wind and concentration fields. J Appl Meteorol 18:761–771
Gustafsson N (1981) A review of methods for objective analysis. In: Dynamic meteorology, data assimilation methods. Springer-Verlag. pp 17–76
Hoeksema RJ, Kitandis PK (1984) An application of the geostatistical approach to the inverse problem in two dimensional groundwater modeling. Water Resour Res 20(7):1003–1020
Inman RL (1970) Operational objective analysis schemes at the National Severe Storms Forecast Center. U.S. National Severe Storms Laboratory Tech. Circular 10, Norman, OK, 50 pp
Journel AJ (1985) The deterministic side of geostatistics. Math Geol 17(1):1–15
Journel AG, Huijbregts CI (1978) Mining geostatistics. Academic, London, p 710
Kitanidis PK (1997) Introduction to geostatistics: applications in hydrogeology. Cambridge University Press, Cambridge, 249 pp
Koch SE, DesJardins M, Kocin PJ (1983) An iterative Barnes objective map analysis scheme for use with satellite and conventional data. J Appl Meteorol 22:1487–1503
Krige DG (1982) Geostatistical case studies of the advantages of log-normal, De Wijsian Kriging with mean for a base metal mine and a gold mine. Math Geol 14(6):547–555
Kruger HB (1969a) General and special approaches to the problem of objective analysis of meteorological variable. Q J Roy Meteorol Soc 95(403):21–39
Kruger HB (1969b) Objective analysis of pressure height data for the tropics. Mon Weather Rev 94 (4):237–257
Lorenc AC (1981) A global three-dimensional multivariate statistical interpolation scheme. Mon Weather Rev 109:701–721
Liebhold AM, Rossi RE, Kemp WP (1993) Geostatistics and geographic information systems in applied insect ecology. Annu Rev Entomol 38:303–327
Ludwig JA, Reynolds JF (1988) Statistical ecology: a primer on methods and computing. Wiley, New York, 337p
Matheron G (1965) Les variables regionalises et leur estimation. Masson et Cie., Paris, 306p
Matheron G (1971) The theory of regionalized variables and its applications. Ecole de Mines, Fontainbleau
Myers DE, Begovich CL, Butz TR, Kane VE (1982) Variogram models for regional groundwater geochemical data. Math Geol 14(6):629–644
Pannatier Y (1996) VARIOWIN - Software for spatial data analysis in 2D. Springer Verlag, New York, p 91 p. ISBN 0-387-94679-9

Panofsky HA (1949) Objective weather map analysis. J Meteor 6:386–392
Pedder MA (1993) Interpolation and filtering of spatial observations using successive correlations and Gaussian filters. Mon Weather Rev 121:2889–2902
Perrie W, Toulany B (1989) Correlations of sea level pressure field for objective analysis. Mon Weather Rev 117:1965–1974
Pham TD (1997) Grade estimation using fuzzy-set algorithms. Math Geol 29(2):291–305
Powers RW, Ramirez LF, Redmond CD, Elberg EL (1966) Geology of the Arabian Peninsula. Sedimentary geology of Saudi Arabia. U.S. Geol. Survey, Prof Paper 560-D, New York, pp 1–47
Rutherford ID (1976) An operational three-dimensional multivariate statistical objective analysis scheme. Proceedings of the JOC Study Group, Conference on four-dimensional Data Assimilation, Paris, November 17–21,1975. The GARP Programme on Numerical Experimentation, Report No. 11, January 1976
Sashegyi Y (1960) An objective analysis for determining initial conditional for the primitive equations. Tech Rep:60–16T
Schlatter TW (1975) Some experiments with a multivariate statistical objective analysis scheme. Mon Weather Rev 103:246–257
Seaman RS (1988) Some real data tests of the interpolation accuracy of Bratseth's successive correction method. Tellus 40A:173–176
Şen Z (1978) Autorun analysis of hydrologic time series. J Hydrol 36:7585
Sen Z (1989) Cumulative semivariogram models of regionalized variables. Int J Math Geol 21 (3):891–903
Şen Z (1992) Standard cumulative semivariograms of stationary stochastic processes and regional correlation. Math Geol 24:417–435
Şen Z (1995) Applied hydrogeology for Scientists and Engineers. CRC Lewis Publishers, Boca Raton. 495 pp
Şen Z (1997) Objective analysis by cumulative semivariogram technique and its application in Turkey. J Appl Meteorol 36(12):1712–1724
Şen Z (2002) İstatistik Veri İşleme Yöntemleri (Hidroloji ve Meteoroloji). Su Vakfı yayınları,243 sayfa
Şen Z (2008) Solar energy fundamentals and modeling techniques. Atmosphere, environment, climate change and renewable energy. Springer, London, 276 pp
Sen Z, Al-Dakheel AR (1985) Hydrochemical facies evaluation in Umm Er Radhuma limestone-Eastern Saudi Arabia. Ground Water 24(5):626–635
Şen Z, Habib ZZ (1998) Point cumulative semivariogram of areal precipitation in mountainous regions. J Hydrol 205:81–91
Şen Z, Habib Z (2001) Monthly spatial rainfall correlation functions and interpretations for Turkey. Hydrol Sci J 46(5):829–829
Shenfield L, Bayer AE (1974) The utilization of an urban air pollution model in air management. Proc. 15th Meeting of the Expert Panel on Air Pollution Modeling, NATO Committee on the Challenges to Modern Society, Brussels, Belgium, 35 pp
Skibin D (1984) Comments on « on the evaluation of wind power from short wind records ». J Clim Appl Meteorol 23:1477–1479
Slivitzky M, Mathier L (1993) Climatic changes during the 20th century on the Laurentian Great Lakes and their impacts on hydrologic regime. NATO Advanced Study Institute, Deauaille
Subyani AM (1987) Hydrogeology of the Wasia aquifer and its geostatistical modeling. Unpublished M.Sc. Thesis, Faculty of Earth Sciences, King Abdulaziz University, 170pp
Taylor GI (1915) Eddy motion in the atmosphere. Phil Trans R Soc A 215:1
Thiebaux HJ, Pedder MA (1987) Spatial objective analysis. Academic, London, 299 pp

Chapter 6
Spatial Modeling

Abstract In general, *spatial variability* is concerned with different values for any property, which is measured at a set of irregularly distributed geographic locations in an area. The aim is to construct a regional model on the basis of measurement locations with records and then to use this model for regional estimations at any desired point within the area.

Earth sciences phenomenon varies both in time and space; and its sampling is based on measurement stations' configuration. In many practical applications, measured data are seldom available at the point of interest, and consequently the only way to transfer the solar irradiation data from the measurement sites to the estimation point is through regional interpolation techniques using powerful models. The spatial variability is measured in the most common way through the recorded solar irradiation time series at individual points. The square differences between each pairs of spatial variable location values help to construct the spatial autocorrelation function as representative of regional dependence function.

Spatial variability is the main feature of regionalized variables, which are very common in the physical sciences. In practical applications, the spatial variation rates of the phenomenon concerned are of great significance in fields such as in solar engineering, agriculture, remote sensing, and other earth and planetary sciences. A set of measurement stations during a fixed time interval (hour, day, month, etc.) provides records of the regionalized variable at irregular sites, and there are few methodologies to deal with this type of scattered data. There are various difficulties in making spatial estimations originating not only from the regionalized random behavior but also from the irregular site configuration.

Optimum interpolation modeling technique is presented for spatio-temporal prediction of regionalized variable (ReV) with application to precipitation data from Turkey. Kriging method is explained with simple basic concepts and graphics, and then various Kriging application methodologies are explained for the spatial data modeling. The distinctions between simple, ordinary, universal, and block Kriging alternatives are presented in detail. Finally, triple diagram mapping methodology for spatial modeling is presented with applications at five different climate locations.

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_6

Keywords Spatial estimation • Optimum interpolation • Spatial correlation function • Cross-validation • Geostatistics • Kriging • Intrinsic property • Simple Kriging • Ordinary Kriging • Universal Kriging • Block Kriging • Triple diagram • Regional rainfall pattern

6.1 General

Modeling is a procedure that helps researchers, planners, politicians, and many experts alike to make future predictions in time or spatial estimations in a region. It is an interactive procedure where the natural event records (data), modeler's expert view, and prediction (estimation) stages work parallel at times and places in a serial manner for the best conclusions and numerical result. During a modeling procedure, three different worlds are entered in a sequential manner. These are real world affairs, interpretation in conceptual world, and mathematical formulation stage in symbolic world. The real world can be grasped by visual observations, experiments, and numerical sampling. It is, therefore, very useful to make field trips to the study area and have verbal (linguistic) information directly from the local authorities and settlers about the phenomenon and its ReV behaviors. This very first stage includes verbal and numerical data collection. Verbal information can be in the form of logical premises which relate the concerned output variable to various input effective factors. The second stage consists of simple assumptions for the simplification of the phenomenon investigation which is very complex especially in the earth sciences. Among such simplifications procedures are the isotropy, homogeneity, uniformity, and geometrical shape (rectangle, cylindrical, circular, etc.) in addition to a set of necessary assumptions. Depending on the type of the problem, it is possible first to try and apply conventional methods such as scatter plots between input and output variables so as to acquire insight about the internal behavior of the phenomenon, basic histograms, and simple regressions, which may provide some further illuminations about the internal generating mechanism of the ReV functionality. After such basic and simple information, more representative formulations for the modeling are sought and questioned by updating basic assumptions and simplifications. Once a satisfaction is felt, then the formulation is applied with the data available and then the prediction (estimation) is achieved, which must be checked against the observation values (cross-validation). Whatever is the complexity, suitability, and verifiability of the model, there is always a difference between the model output and the measurements. This is the model error and then the model must be trained according to the desired error level. For instance, if the relative error is not less than desired level, say 5 %, then the existing model must be updated either by releasing the basic assumptions or by changing the model formulation structure. Hence, the validation of the model is reiterated according to the previous sequence, and this procedure is repeated until the desired level of error or less is reached. This is tantamount to saying that the model is verified with available numerical and verbal information and finally it is ready for temporal

and/or spatial predictions (estimations) at locations where the data measurements are not available.

The first spatial model is performed by Student (1907) who summed up the number of the particles per unit area instead of the analysis of the spatial positions of the particles in a liquid. He divided the 1 mm^2 area of a hemocytometer into 400 equal squares and counted the yeast cells. Later, Fisher (1935) used the spatial analysis in agricultural area. Yates (1938) searched about the influence of spatial correlation at the randomization process, where completely random regionalized variables are considered for the modeling purposes.

This chapter is concerned with the spatial prediction of the ReV, and two of the most important procedures, namely, the optimum interpolation and Kriging models, are presented in detail.

6.2 Spatial Estimation of ReV

The spatial interpolation of earth sciences data aims at estimating the ReV value at a given site based on the nearby observation (measurements) sites (Fig. 6.1). Most often the input variables are longitudes and latitudes (easting and northing), and the ReV is one of the earth sciences variables, i.e., triplicate as mentioned in Chap. 2.

This is a problem of operational earth sciences that is regularly encountered in the spatial estimation procedures. 2D statistically optimum interpolation models are useful in the analysis and modeling ReV, where spatial dependence functions (SDF) are used for depicting the radius of influence in the case of isotropic ReV or area of influence for anisotropic ReV variability (Chaps. 2 and 4). One of the fundamental, effective, and objective analyses is the transformation of

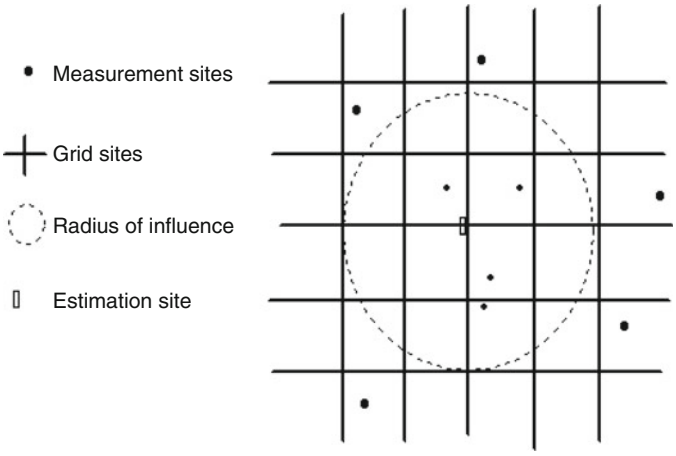


Fig. 6.1 Elements in a spatial estimation

measurements from irregularly distributed sites to regular gridded networks for use in the numerical prediction schemes or mapping methodology through spatial models such as the Kriging methodology, which assigns a spatial estimation value to any point without measurement through a suitable estimation procedure. Hence, in practice, there are twofold purposes in describing the spatial variability in earth sciences ReV by objective analysis:

1. A regular grid of nodes serves as the initial data for numerical forecast models or mapping methodologies.
2. Coordinates of a rather great number of isoline points help to construct equal-value lines (contour lines) such as isohyetal maps.

In any spatial prediction, the problems are with spatial (areal and elevation) modeling and with the transfer of information from available irregular measurement sites to regular grid nodes or to any desired point of interest. In general, ReV at any location bears some relationship to nearby locations. The strength of the relationship normally decreases as the distance between the location increases (Chap. 4, Sect. 4.10). There are various spatial data modeling techniques for data interpolation from measurement sites to any desired point as follows (Schlatter 1988):

1. Surface fitting methods: The first objective analysis method in meteorology was the surface fitting type devised by Panofsky (1949). In this method, the analysis value is represented as a continuous mathematical function, which fits irregularly scattered measurements. Among these methods are the polynomial interpolation (Panofsky 1949), orthogonal polynomials (Dixon 1969, 1976), splines (Fritsch 1971), and, finally, spectral approaches (Flattery 1970).
2. Empirical linear interpolations: The value of any variable at a particular location is estimated as a weighted sum of observations. Among such interpolation techniques are the iterative successive correction methods (Cressman 1959; Barnes 1964), which are already explained in Chaps. 2 and 4.
3. Statistical objective analysis: These are estimation methods at any desired point where spatial correlation (dependence) structure determines the weights applicable to each observation. The major approaches in this category are the optimal interpolation (Gandin 1963), the covariance models for atmospheric variable (Buell 1958), adaptive filtering (Kalman 1960; Şen 1980, 1983), and, recently, the CSV method (Şen 1997).
4. Variational techniques: These include more mathematical abstractions than other methods, and two of them are the incorporation of dynamic constraints (Sasaki 1958) and the fitting models to data at different times (Ledimet and Talagrand 1986).
5. Geostatistical approaches: These models are based on the classical SV and Kriging techniques (Matheron 1963; Clark 1979; Journel and Huijbregts 1978). Various alternatives of SV are presented by different authors (see Chap. 4).
6. Objective modeling: The analysis of ReV produces the optimum solution in the sense that the interpolation error is minimized on the average. This method

allows for the extraction of as much useful information as possible from the measurements. The problems associated with optimum interpolation analysis can be summarized as follows:

(a) It requires knowledge of covariance, which is often not known, and thus its estimate is necessary from the available data. Establishing such an estimate is often fraught with difficulty as a host of local factors are involved.
(b) Essentially, one must determine the priority about which station measurements are significantly correlated with the value at the point of estimation (interpolation), i.e., one must determine a region of influence around the interpolation point as in Fig. 6.1.

6.3 Optimum Interpolation Model (OIM)

A technique commonly used for meteorological analysis is the "optimal interpolation model" (OIM), which is very similar to Kriging method except that some prior knowledge of climatological (obtained from a 6 or 12 h numerical forecast) background field are assumed. Apparently, the optimum interpolation method was suggested first by Eliassen (1954). However, Gandin (1963) developed the method over a number of years, and his name is associated with it in the literature. Eddy (1964) was also one of the early developers of this method. Detailed literature review on this subject is given by Şen and Habib (2001).

In any numerical analysis technique, the main idea is that estimation at any point is considered as a weighted average of the measured values at irregular sites. Hence, if there are $Z_i^o (i = 1, \ldots, n)$ observation stations with the number of prediction site k, then the general prediction in the form of weighted average formula becomes similar to Eq. 2.27 as

$$Z_k^a = \frac{\sum_{i=1}^{n} W_i Z_i^o}{\sum_{i=1}^{n} W_i} \tag{6.1}$$

where Z_k^a is the estimate value and W_i's show the weightings between the ith site and the grid point k (Fig. 6.1). In the literature, all weighting functions that are proposed by various researchers appear as functions of the distance between the sites only. The major drawbacks of such weighting functions are already presented in Chap. 2.

OIM assumes that analyses of data are represented by first-guess data plus corrections, which are linear combinations of differences between the first-guess and observed data. The coefficients of linear combination are determined statistically so as to minimize expected square error of data. The coefficients of linear

combination are expressed by error covariance matrices of observed and forecast data (estimated data), when forecast data are used as "first guess." Thus, the covariance matrix has a great influence on the final estimation. It is assumed that measurements are spatially correlated and such correlations are calculated as in Chap. 4 (Figs. 4.5 and 4.6). This implies that measurements are close together in clusters, i.e., highly correlated and that as they get farther apart they become independent similar to what has been explained in the SV definition in Chap. 4. Although the method presented here is a fully 2D version of optimum interpolation, similar approach is applicable to 3D or to multivariate problems. In all OIM, the following points should be kept in mind:

1. One of the significant advantages in the use of OIM is the ability to estimate a variability at any site from observations at adjacent sites.
2. The interpolation weights depend on the statistical structure of the ReV measured as a sequence of time series and not on the individual measurement values at a given time.
3. The expected analysis errors are produced as a function of the distribution and accuracy of the data.
4. It is more expensive computationally than other commonly used methods. The method is computer intensive as the number of computations and amount of computer storage are concerned.
5. The correlation (dependence) models require a long history of first-guess field for accurate determination of empirical coefficients.
6. The statistical error functions are estimates, not exact values.

Furthermore, the basic assumptions that are embedded in any OIM can be summarized as follows:

1. The measurements have a spatial dependence, which implies that as long as they are close together, the spatial dependence is high; otherwise, they become more independent.
2. There is no dependence (no correlation) between the measurements and the first-guess field errors.
3. OIM as described here may be applied to any scalar field when the correlation and error patterns for that field are known.
4. The field to be analyzed is statistically homogeneous and isotropic.

Consider any grid point, k, with n observations, $Z_i^o (i = 1, \ldots, n)$ around it, which will be used to calculate the analyzed value, Z_k^a, at this grid point. The main idea is that Z_k^a is determined as the sum of a first-guess value, Z_k^g, at grid point k plus a linear combination of the deviations of the observed values from the first-guess values $(Z_i^o - Z_i^g)$ (see Fig. 6.2). This is to say that

$$Z_k^a = Z_k^g + \sum_{i=1}^{n} W_i \left(Z_i^o - Z_i^g \right) \tag{6.2}$$

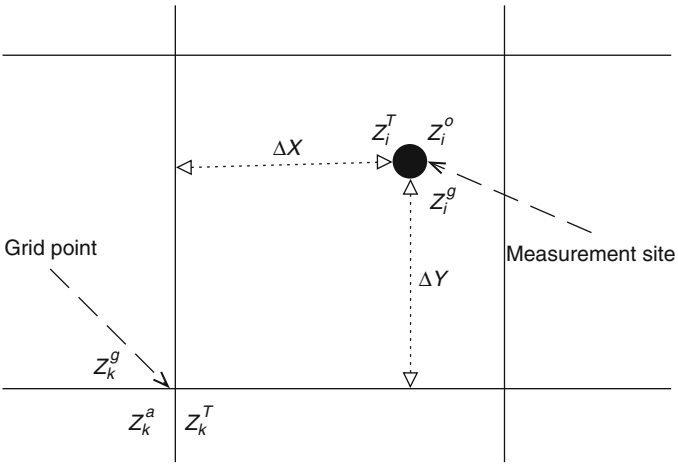


Fig. 6.2 Location of grid point and measurement site

where n is the number of stations (in other words, it may also be considered as the number of influencing stations). The W_i's are the interpolation weights as intuitively mentioned above one would expect the W_i's to be positive and decrease monotonically with increasing distance from the grid point, k.

Assume that the true value, Z_k^T, can be subtracted from both sides of the above equation. Usually, the true value is not known, but some knowledge may be known about its statistical parameters, and that will prove useful in the following method. With these considerations, Eq. 6.2 becomes

$$\left(Z_k^T - Z_k^a\right) = \left(Z_k^T - Z_k^g\right) - \sum_{i=1}^{n} W_i \left(Z_i^o - Z_i^g\right) \tag{6.3}$$

The values of the guess fields at each observation location may be determined by using bilinear interpolation. The difference between each observation and corresponding guess values may be computed as

$$Z_i^o - Z_i^g = \left(Z_i^o - Z_i^T\right) + \left(Z_i^T - Z_i^g\right) = \widetilde{Z}_i + \hat{Z}_i \tag{6.4}$$

where the deviation of the true value Z_i^T from the first-guess value at observed station is $\hat{Z}_i$ (guess error at station i), the deviation of true value from the first guess at grid point is $\hat{Z}_k$ (guess error at grid k), and, finally, $\widetilde{Z}_i$ denotes deviation of true value from the observed value at station i (the observational errors at station i), which can be written as

$$\left.\begin{aligned}\hat{Z}_i &= \left(Z_i^T - Z_i^g\right)\\ \hat{Z}_k &= \left(Z_k^T - Z_k^g\right)\\ \widetilde{Z}_i &= \left(Z_i^o - Z_i^T\right)\end{aligned}\right\} \tag{6.5}$$

By substituting Eqs. 6.4 and 6.5 into Eq. 6.3, one can obtain

$$\left(Z_k^T - Z_k^a\right) = \hat{Z}_k - \sum_{i=1}^{n} W_i\left(\widetilde{Z}_i + \hat{Z}_i\right) \tag{6.6}$$

Furthermore, the interpolation weights W_i are obtained by condition that the mean square error (MSE) of interpolation is the minimum, which corresponds to the method of least squares. This condition can be expressed as

$$E = \overline{\left(Z_k^T - Z_k^a\right)}^2 = \text{Minimum} \tag{6.7}$$

where the over bar denotes an ensemble average in the case of a large grid point number. Insertion of Eq. 6.6 into Eq. 6.7 and evaluation of the square term lead to

$$\begin{aligned}E = &\overline{\left\{\hat{Z}_k - \sum_{i=1}^{n} W_i\left(\widetilde{Z}_i + \hat{Z}_i\right)\right\}^2} = \overline{\hat{Z}_k^2} + \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j \overline{\left(\widetilde{Z}_i + \hat{Z}_i\right)\left(\widetilde{Z}_j + \hat{Z}_j\right)}\\ &-2\sum_{i=1}^{n} W_i\left(\overline{\hat{Z}_k\widetilde{Z}_i} + \overline{\hat{Z}_k\hat{Z}_i}\right) = \overline{\hat{Z}_k^2} + \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j\left(\overline{\widetilde{Z}_i\widetilde{Z}_j} + \overline{\widetilde{Z}_i\hat{Z}_j} + \overline{\hat{Z}_i\widetilde{Z}_j} + \overline{\hat{Z}_i\hat{Z}_j}\right)\\ &-2\sum_{i=1}^{n} W_i\left(\overline{\hat{Z}_k\widetilde{Z}_i} + \overline{\hat{Z}_k\hat{Z}_i}\right)\end{aligned} \tag{6.8}$$

where $i = 1, \ldots, , n$ and $j = 1, .., , n$. By assuming that there is no dependence (no correlation) between the observational and the first-guess field errors, one can obtain

$$\overline{\widetilde{Z}_i\hat{Z}_j} = \overline{\hat{Z}_i\widetilde{Z}_j} = \overline{\hat{Z}_k\widetilde{Z}_i} = 0 \tag{6.9}$$

Consideration of these equations renders Eq. 6.8 to a simplified form as

$$E = \overline{\hat{Z}_k^2} + \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j\left(\overline{\widetilde{Z}_i\widetilde{Z}_j} + \overline{\hat{Z}_i\hat{Z}_j}\right) - 2\sum_{i=1}^{n} W_i\left(\overline{\hat{Z}_k\hat{Z}_i}\right) \tag{6.10}$$

which can be rewritten as follows:

$$E = \gamma_{kk} + \sum_{i=1}^{n}\sum_{j=1}^{n} W_i W_j\left(\gamma_{ij} + \widetilde{\gamma}_{ij}\right) - 2\sum_{i=1}^{n} \mathrm{W}_i\gamma_{ik} \tag{6.11}$$

The elements in this last expression are defined as

$$\gamma_{kk} = \overline{\left(Z_k^T - Z_k^g\right)^2} \tag{6.12}$$

which represents the variance of the guess error at the grid point k, and

$$\gamma_{ij} = \overline{\left(Z_i^T - Z_i^g\right)\left(Z_j^T - Z_j^g\right)} \tag{6.13}$$

is the covariance of the guess error at location, i, and with guess error variance at location, j, as

$$\gamma_{ki} = \overline{\left(Z_k^T - Z_k^g\right)\left(Z_i^T - Z_i^g\right)} \tag{6.14}$$

which is the guess error covariance at the grid point k with the guess error at location i

$$\widetilde{\gamma}_{ij} = \overline{\left(Z_i^o - Z_i^T\right)\left(Z_j^o - Z_j^T\right)} \tag{6.15}$$

This is the covariance of the observation error at the location i with the observation error at station j.

The statistical interpolation is a minimum variance estimation procedure and attempts to minimize the expected analysis error variance. The problem is to find the weights W_i that minimize the variance. Hence, differentiating Eq. 6.11 with respect to each of the weights W_i and equating the result to zero after necessary algebraic manipulations lead succinctly to

$$\sum_{i=1}^{n} W_i\left(\gamma_{ij} + \widetilde{\gamma}_{ij}\right) = \gamma_{kj} \tag{6.16}$$

Interpolation of Eq. 6.2 using weights from Eq. 6.16 is called "optimum interpolation." However, these weights are optimal only if the observation and first-guess error covariance are correct. On the other hand, if the assumed observation and first-guess error covariance are not correct, then Eq. 6.7 is not minimized strictly. In this case, the interpolation weights in Eq. 6.10 are not optimal and it is called statistical interpolation.

Multiplication of Eq. 6.16 by W_j and summation for $j = 1, \ldots, n$ lead to

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\gamma_{ij}+\widetilde{\gamma}_{ij}\right)W_iW_j=\sum_{j=1}^{n}W_j\gamma_{kj} \tag{6.17}$$

Finally, subtraction of this expression from Eq. 6.11 gives the minimum interpolation error as

$$E=\gamma_{kk}-\sum_{i=1}^{n}W_i\gamma_{ik} \tag{6.18}$$

6.3.1 Data and Application

The region of application is Turkey with stations as shown in Fig. 6.3. The grid for this analysis is on a polar stereographic projection oriented along the 35°E longitude line (Habib 1999). There are 16 columns and 8 rows in the array, for a total of 128 grid points.

The data employed in this study is the mean monthly precipitation records collected from the statistics published by Turkish State Meteorological Office. Fifty-two stations for 30-year period from 1960 to 1990 are selected with monthly rainfall measurements.

The spatial scatter distribution of these measurement sites is given in Fig. 6.4, which shows an irregular pattern. In general, the transfer of information from the measurement sites to grid points is necessary for any type of modeling as numerical solution or mapping the ReV. Here each site monthly time series, which may also

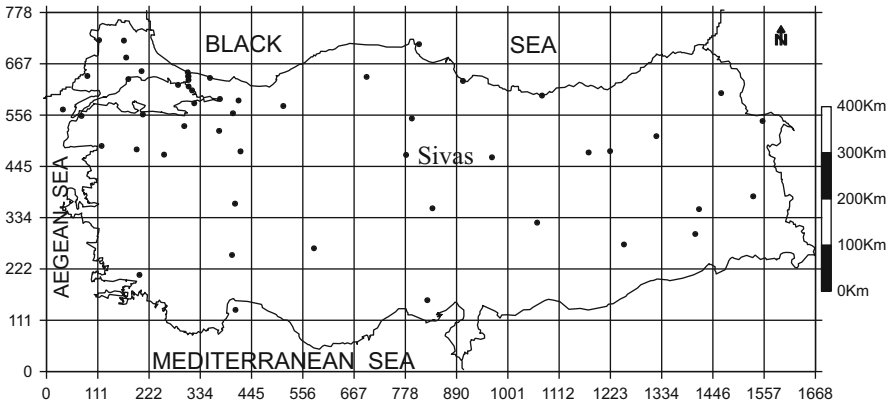


Fig. 6.3 Grid points and measurement sites in Turkey

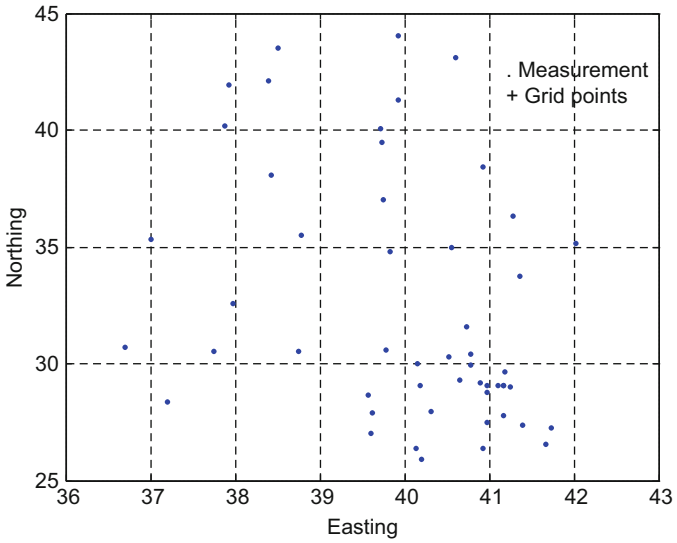


Fig. 6.4 Spatial distributions of the data points

be monthly average groundwater levels, groundwater quality values (calcium, magnesium, sodium, potassium, sulfate, bicarbonate, chloride, nitrate, total dissolved solids, electric conductivity, etc.), and alike.

In practical applications of the optimum interpolation such as the analysis of rainfall, one uses the climatological mean as the first-guess value. Hence, the following expression is considered for the interpolation point value (Habib 1999).

$$Z_k^a = \overline{Z}_k + \sum_{i=1}^{n} W_i \left(Z_i^o - \overline{Z}_i\right) \tag{6.19}$$

where $Z^a{}_k$ and $Z^0{}_i$ are the calculated and observed rainfall values corresponding to the arithmetic averages of $\overline{Z}_k$ and $\overline{Z}_i$, respectively, at interpolation point k and observation stations, $i = 1, 2, .., , n$; and W_i's are the interpolation weights. In order to calculate these weights, the interpolation formula can be obtained by multiplying both sides of Eq. 6.19 by $(Z^0{}_j - Z_j)$ and taking the expectation, leading to

$$\sum_{i=1}^{n} W_i \rho_{ij} = \rho_{kj} \tag{6.20}$$

where ρ_{ij} is the spatial correlation coefficients between stations i and j and ρ_{kj} between stations j and k. This equation is valid in the case when $E(Z_k) = E(Z_k{}^a)$, which implies unbiased estimator. In short, interpolation weights, W_i, are dependent on the statistical structure of the spatial correlation function (SCF) of the rainfall records at irregular sites. Once the SCFs are obtained from the available data, then

the value of ρ_{kj} can be read from this function depending on the distance between k and j, and consequently, the only unknowns in Eq. 6.20 are the weights which can be calculated from the set of n linear equations. The correlation functions are presented in the previous chapter (Figs. 4.5 and 4.6). The expected analysis error, ε_{kj}, at grid point k that results from the introduction by using information at location j can be expressed as (Habib 1999)

$$\varepsilon_k = 1 - \sum_{i=1}^{n} W_i \rho_{ki} = 1 - \rho_{kj} \tag{6.21}$$

Most often in practice, $0 < \rho_{kj} < 1$ and therefore $0 \leq \varepsilon_{kj} \leq 1$. It is obvious that the expected error does not depend directly on the observed values but again on the spatial statistical structure of the rainfall amounts. Under the light of the aforementioned discussion, the following OIM steps are necessary for practical applications:

1. Specify the geographical locations (longitude and latitude) of interpolation points.
2. Specify the estimation locations.
3. Compute a first-guess field value (climatologic means, i.e., arithmetic averages) for each station locations.
4. Compute the background error correlations, which correspond to the differences between observed and average values.
5. Find a suitable model for the background error SCF.
6. Select the measurement sites that will influence the interpolation point.
7. Solve the system in Eq. 6.20 and obtain the interpolation weights.
8. Compute the interpolation point value by using Eq. 6.19.
9. Calculate the expected analysis error at the interpolation point by using Eq. 6.21.
10. Repeat steps (5–8) for all desired interpolation points. Figure 6.5 shows the necessary steps in the execution of a regional modeling procedure according to OIM approach.

6.3.1.1 Spatial Correlation Function (SCF)

Although there are several models to represent the correlation by continuous functions, the basic modeling form is adapted as "the negative exponential" for all monthly spatial correlation types. The full correlation array is reduced to distance-interval-averaged values (Fig. 4.6). Hence, most of the heterogeneity is averaged out in the computed lag correlation and computation time is greatly reduced. In general, the negative exponential model has the following mathematical expression:

$$R(d) = a + be^{-d/c} \tag{6.22}$$

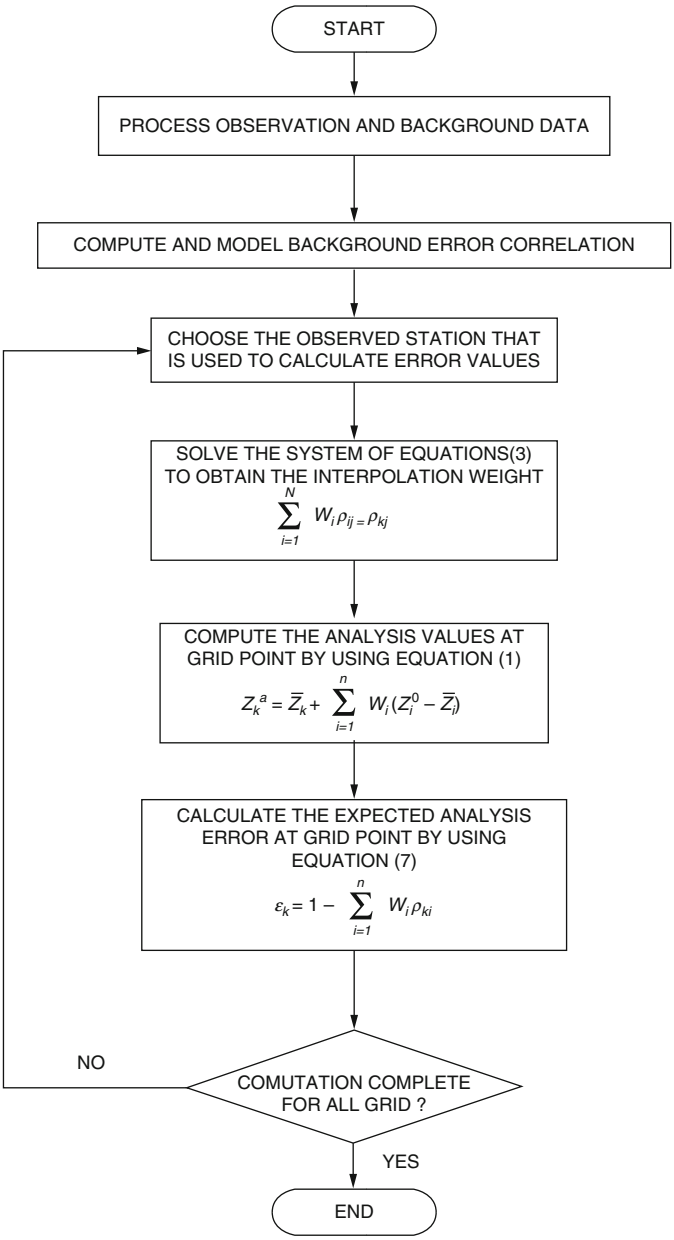


Fig. 6.5 OIM execution steps

Table 6.1 Spatial correlation model parameters

Month	*a*	*b*	*c*
January	−0.15553	0.90883	1352.15221
February	0.27266	0.55969	270.50807
March	0.00000	0.85671	575.8087
April	0.12362	0.71309	255.97296
May	0.06599	0.78690	220.55975
June	−0.06992	0.62878	506.51055
July	0.11930	0.65853	131.18509
August	−0.08718	0.61655	653.76575
September	−0.29293	0.95016	909.51275
October	0.11815	0.65535	370.6699
November	−0.33159	0.82909	1707.53125
December	0.03030	0.51511	636.63236
Average	−0.01650	0.72650	623.2176

where $R(d)$ represents the theoretical SCF and d is considered as the distance between ith and jth stations; a, b, and c are the model parameters. These parameters are determined by fitting a mathematical function to the array of computed correlation versus corresponding distances between measurement sites. Each one of the monthly average SCF is fitted to this model by employing the ordinary least squares regression approach. The resulting parameter values for each month are presented in Table 6.1 in addition to the overall monthly average parameter values.

All the monthly spatial correlation functions (SCF) for, say, Sivas station (see Fig. 6.3) are confined between January and July as one can see from Fig. 6.6. It is also clear from this figure that summer months have less SCF values than winter months.

This is due to the fact that in this region cyclonic rainfall types are dominant in winter, whereas convectional type of rather local rainfall occurs in summer season. Of course, orographic rainfall effects play more effective role in winter season than summer. However, it is not possible to identify this type of rainfall from the SCFs. In Fig. 6.6, one can note that the maximum continuity appears in winter months. This is due to the fact that in this region frontal rainfall types are dominant in winter, whereas convectional type of rather local rainfall occurs in summer season.

At large distances, February has strong persistent SCF, which implies that in this month there are cyclonic rainfall types because they cover large areas. Similar trends are also observed in April and May. On the other hand, comparatively steeper SCFs indicate convectional rainfall types, which appear over rather smaller area during summer months.

It is possible to show the variation of SDF for any measurement station as the center of concentric contour. The extreme and average SCFs are shown for Sivas City as given in Fig. 6.7.

The benefit from these figures is that at any given month of the year and a given correlation level, say 0.050, it is possible to determine the influence area around the

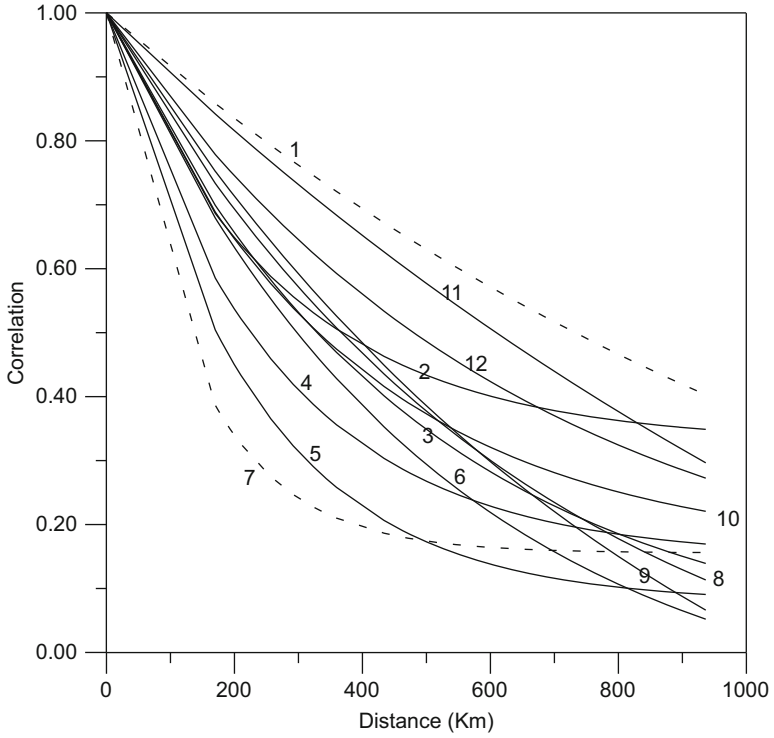


Fig. 6.6 Spatial correlation function of Sivas City for different months

station. On the average, the minimum correlation value for 200 km (other distances have similar interpretations) appears in July, and the maximum correlation value is in January. It is also clear from this figure that summer months have less spatial correlation values than winter months. Table 6.2 shows the values of the percentage of the variance of observational error which has a minimum of 14.32 in March and a maximum of 50.46 in November. Similarly, the correlation of the observed values varies between a minimum of 0.4975 in November and a maximum of 0.8567 in March.

It is possible to find the correlation $\rho(d)$ of the true (first-guess) values of the meteorological variables at distance d as

$$\rho(d) = \frac{R(d)}{R(0)} \tag{6.23}$$

where $R(d)$ and $R(0)$ are the correlations of the observed values at d and at zero distances, respectively. This equation is used to calculate the correlation of the

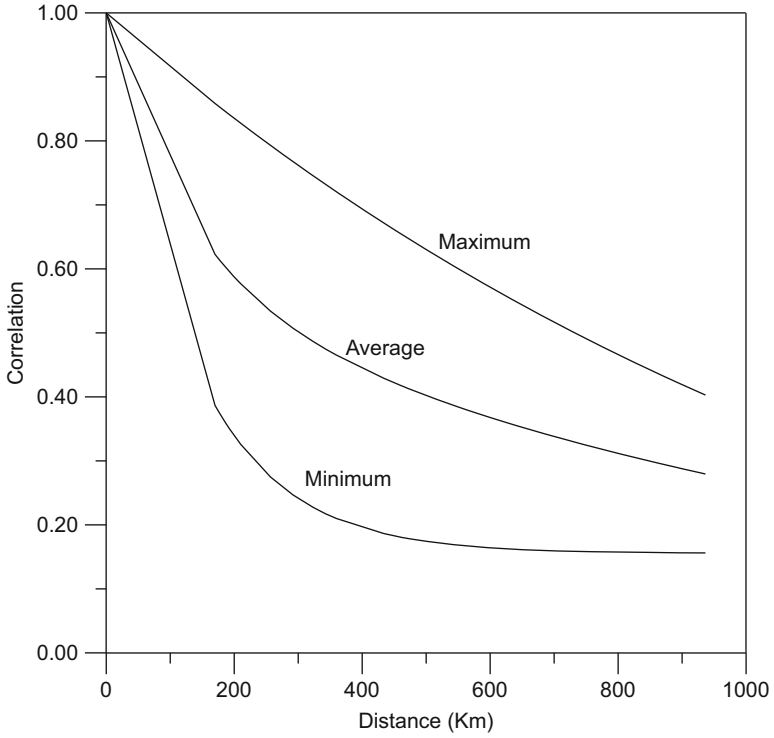


Fig. 6.7 Sivas correlation contour lines

Table 6.2 Some pertinent statistics

Month	R(0)	Percentage of σ_e^2
January	0.7533	24.67
February	0.8323	16.76
March	0.8567	14.32
April	0.8367	16.32
May	0.8528	14.71
June	0.5588	44.11
July	0.7775	22.24
August	0.5293	47.06
September	0.6572	34.27
October	0.7735	22.65
November	0.4975	50.25
December	0.5454	45.46
Average	0.7059	29.40

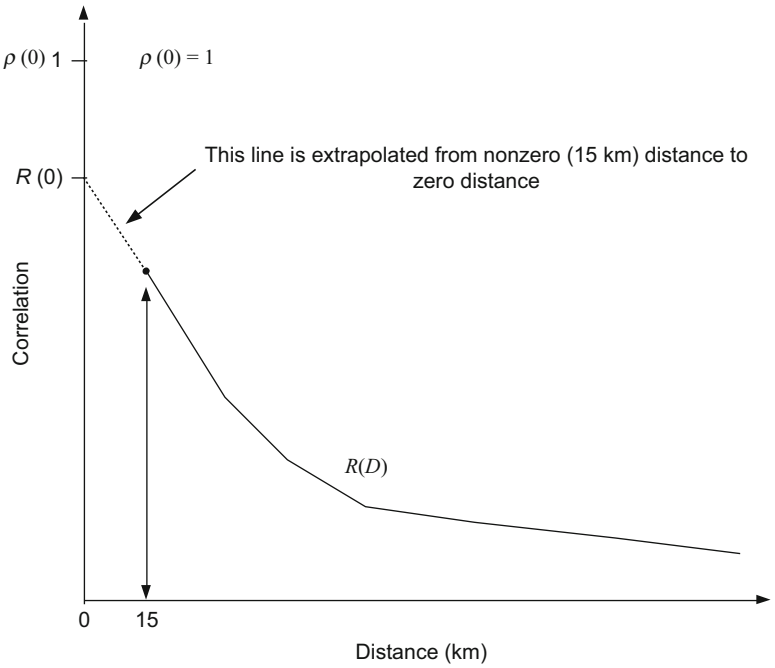


Fig. 6.8 Estimation of zero intercept value

first-guess (true) values of the ReV. The negative exponential model in Eq. 6.22 at zero distance can be defined as

$$R(0) = a + b \tag{6.24}$$

and, finally, by substituting Eqs. 6.22 and 3.24 into Eq. 6.23, it is possible to obtain

$$\rho(d) = \frac{\left(a + be^{-d/c}\right)}{a + b} \tag{6.25}$$

This expression can be used for the indirect evaluation of the correlation of the first-guess (true) values of the ReV by means of extrapolating the correlation of the observed values at zero distance $d = 0$, as shown in Fig. 6.8 or by using Eq. 6.23 or Eq. 6.25 for finding the correlation of the true values.

6.3.1.2 Expected Error

The expected analysis error is a by-product of the OIM analysis procedure as already stated in Eq. 6.18. It is computed at each analysis point and is a function of the amount, quality, and distribution of the data near each point. If no data are

present, the expected error remains unchanged from its initial value. The calculation and modeling of errors result in an analysis that is not necessarily optimum (Lorenc 1981). The expected error is thus a measure of what the analyzer "thinks" the error is (Schlatter and Branstator 1987).

Figure 6.9 includes maps of the expected error of rainfall calculated from Eq. 6.18 for whole months at Sivas station. It is obvious that there are a number of areas where expected error changes quite rapidly. For instance, there is a bull's-eye feature in all months, which is enclosed by the minimum expected error. There is also a similar area of less circular contours surrounding the region at the north center of the study area.

At the first glance, it is possible to depict that contours are less densely occurring over terrestrial area during September–February winter period. However, the contours are very dense from March to October. The reason for such a digression is due to the fact that the rainfall occurrences are comparatively more sporadic (i.e., regionally random) during March–October duration. This is also because of the convective rainfall occurrences appearing almost independently from each other. However, areal continuity of winter rainfall is a signature of cyclonic weather movements. Such continuity results in comparatively very small expected error amounts, say, for instance, in January (Fig. 6.9) where errors vary between 0.05 and 0.15 over Anatolia within Turkey. On the contrary, in July (Fig. 6.9), error band varies between 0.20 and 0.70. Table 6.3 shows the relation between expected mean square error (MSE) and months.

One can note that the maximum expected error appears in summer months (e.g., in July, 0.3480) because the rainfall has more discontinuous ReV as explained above. On the other hand, the minimum expected error is in winter months (e.g., in January, 0.0193) because the rainfall in this season is really extensive and more continuous.

6.3.1.3 Data Search and Selection Procedure

In order to find the weights W_i, the inversion of the correlation (covariance) matrix must be computed at each grid point. The size of the correlation matrix is directly related to the number of the observations permitted to influence each grid point. The choice of a search strategy that controls the stations which are included in the interpolation procedure is an important consideration in any approach to OIMs.

The most common approach in choosing the stations that contribute to the interpolation is to define a search neighborhood within which all available stations will be used. Herein, a simple search strategy is adapted using all station within a circular search neighborhood with a limited radius of influence.

Meleshko and Prigodich (1964) have shown that the interpolation error reaches a minimum at about six to eight measurements and shows no further improvement with the inclusion of more sites. In order to fix this idea with the data at hand, the change of expected MSE is plotted versus number of neighboring station for each month. However, it appeared that such graphs are very similar to each other. It is

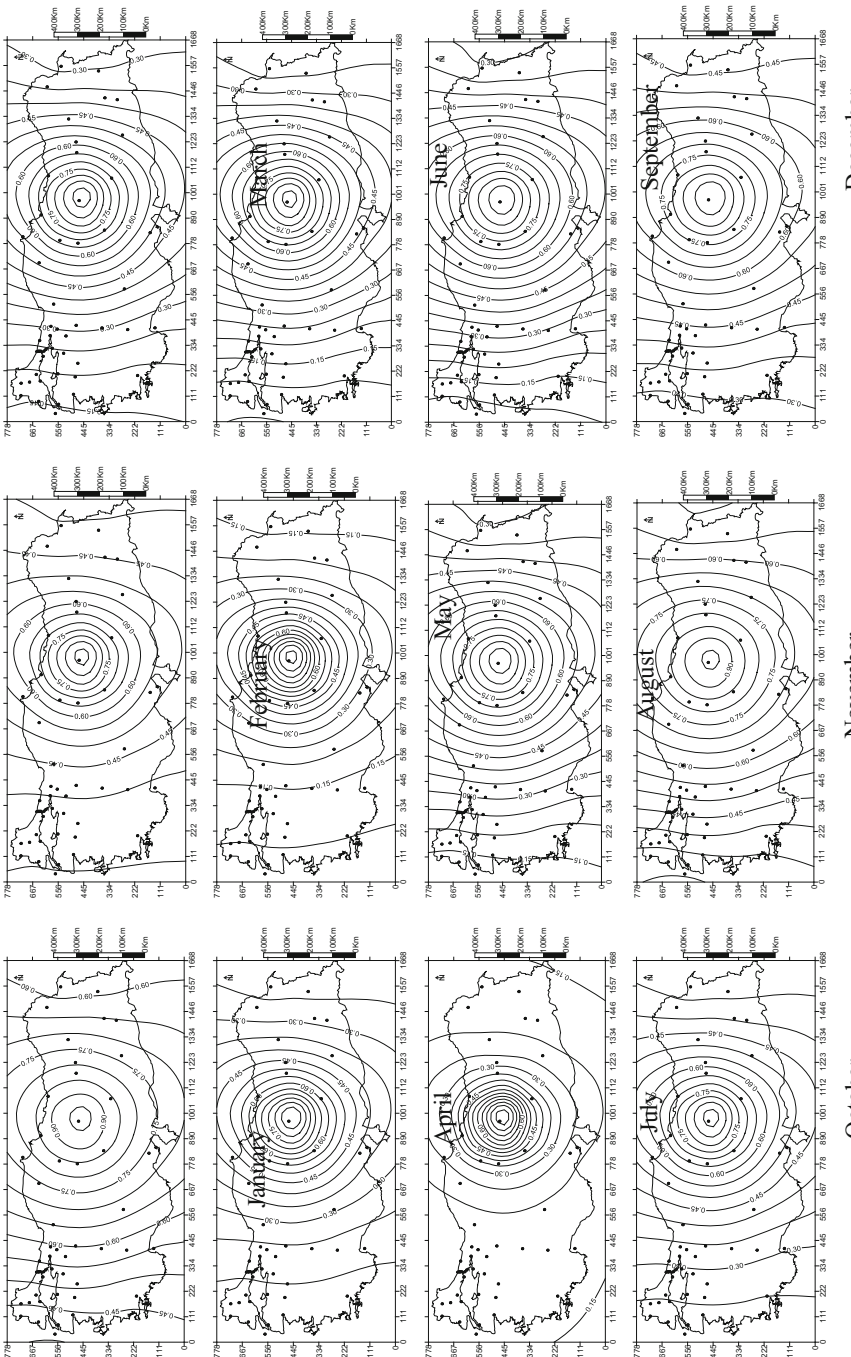


Fig. 6.9 Equal correlation lines between Sivas City and other stations

Table 6.3 Monthly expected errors

Month	MSE
January	0.0193
February	0.0935
March	0.0834
April	0.1700
May	0.2500
June	0.1000
July	0.3480
August	0.1200
September	0.0657
October	0.0994
November	0.0238
December	0.0463
Average	0.1141

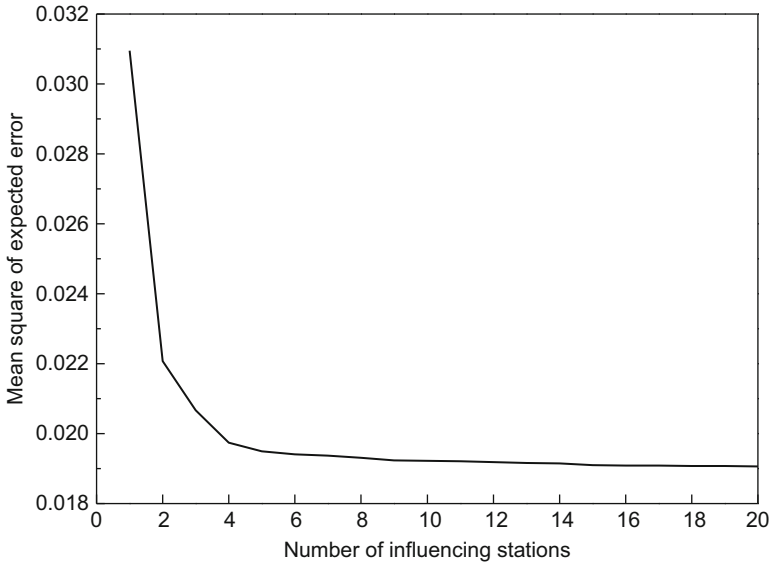


Fig. 6.10 Relationship between number of influencing stations an

objectively seen from Fig. 6.10 that on the average the number of influencing stations does not change significantly above four sites. For more station numbers, the expected error mean square remains on almost the same minimum level.

6.3.1.4 Cross-Validation of the Model

The accuracy of the OIM is investigated with the help of numerical experiments from the surrounding sites and in comparison to the interpolated values with the

Table 6.4 Univariate statistics of optimum interpolation analysis

Parameters	Observation	Estimation
n	52	52
Mean	76.6	74.70
Minimum	5	0
Maximum	361	266.9
Range	356	267
Variance	3407.60	2905.61
Std. dev.	58.37	53.90

measurements (Gandin 1965). Similarly, cross-validation techniques allow comparison of estimation and true values using only the information available in the measurements (Isaaks and Srivastava 1989). In a cross-validation technique, the interpolation method is tested at the measurement site, and the measurement value at a particular station is discarded temporarily from the surrounding sites. Once the estimate is calculated, it is possible to compare it with the measurement value that is initially removed from the station. This procedure is repeated for all available sites, with comparison of observed, interpolated, and actual values by using simple statistical parameters such as the means and standard deviations. The statistical parameters of the comparison between various estimation methods and of the true values are summarized in Table 6.4 for January 1984.

In general, the difference between the average estimate from the OIM and the true average reflects a tendency toward overestimation or underestimation. However, in this case, the analysis method has a slight tendency toward overestimation as seen from Table 6.4. The estimations have less variability than measurements. However, this reduced variability of analysis values is often referred to as smoothing. Therefore, the measurements are less smoothed than the analysis values. A scatter plot of measurements versus estimations provides additional evidence on how well an estimation method has been performed. Figure 6.11 shows scatter plot of measurements versus estimations for January 1984, which indicates that there is a good relationship between estimations and measurements especially for less than 200 mm. This is also supported by the root MSE, which is 29.9, and the mean percentage error is −9.07 %. Furthermore, the square of the correlation which is a measure of the variance explained by the model is equal to 0.76.

In order to see the correspondence matching between measurements and estimations as well as monthly rainfall amounts, the monthly statistical parameter variations are presented in Figs. 6.12 and 6.13. It is seen that on the basis of averages, measurements and estimations are very close to each other with less than 1 % error. Although there are more discrepancies for monthly standard deviations, generally 2.5 % but high errors appear in November and December.

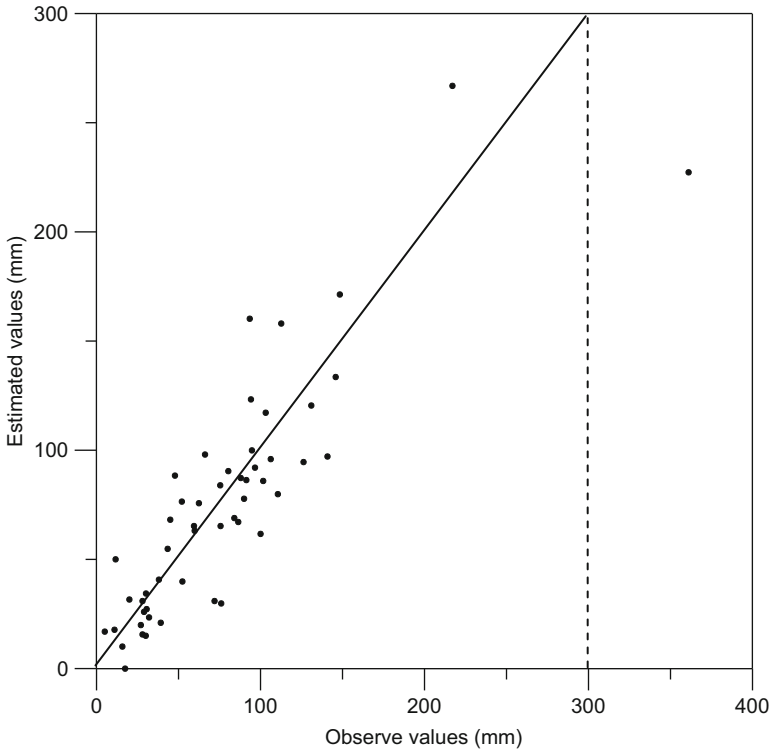


Fig. 6.11 Scatter plot of observed versus estimated values for January 1984

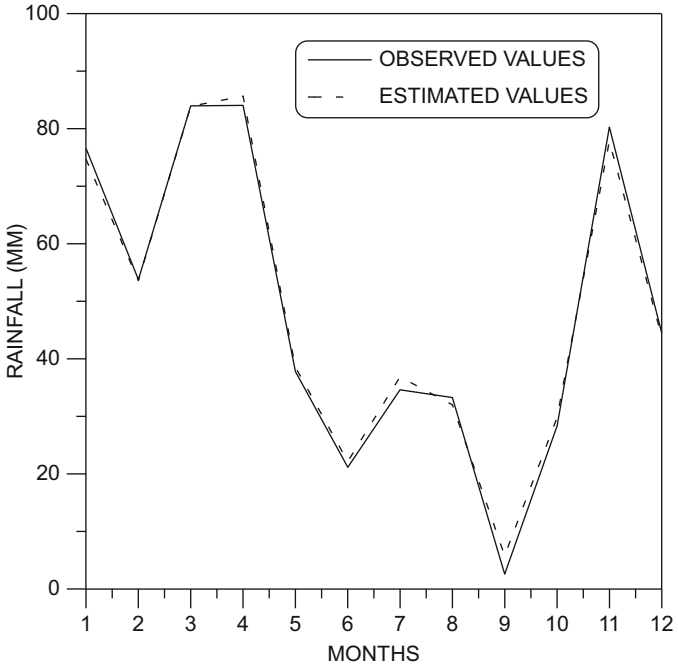


Fig. 6.12 Arithmetic mean of observed and predicted monthly values

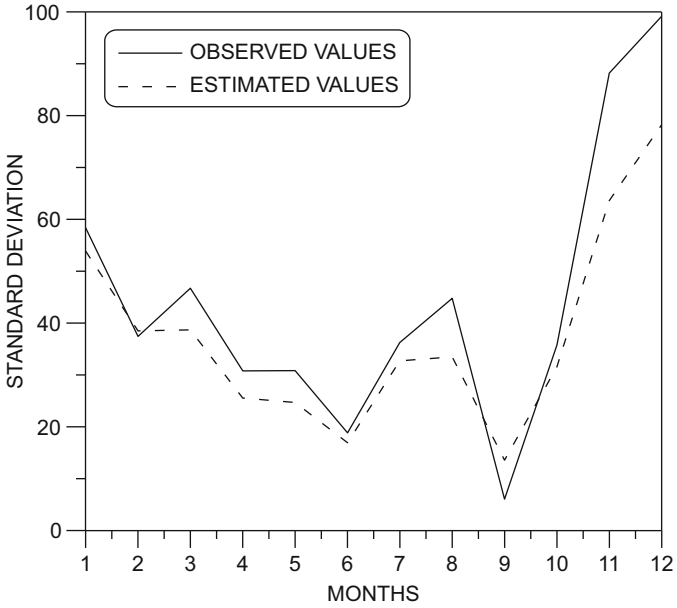


Fig. 6.13 Standard deviation of observed and predicted monthly values

6.4 Geostatistical Analysis

It is a branch of applied statistics with the original purpose of spatial data processing for estimating changes in ReV. However, the principles have been applied to a variety of areas in geology and other scientific disciplines. A unique aspect of geostatistics is the use of ReV, which are variables that fall between random and completely deterministic behaviors (Chap. 1). ReV describe phenomena with geographical distribution (e.g., elevation of ground surface) and exhibits spatial continuity, which is not always possible to sample at every location. Therefore, unknown values must be estimated from data taken at specific locations. The size, shape, orientation, and spatial arrangement of the sample locations are termed as the support and influence the capability to predict the unknown samples. If any of these characteristics change, then the unknown values also change. The sampling and estimation of ReV are done, so that a pattern of variation in the phenomenon under investigation can be mapped such as a contour map for a geographical region. In addition to spatial and temporal variation in nature, spatial and temporal dependence or continuity also occurs. Geostatistics include a set of statistical procedures that can be used to analyze and model spatial relationships in nature. The spatial variation is somewhat similar to the time variation in some properties, time observations are dependent and occur in equally spaced time intervals in 1D, whereas the spatial observations are dependent and occur in 2D or 3D, and the

distance between points are not necessarily equal intervals as they are in time variations. In earth sciences, most of the natural phenomena behave simultaneously in space and time as soil moisture content, water level, and precipitation data may consist of long time series at various locations. The spatiotemporal variation of the natural phenomena implies a significant amount of uncertainty. Moreover, this variation is, in general, nonhomogeneous in space and nonstationary in time.

Matheron (1963) was the first to use the term "geostatistics" which was defined by him as "the application of the formalism of random function to the reconnaissance and estimation of natural phenomena." The basis of the method is the ReV theory and concepts of this theory can be split into two main successive steps:

1. Establish the theoretical basis for expressing the structural properties of a natural phenomenon in a useful mathematical form as SV or CSV.
2. Provide a particular means for solving various problems of estimation such as the Kriging methodology, which guarantees a solution to the estimation problem (Kriging) and deals with the ReV by using the probabilistic theory of random function.

By now, however, there are number of excellent books on the subject including both introductory (Clark 1979) and advanced aspects (David 1977; Journal and Hujibregts 1978). The classical parametric and nonparametric geostatistics are introduced recently for application in the field of earthquake ground motion evaluation (Glass 1978; Carr and Glass 1984, 1985; Carr et al. 1985).

Almost all variables encountered in the earth sciences can be regarded as ReV. For seismic zonation, this is an ideal one in describing earthquake ground motions. Each observation of ground motion can be considered simply as a unique realization of a Rev, which adequately represents local random behavior tempered by global attenuation. Furthermore, provided that a valid SV can be developed, Kriging is well suited for the estimation process to result in data regionalization.

6.4.1 Kriging Technique

Geostatistics, which is also referred as ReV, was basically developed for statistical study of the natural phenomena in mining field including other spatial phenomena in earth sciences (Matheron 1963; Davis 1986). The natural phenomena that have spatial distribution vary from one place to another with apparent continuity structure. The first step in geostatistics is to study and determine the spatial structure and dependence from measurement characteristics of the natural phenomena (Chap. 4). The spatial structure can be established by SV, CSV, or PCSV as convenient. Using the underlining generation structure, it is possible to estimate the characteristics at unsampled locations and extend findings of the regional behavior in natural phenomena through Kriging techniques.

In any geostatistical modeling, there are two stages as the estimation of SV (CSV or PCSV) from a given set of regionalized data for extraction of spatial dependence

and then Kriging methodology, which provides estimates for locations where no data are available. The SV is the source of information used in Kriging to achieve optimal weighting functions for mapping. Kriging uses theoretical SV in calculating estimates at a set of grid nodes. These Kriging estimates are best linear unbiased estimator (BLUE) of the ReV at a set of locations, provided that the surface is stationary and the correct form of the theoretical SV has been determined (Chap. 4).

In many disciplines such as petroleum exploration, mining, and groundwater and water pollution, analysis data are available at a set of predetermined spatial locations (water and oil wells, meteorology stations, etc.). The purpose is to make regional estimation at any location based on the available data at these locations. It is necessary to have often maps based on a regular grid, and the estimations are used to produce 2D contour maps or 3D surface plots. In theory, only Kriging method of grid generation can produce better estimates (in the sense of being unbiased and having minimum error). In practice, the effectiveness of the technique depends on the correct specification of several parameters that describe the SV and the model of the drift (regional trend). However, because Kriging is robust, even with a naive selection of parameters, it will do no worse than conventional grid estimation procedures (Chap. 2).

The price that must be paid for optimality in estimation is computational complexity compared to techniques presented in Chap. 3. A large set of simultaneous equations must be solved for every grid node estimated by Kriging. Therefore, computer run times will be significantly longer if a map is produced by Kriging rather than by conventional gridding. Kriging can be computationally very intense but increasingly available in software packages, and it is the best method for many purposes. In addition, an extensive prior study of the data must be made to test for stationarity, determine the form of the SV, set the neighborhood size, and select the proper order of the drift, if it ever exists. These properties are not independent, and because the system is underdetermined, trial-and-error experimentation may be necessary to determine the best combination. For this reason, to warrant the additional costs of analysis and processing, Kriging probably should be applied in those instances where the best possible estimates of the surface are essential, the data are of reasonably good quality, and estimates of the error are needed.

6.4.1.1 Intrinsic Property

Mathematically speaking, let x be a point in space and $Z(x)$ the value of the function at point x. It is usually highly variable and noncontinuous and therefore cannot be estimated directly. Thus, the structure with variation will be studied by examining its increment. The basic idea of the theory is to consider the function $Z(x)$ as one realization of random function from the exhaustive ensemble. This is tantamount to assuming a well-defined and unique sequence of numerical values into a realization of random process. Only one realization of that process is available, and the problem is to find the characteristics of the ReV in order to make the estimation

of unknown points possible. Obviously, it is not rigorously possible to infer the probability law of any random function from its single realization because it has limited finite number of sample points. Thus, many realizations $Z_1(x), Z_2(x), \ldots,$ $Z_k(x)$ of the ReV are required for identifying the probability law. In order to render the situation into a practically tractable one, the assumptions of spatial homogeneity and stationarity are needed. A ReV is said to be strictly stationary if any conventional statistical properties of the medium, such as the mean, variance, covariance, etc., do not change with distance. This assumption is seldom encountered in natural phenomena. Therefore, a ReV is to be first-order stationary when the random variable has the same mean value irrespective of location. However, in the linear geostatistics, it will be enough to assume first of all that the weak or second-order stationarity exists. The second-order stationarity assumption consists of the two following conditions:

1. The arithmetic average, m (expected value for any term), value of the ReV is the same all over the area

$$E[Z(x)] = m \tag{6.26}$$

Or, equivalently,

$$E[Z(x) - Z(x+d)] = 0 \tag{6.27}$$

2. The spatial covariance of the ReV is the same all over the field of interest

$$\mathrm{Cov}\{[Z(x) - m][Z(x+d) - m]\} = \mathrm{Cov}(d) \tag{6.28}$$

where d is the distance separating between two sites at locations x and $x + d$.

If it is assumed that for every d this spatial covariance is independent of the location, then the second-order stationarity is valid. By holding the full second-order stationarity, the covariance $\mathrm{Cov}(d)$ approaches the variance σ_Z^2 of the random variable as $d \to 0$

$$\mathrm{Cov}(0) = \{[Z(x) - m][Z(x) - m]\} = E\left\{[Z(x) - m]^2\right\} = \mathrm{Var}[Z(x)] = \sigma_Z^2$$

A second-order stationarity is, however, really justifiable, and a weaker assumption may be adopted instead. The mean value, m, is always unknown and may not be constant, so that the variance and covariance cannot be computed directly. It is, therefore, beneficial to define an alternative statistic, which does not require the mean value. For instance, successive difference as in Eq. 6.27 may be zero but its square is not essentially zero. Considering these characteristics, a consistent set of assumptions weaker than second-order stationarity have been made and called by

Matheron (1963) as the "intrinsic hypothesis" which subsumes the following expressions:

$$\mathrm{Var}[Z(x) - Z(x+d)] = E\Big\{Z(x) - Z(x+d) - E[Z(x) - Z(x+d)]^2\Big\}$$

Consideration of Eq. 6.27 leads to

$$\mathrm{Var}[Z(x) - Z(x+d)] = E[Z(x) - Z(x+d)]^2 = 2\gamma(d) \tag{6.29}$$

where $2\gamma(d)$ is the variogram and $\gamma(d)$ is the SV or half-variogram. It should be noted that second-order stationarity implies the intrinsic hypothesis but the converse is not true.

In addition, under the hypothesis of second-order stationarity, the covariance and SV are two equivalent tools that characterize the autocorrelation between two variables $Z(x+d)$ and $Z(x)$ separated by d as

$$\gamma(d) = \mathrm{Cov}(0) - \mathrm{Cov}(d) \tag{6.30}$$

The validity of this expression is true especially if the ReV has a Gaussian (normal) PDF. Estimation of the SV is preferable to estimation of the covariance because the experimental SV does not require a prior estimate of the population mean. Besides, SV calculation requires a set of measurements with no time variation. Under the same condition, the relationship between the model autocovariance, $\rho(d)$, and SV is

$$\gamma(d) = \sigma_Z^2[1 - \rho(d)] \tag{6.31}$$

6.5 Geostatistical Estimator (Kriging)

After successful matching of theoretical SV model, one can use Kriging to estimate what weights should be applied to each surrounding point to estimate the value of the ReV at an unknown location (e.g., a grid node, see Fig. 6.1). Kriging calculates the weighted average of neighboring values and assigns it to each grid node. Kriging can also calculate the error variance for each estimated value, which means that it is possible to know what the most likely value is at each location and how likely that it is to be the actual value. In its most basic form, Kriging does not work where there is a trend in the field (high values on one end and low on another). Universal Kriging addresses this limitation by first calculating a trend surface and then using Kriging to estimate the values of the difference between the real surface and the trend surface (called residuals). The problem of local estimation appears in finding the best estimator of the mean ReV value over a limited domain.

The use of Kriging estimations is advantageous because it is not limited to a simple point estimation of the region magnitude (Delhomme 1978; Marsily 1986).

Kriging is the name given to a local estimation technique. Provided that the basic assumptions of no trend and a model for the SV (or CSV) are known or determined in some way from the data, then the Kriging produces always the best linear unbiased estimator (BLUE) of the unknown ReV characteristics. In the case of nonstationarity, several non-bias conditions are required, and this leads to the technique known as "universal Kriging" or unbiased Kriging of order, say, n. When groups of variables are correlated, the co-Kriging technique can be used to estimate any ReV (e.g., response spectra) from the data available on all the correlated variables (e.g., peak acceleration, velocity, and displacement in addition to the response spectra). Other classes of estimators are also used in practice, for instance, nonlinear estimator can be built by prior transformation of the data. Disjunctive Kriging (Matheron 1971), Lognormal Kriging (Krige 1951), and indicator Kriging (Journal 1983) are examples of nonlinear estimators. All Kriging techniques are based on the simple linear models as

$$Z_E = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n \tag{6.32}$$

where Z_E is the estimator of the true value at location E, and λ_i are the weights allocated to each observation such that

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{6.33}$$

The technique minimizes estimation variables by solving a set of Kriging equations, which include covariance between the point or volume to be estimated and the sample points and covariance between each pair of sample points. The calculated weights after the solution of the equation system depend on various factors such as the size and shape of volume estimation, the distance and direction of each sample, the distances between sample locations and the SV or CSV.

The equations and their full derivation as well as attributes and advantages of Kriging can be found in numerous publications such as Matheron (1971), David (1977), Journel and Huijbregts (1978), and Clark (1979). The following points are some of the characteristics and advantages of Kriging technique:

1. If one has a model for the SV (or CSV), he/she can produce the minimum variance using the Kriging technique.
2. If the proper models are used for the SV (or CSV) and the system is set up correctly, then there is always a unique solution to the Kriging system.
3. If one has regular sampling, and hence the same sampling set up at many different positions within the region, it is not necessary to recalculate the Kriging system each time.

4. Kriging is not limited to a single point estimation of the given magnitude Z but can also be used:
   (a) To estimate the mean value, $\overline{Z}$, on a given block, e.g., on the mesh of model or a sub-domain of any shape of watershed
   (b) To obtain the estimation variance of magnitude Z, i.e., roughly the confidence interval of this estimation
   (c) To locate the best situation for a new measurement point, e.g., by minimizing the overall uncertainty in the field under consideration
5. Kriging is advantageous because it considers the following points explicitly:
   (a) The number of spatial configuration of observation points within the study region
   (b) The position of the data points within the region of interest
   (c) The distance between the data points with respect to area of interest
   (d) The spectral continuity of the interpolated variable

For instance, in an earthquake ground motion study, Kriging can be used for evaluating the earthquake ground motion hazard within the region of interest and for estimating the tripartite earthquake response spectra at a site of interest. It is already stated that potential errors in the data collection may lead to nugget effect (Chap. 4), which becomes evident especially from the sample SV, and its existence causes smoothing operation through the Kriging and less confidence in individual data points versus the overall trend of the data. It has the same unit as the SV. There are two components that give rise to nugget effect, namely, variances due to error, σ_e^2, and separation, σ_s^2. The latter variance is a measure of variation that occurs at separation distances of less than the typical nearest neighbor sample spacing. The more the random fluctuation of the same data at a given location is, the greater the error variance and the lesser the prediction reliability. Consequently, Kriging tends to smooth the surface, and therefore, it is not a perfect estimator.

6.5.1 Kriging Methodologies and Advantages

Once the degree of spatial dependence has been established, then the SV can be used to interpolate values for points not measured through the process of Kriging, which is an interpolation method that uses the sample (empirical, experimental) SV to weight sample points based on their locations in space relative to the point value that is to be estimated. Therefore, a first step in Kriging is to fit a theoretical function to the SV model that describes the theoretical SV. Kriging has many things in common with traditional point interpolation methods such as inverse distance (or square) weighting, triangulation, polygonization, methodologies, etc.

which are explained in Chap. 2. The result of Kriging is an interpolated surface map of ReV. There are several types of Kriging methodology:

1. Ordinary Kriging: It is similar to multiple linear regression and interpolates' values based on point estimates.
2. Indicator Kriging: It is used to estimate indicator variables. Indicator Kriging is simply an ordinary Kriging performed on the indicator-transformed data.
3. Punctual Kriging: Both ordinary and indicator Kriging methodologies are forms of punctual Kriging because they are used to estimate values for exact points within the sampling unit. Punctual Kriging is the most common method used in earth sciences. It helps to estimate the value at a point from a set of nearby sample values using Kriging. The kriged estimate for a point will usually be quite similar to the kriged estimate for a relatively small block centered on the point, but the computed Kriging standard deviation will be higher. When a kriged point happens to coincide with a sample location, the kriged estimate will equal the sample value.
4. Block Kriging: Its interpolations are based on values in a particular finite area. It is a more accurate and intensive computation that uses point estimates within a block to derive an average estimate for the block.
5. Co-Kriging is a modification of ordinary Kriging that relies on the fact that many phenomena are multivariate and that the primary variable of interest is undersampled. Co-Kriging estimation is done by minimizing the variance of the estimation error using the cross-correlation (dependence) between several variables. Estimates are derived for both the primary and secondary variables. The co-Kriging technique is a modification of the simpler technique of Kriging. It is used to merge two variables or more. Estimation of co-Kriging contains a primary variable of interest and one or more secondary variables. Improvement in the interpolation of one variable by using other variables is important (David 1977; Journel and Huijbregts 1978; Myers 1982). There are two steps in co-Kriging estimation:

   (a) Evaluation of the cross-correlation (or co-SVs) between variables to obtain information about continuity and dependencies
   (b) Construction of contour maps for the primary variable

   Seo (1998) used linear co-Kriging to interpret rainfall data from a set of rain gages and radar information. He concludes that the consistency of the improvement by gage-radar estimation makes co-Kriging an attractive tool in rainfall estimation. Martinez (1996) applied co-Kriging to improve the accuracy of evapotranspiration estimation over a regular grid by including the effects of topography.

Among other spatial estimation techniques, Kriging methodology has the following advantages:

1. It yields best linear unbiased estimator (BLUE) for the ReV modeling in earth sciences.

2. It is best because it minimizes the error variance in the estimate, unbiased because the weights sum to 1, and linear because it is a simple weighted average.
3. Also uses a weighted average method to calculate the value at unsampled locations (Eq. 2.27).
4. Weights are determined by considering the representative sample, and then the theoretical SV (or CSV) and the estimation problem are solved by using matrix algebra.

In the application of Kriging algorithm, there are four successive essential steps as follows:

1. When computing the interpolation weights, the algorithm considers the spacing between the point of estimation and the data sites. The algorithm considers also the inter-data spacing, which allows for de-clustering.
2. When computing the interpolation weights, the algorithm considers the inherent length scale of the data. For example, there are regions where the topography varies much more slowly than some other regions. If two points' elevations at the same distance are considered in these two different topographies, then in the slowly changing regional elevation case, it would be reasonable to assume a linear variation between these two observations, while in the other region, such an assumed linear variation would be unrealistic. The algorithm adjusts the interpolation weights accordingly.
3. When computing the interpolation weights, the algorithm considers the inherent trustworthiness of the data. If the data measurements are exceedingly precise and accurate, the interpolated surface goes through each and every observed value. If the data measurements are suspected, the interpolated surface may not go through an observed value, especially if a particular value is in stark disagreement with neighboring observed values. This is an issue of data repeatability.
4. Natural phenomena are created by physical processes, which have often preferred orientations. For example, at the mouth of a river, the coarse material settles out fastest, while the finer material takes longer to settle. Thus, the closer one is to the shoreline, the coarser the sediments, while the further from the shoreline, the finer the sediments. When computing the interpolation weights, the algorithm incorporates this natural anisotropy. When interpolating at a point, an observation 100 m away but in a direction parallel to the shoreline is more likely to be similar to the value at the interpolation point than is an equidistant observation in a direction perpendicular to the shoreline.

The last three items incorporate something about the underlying process from which the observations are taken. The length scale, data repeatability, and anisotropy are not a function of the data locations. These enter into the Kriging algorithm via the SV (or CSV). The length scale is given by the SV range (or slope), the data repeatability is specified by the nugget effect, and the anisotropy is given by the anisotropy (Chap. 4).

6.6 Simple Kriging (SK)

This Kriging method has straightforward mathematical derivations based on the following three basic assumptions:

1. The spatial sampling points are representatives of the ReV at a set of given locations with measurement values.
2. The ReV is considered as a second-order random field variable with mean, variance, and SV.
3. The mean of ReV is known, which limits the application of this Kriging modeling alternative severely.

In practical applications, there are many cases where the areal mean of the ReV is known, and hence direct application of the simple Kriging methodology becomes attractive. Especially in the case of second-order stationarity (mean and variance constancy), the whole ReV samples can be standardized according to classical statistical standardization formulation, and hence the standardized ReV has zero mean and unit variance. Let a set of measurement locations scattered irregularly over the study area as in Fig. 6.14, where there are n measurement and one estimation sites.

The regional variability of the ReV is recognized by a suitable regional dependence (covariance or SV) function between each pair of the measuring sites. Kriging estimation is considered as the weighted average of the measurement values on the estimation point with distant-dependent weighting values, λ_i. The Kriging estimation, Z_E is a linear weighted average of the surrounding measurements Z_i ($i = 1, 2, \ldots, n$) as follows:

$$Z_E = \overline{Z} + \sum_{i=1}^{n} \lambda_i \left(Z_i - \overline{Z}\right) \tag{6.34}$$

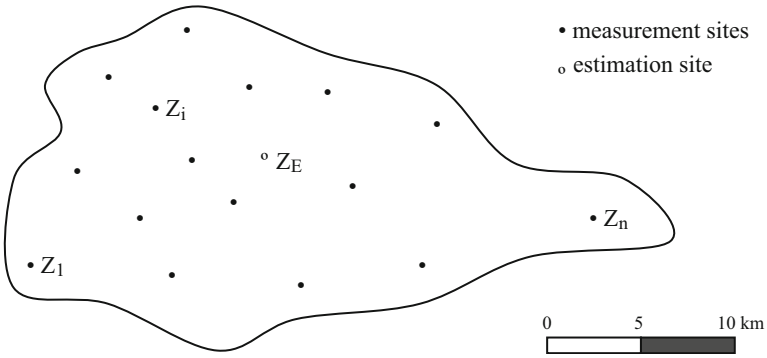


Fig. 6.14 ReV sample sites and estimation site

where $\overline{Z}$ indicates the regional constant mean value of the ReV. If there are n neighboring sites for the estimation calculation, then there will be n^2 elements in the covariance (or SV) matrix with the variances on the main diagonal. Due to the diagonal symmetry, the number of different elements in the matrix is equal to $n(n-1)/2$ as follows:

$$C = \begin{bmatrix} \text{var}(z_1) & \text{cov}(z_1, z_2) & \ldots & \text{cov}(z_1, z_n) \\ \text{cov}(z_2, z_1) & \text{var}(z_2) & \ldots & \text{cov}(z_2, z_n) \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ \text{cov}(z_n, z_1) & \text{cov}(z_n, z_2) & \ldots & \text{var}(z_n) \end{bmatrix} \tag{6.35}$$

In this matrix, each element is dependent also on the distance difference (relative distance) between the two sites. It is obvious that $\text{Cov}(z_i,\ z_j) = \text{Cov}(z_j,\ z_i)$ and furthermore $\text{Cov}(z_i,\ z_i) = \sigma_i^2$ which is the variance at site i. However, if the ReV is standardized with constant regional mean, $\overline{Z}$, and variance, σ_Z^2 then $\text{Cov}(z_i,\ z_i) = 1$, and the covariance corresponds to dependence (correlation) coefficient as $\rho(z_1,\ z_2) = \text{cov}(z_1,\ z_2)$. For standardized ReV, Eq. 6.35 takes the following form:

$$\rho = \begin{bmatrix} 1 & \rho(z_1, z_2) & \ldots & \rho(z_1, z_n) \\ \rho(z_2, z_1) & 1 & \ldots & \rho(z_2, z_n) \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots .1 & . \\ \rho(z_n, z_1) & \rho(z_n, z_2) & \ldots & 1 \end{bmatrix} \tag{6.36}$$

This is the regional correlation matrix for ReV. Similar to this matrix, one can write also the distance matrix, D, between these n sites with zero distances along the main diagonal as follows:

$$D = \begin{bmatrix} 0 & \text{dis}(z_1, z_2) & \ldots & \text{dis}(z_1, z_n) \\ \text{dis}(z_2, z_1) & 0 & \ldots & \text{dis}(z_2, z_n) \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & 0 & . \\ dis(z_n, z_1) & dis(z_n, z_2) & \ldots & 0 \end{bmatrix} \tag{6.37}$$

This matrix is also symmetrical with respect to the main diagonal. Hence, both the covariance and distance matrices provide information in the form of say upper triangular matrices. The plot of distance matrix values on the horizontal axis versus corresponding regional correlation coefficients from the correlation matrix provides a general shape as in Fig. 6.15, which may be referred to as the regional dependence (correlation) function. Logically, as the distance increases, the correlation

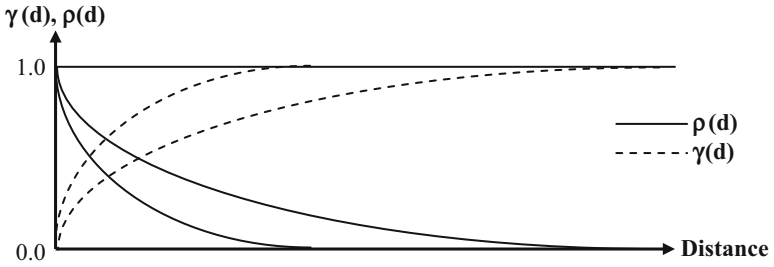


Fig. 6.15 Covariance-distance graph

coefficient between the ReV values decreases, and therefore, Fig. 6.15 has a decreasing trend with distance, and theoretically this function should be asymptotic to the horizontal axis. Consideration of Eq. 6.31 with unit variance yields the corresponding SVs in the same figure.

The general expression in Eq. 6.34 for the simple Kriging can be rewritten for a standardized ReV as

$$z_E = \sum_{i=1}^{n} \lambda_i z_i \tag{6.38}$$

Here there are n unknowns and accordingly n equations are necessary for the simultaneous solution. For this purpose, both sides of Eq. 6.38 are multiplied by each measurement ReV variable, and then the averages (expectations) are taken. The resultant set of equations becomes

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i \rho(z_i, z_1) &= \rho(z_E, z_1) \\ \sum_{i=1}^{n} \lambda_i \rho(z_i, z_2) &= \rho(z_E, z_2) \\ \sum_{i=1}^{n} \lambda_i \rho(z_i, z_k) &= \rho(z_E, z_k) \end{aligned} \tag{6.39}$$

In order to bring this set of simultaneous equations into a matrix form, the following additional succinct vector definitions are necessary. The unknown column vector is

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ \lambda_n \end{bmatrix} \tag{6.40}$$

Finally, the right-hand side of Eq. 6.39 represents the known part, say, column vector, B, which is defined as

$$B = \begin{bmatrix} \rho(z_E, z_1) \\ \rho(z_E, z_2) \\ . \\ . \\ \rho(z_E, z_n) \end{bmatrix} \tag{6.41}$$

With these notations at hand, Eq. 6.39 can be written shortly as

$$C\Lambda = B$$

or the inversion operation gives the solution implicitly as

$$\Lambda = C^{-1}B \tag{6.42}$$

After the determination of the weighting values, λ_i from this last expression, their substitution into Eq. 6.38 leads to the estimation of the standard ReV, which is then converted to nonstandard (original) ReV as

$$Z_E = \overline{Z} + \sigma_E^2 z_E \tag{6.43}$$

where ZE is an $(n \times 1)$ matrix of the measured ReV values with zero mean and unit variance.

The simple Kriging procedure can also be applied by considering the relationship between the SV and the covariance function as given in Eq. 6.31. However, the application of this transition between the covariance and the corresponding SV will be reliable only in the case of the normally distributed ReV. Otherwise, the results obtained from the use of the covariance will be biased. The reader can see the difference by applying the simple Kriging once with the covariance and then the SV functions of the same set ReV measurements. The simple Kriging is equivalent to multiple regression procedures where the covariance is used for the parameter estimation.

All that has been explained in this section is based on the covariance function for the depiction of spatial dependence. Since there is a relationship between the covariance and SV functions in the case of standardized ReV according to Eq. 6.31 as $\rho(d) = 1 - \gamma(d)$, then the replacement of the covariance terms in all the equations of this section provides an alternative spatial modeling of ReV based on SV. The variance of the estimation in case of covariance use is

$$\sigma_E^2 = 1 - B'\Lambda \tag{6.44}$$

When the SV is used for the spatial modeling, the same estimation variance becomes as

$$\sigma_E^2 = B'\Lambda \tag{6.45}$$

The critic of the simple Kriging is that it depends on the statistical property of the covariance (or SV) function preservation in the final estimations. In other words, the spatial estimation is achieved in such a way that the overall spatial dependence function (SDF) of the ReV is preserved throughout the procedure. Unfortunately, neither the cross-validation nor the unbiasedness procedures are applied explicitly in the simple Kriging procedure.

Example 6.1 The earthquake magnitude measurements at five stations (Z_1, Z_2, Z_3, Z_4 and Z_5) are presented in Table 6.5 with their positions in Fig. 6.16. The spatial estimation, *ZE*, is obtained from these measurements.

The distance matrix between each pair of data is calculated with the following results:

$$D = \begin{bmatrix} 0 & & & & \\ 0.534689 & 0 & & & \\ 0.349005 & 0.326195 & 0 & & \\ 0.664208 & 0.185368 & 0.370756 & 0 & \\ 0.483628 & 1.015764 & 0.772863 & 1.129264 & 0 \end{bmatrix}$$

On the other hand, the distances between the estimation site, ZE, and measurement sites are as follows:

	1	**2**	**3**	**4**	**5**
E	0.297151	0.284038	0.092392	0.374746	0.754616

Provided that the ReV is standardized, the corresponding half-square difference (i.e., SV) matrix can be obtained by substituting Eq. 6.31 with $\sigma_Z^2 = 1$ into Eq. 6.36, which leads to

$$\Gamma = \begin{bmatrix} 0 & 1-\gamma(z_1,z_2) & \cdots & 1-\gamma(z_1,z_n) \\ 1-\gamma(z_2,z_1) & 0 & \cdots & 1-\gamma(z_2,z_n) \\ \vdots & \vdots & \cdots & \vdots \\ 1-\gamma(z_n,z_1) & 1-\gamma(z_n,z_2) & \cdots & 0 \end{bmatrix} \tag{6.46}$$

Table 6.5 Spatial data

ReV	Easting (km)	Northing (km)	Magnitude
Z1	24.47950	40.40017	1.87
Z2	24.48400	39.86550	1.92
Z3	24.68783	40.12017	2.2
Z4	24.63183	39.75367	2.15
Z5	24.54383	40.87950	2.07
ZE	24.59567	40.12667	**2.65**

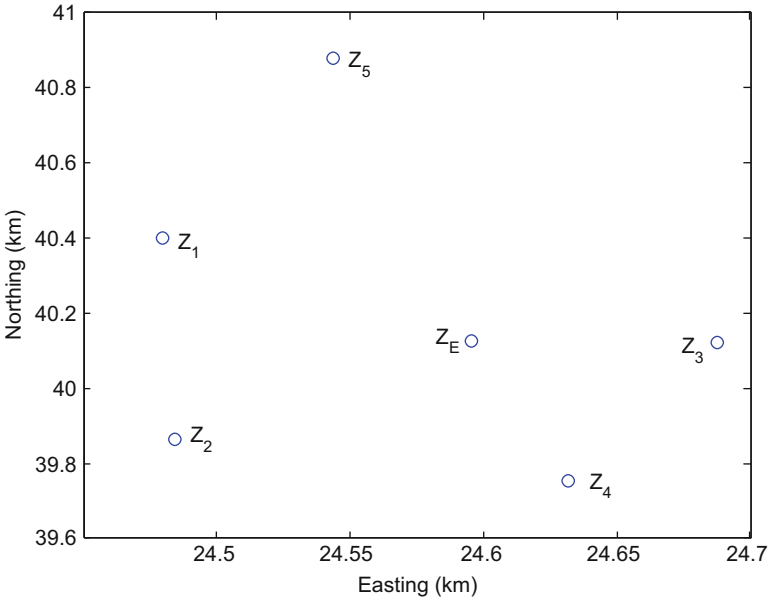


Fig. 6.16 Spatial scatter of data locations

There are two different ways to calculate the SV values in practical works either from a given small sample as in Table 6.6 without knowing the basic structure of the sample SV or after defining the structural form of the sample SV from a large number of data, which is preferred to be more than 30 data values. The former approach yields to a matrix that represents half-square differences and according to Eq. 6.46 the subtraction from 1 results in Table 6.5.

$$\Gamma = \begin{bmatrix} 0 & & & \\ 1\text{-}0.00125 & 0 & & \\ 1\text{-}0.05445 & 1-0.03920 & 0 & \\ 1\text{-}0.03920 & 1-0.02645 & 1-0.00125 & 0 \\ 1\text{-}0.02000 & 1-0.01125 & 1-0.00845 & 1-0.00320 \end{bmatrix}$$

Likewise, Eq. 4.41 can be written by considering Eq. 6.31 as

$$B = \begin{bmatrix} 1-\gamma(z_E, z_1) \\ 1-\gamma(z_E, z_2) \\ . \\ . \\ 1-\gamma(z_E, z_n) \end{bmatrix} \tag{6.47}$$

It is not possible to estimate the SV values in this vector because the earthquake value at prediction location is not known. Therefore, it is necessary to know the global SV that would depend on many location records, and it is assumed herein that from a priori structural analysis had produced the sample SV model as a linear model as

$$\gamma(d) = 0.015 + 0.1d$$

Now one can calculate the SV value from the distances between the estimation point and other surrounding points in Fig. 6.16, which leads to

$$B = \begin{bmatrix} 1 - 0.0447151 \\ 1 - 0.0434038 \\ 1 - 0.0242392 \\ 1 - 0.0524746 \\ 1 - 0.0904616 \end{bmatrix}$$

Of course it is now possible to calculate the matrix in Eq. 6.46 according to the distance matrix above by using large sample SV equation which leads to

$$\Gamma = \begin{bmatrix} 0 & & & \\ 1\text{-}0.0684689 & 0 & & \\ 1\text{-}0.0499005 & 1 - 0.0476195 & 0 & \\ 1\text{-}0.0814208 & 1 - 0.0335368 & 1 - 0.0520756 & 0 \\ 1\text{-}0.0633628 & 1 - 0.1165764 & 1 - 0.0922863 & 1 - 0.279264 \end{bmatrix}$$

The final solution can be found by taking inverse of this Γ, which appears as follows:

$$\Gamma^{-1} = \begin{bmatrix} -0.7747 & 0.3017 & 0.2867 & 0.2763 & 0.2058 \\ 0.3017 & -0.7773 & 0.2822 & 0.2041 & 0.2866 \\ 0.2867 & 0.2822 & -0.7600 & 0.2391 & 0.2558 \\ 0.2763 & 0.2041 & 0.2391 & -0.9102 & 0.4471 \\ 0.2058 & 0.2866 & 0.2558 & 0.4471 & -0.9829 \end{bmatrix}$$

Hence, application of Eq. 6.42 leads to the final weight values as

$$\Lambda = \begin{bmatrix} 0.2801 \\ 0.2668 \\ 0.2641 \\ 0.2387 \\ 0.2526 \end{bmatrix}$$

It is now possible to calculate the prediction value from Eq. 6.32, which gives $ZE = 2.65$.

6.7 Ordinary Kriging (OK)

This procedure has the two assumptions of the simple Kriging with conflict of the third one, which is assumed to be constant regionally but unknown. Hence, it is not possible to apply the standardization procedure to the ReV measurements because the mean value is unknown. Similar expression to Eq. 6.34 can be written but with the consideration of unknown regional mean, m, value as

$$Z_E = m + \sum_{i=1}^{n} \lambda_i (Z_i - m) \tag{6.48}$$

Comparison of this with Eq. 6.34 indicates that rather than the standardized ReV variables, nonstandardized ReV are used. If both sides of this last expression are arithmetically averaged, one can then obtain

$$m = m + m \sum_{i=1}^{n} \lambda_i - m = 0$$

which yields the restrictive condition, as has already been given by Eq. 6.33. Hence, the first rule in the ordinary Kriging is that the summation of all the weights ought to be equal to 1. It indicates that the Kriging weights are independent of the ReV average. This is referred to as the unbiasedness principle in the Kriging literature. It is possible to write Eq. 6.48 as

$$Z_E = m \left(1 - \sum_{i=1}^{n} \lambda_i \right) + \sum_{i=1}^{n} \lambda_i Z_i \tag{6.49}$$

where the parenthesis in the first term on the right-hand side is equal to zero by definition. Equation 6.33 is a condition to the main estimation expression in Eq. 6.48, which can be rewritten with the estimation error term, εE as

$$(Z_E - m) = \sum_{i=1}^{n} (\lambda_i Z_i - m) + \varepsilon_E$$

In the following, this estimation error variance will be minimized. Let us leave the estimation error as a subject

$$\varepsilon_E = (Z_E - m) - \sum_{i=1}^{n} (\lambda_i Z_i - m)$$

It is known by definition that the error estimation overall arithmetic average, $\overline{\varepsilon}_E$, is equal to zero. The square of both sides can be obtained as

$$\varepsilon_E^2 = (Z_E - m)^2 - 2\sum_{i=1}^{n}(\lambda_i Z_i - m)(Z_E - m) + \left[\sum_{i=1}^{n}(\lambda_i Z_i - m)\right]^2$$

or more explicitly

$$\varepsilon_E^2 = (Z_E - m)^2 - 2\sum_{i=1}^{n}(\lambda_i Z_i - m)(Z_E - m) + \sum_{j=1}^{n}\sum_{i=1}^{n}\left(\lambda_j Z_j - m\right)(\lambda_i Z_i - m)$$

If the estimation is made n times, then the average error estimate square (variance) will be obtained as

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{n}\left(\varepsilon_E^2\right)_l = \; & \frac{1}{n}\sum_{l=1}^{n}(Z_E - m)^2 - 2\frac{1}{n}\sum_{l=1}^{n}\sum_{i=1}^{n}(\lambda_i Z_i - m)(Z_E - m) \\ & + \frac{1}{n}\sum_{l=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\left(\lambda_j Z_j - m\right)(\lambda_i Z_i - m)\end{aligned}$$

or

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{n}\left(\varepsilon_E^2\right)_l = \; & \frac{1}{n}\left[\sum_{l=1}^{n}(Z_E - m)^2\right] - 2\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}(\lambda_i Z_i - m)_l (X_E - m)_l\right] \\ & + \sum_{j=1}^{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}\left(\lambda_j Z_j - m\right)\left((\lambda_i Z_i - m)\right)\right]\end{aligned}$$

The first average in the big brackets on the right-hand side is equal to the estimation variance, σ_E^2; the second term average is equivalent to estimation-measurement covariance, Cov (ZE, Zi); and the last term is the covariance between two measurements, Cov (Zi, Zj). In fact, the average terms based on n values are equivalent to their respective expectations as n goes to infinity theoretically. The last expression can be written with these new covariance values as

$$\overline{\varepsilon_{El}^2} = \sigma_E^2 - 2\sum_{i=1}^{n}\mathrm{Cov}(Z_E, Z_i) + \sum_{j=1}^{n}\sum_{i=1}^{n}\mathrm{Cov}\left(Z_i, Z_j\right) \tag{6.50}$$

This expression must be minimized with the condition in Eq. 6.33, and therefore, the minimization equation can be written with the Lagrange multiplier, μ, as

$$\overline{\varepsilon_{El}^2} = \sigma_E^2 - 2\sum_{i=1}^{n}\mathrm{Cov}(Z_E, Z_i) + \sum_{j=1}^{n}\sum_{i=1}^{n}\mathrm{Cov}\left(Z_i, Z_j\right) + \mu\sum_{i=1}^{n}\lambda_i \tag{6.51}$$

This can be written in the matrix and vector form after the definition of the following quantities:

$$C = \begin{bmatrix} \mathrm{cov}(z_1, z_1) & \mathrm{cov}(z_1, z_2) & \ldots & \mathrm{cov}(z_1, z_n) & 1 \\ \mathrm{cov}(z_2, z_1) & \mathrm{cov}(z_2, z_2) & \ldots & \mathrm{cov}(z_2, z_n) & 1 \\ . & . & \ldots & . & . \\ . & . & \ldots & . & . \\ . & . & \ldots & . & . \\ \mathrm{cov}(z_n, z_1) & \mathrm{cov}(z_n, z_2) & \ldots & \mathrm{cov}(z_n, z_n) & 1 \\ 1 & 1 & 1 & & 0 \end{bmatrix} \tag{6.52}$$

which can be written based on the SV formulation alternatively as

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ \lambda_n \\ \mu \end{bmatrix} \tag{6.53}$$

and finally

$$B = \begin{bmatrix} \mathrm{cov}(z_E, z_1) \\ \mathrm{cov}(z_E, z_2) \\ . \\ . \\ \mathrm{cov}(z_E, z_n) \\ 1 \end{bmatrix} \tag{6.54}$$

The measurement vector that includes n nearest values is given as

$$M = \begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ Z_n \\ 0 \end{bmatrix} \tag{6.55}$$

With these notations, the first ordinary Kriging weights may be estimated by using either the covariance or SV values from

$$\Lambda = C^{-1}B \tag{6.56}$$

The ordinary Kriging estimate of the regionalized variable at location E can be obtained as

$$Z_E = M^T \Lambda = M^T C^{-1} B \tag{6.57}$$

Finally, the ordinary Kriging estimation variance is

$$\sigma_E^2 = B^T \Lambda = B^T C^{-1} B \tag{6.58}$$

The estimate and estimation error depend on the weights chosen. Ideally, Kriging tries to choose the optimal weights that produce the minimum estimation error. In order to derive the necessary equations for Kriging, extensive calculus use is required, which is not included here; however, information about the derivation can be found in various textbooks such as by Clark (1979) and Olea (1975). Optimal weights produce unbiased estimates and have a minimum estimation variance, which are obtained by solving a set of simultaneous equations. In the case of SV, the corresponding expressions to Eqs. 6.52, 6.53, and 6.54 are respectively.

$$\begin{bmatrix} \gamma(z_1,z_1) & \gamma(z_1,z_2) & \cdots & \gamma(z_1,z_n) & 1 \\ \gamma(z_2,z_1) & \gamma(z_2,z_2) & \cdots & \gamma(z_2,z_n) & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \gamma(z_n,z_1) & \gamma(z_n,z_2) & \cdots & \gamma(z_n,z_n) & 1 \\ 1 & 1 & 1 & & 0 \end{bmatrix} \tag{6.59}$$

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\mu \end{bmatrix} \tag{6.60}$$

$$B = \begin{bmatrix} \gamma(z_E,z_1) \\ \gamma(z_E,z_2) \\ \vdots \\ \gamma(z_E,z_n) \\ 1 \end{bmatrix} \tag{6.61}$$

Example 6.2 Ordinary Kriging methodology is demonstrated by considering groundwater levels at five sites as given in Table 6.6 with easting, northing, and groundwater level elevation values. The graphical locations are shown in Fig. 6.17.

Table 6.6 Well coordinates and groundwater levels

Well location	Easting (km)	Northing (km)	Elevation (m)
Z1	4.2	3.0	89
Z2	3.7	6.5	75
Z3	1.8	2.6	98
Z4	2.9	4.5	81
Z5	4.3	1.5	105
ZE	3.2	3.0	**91.27**

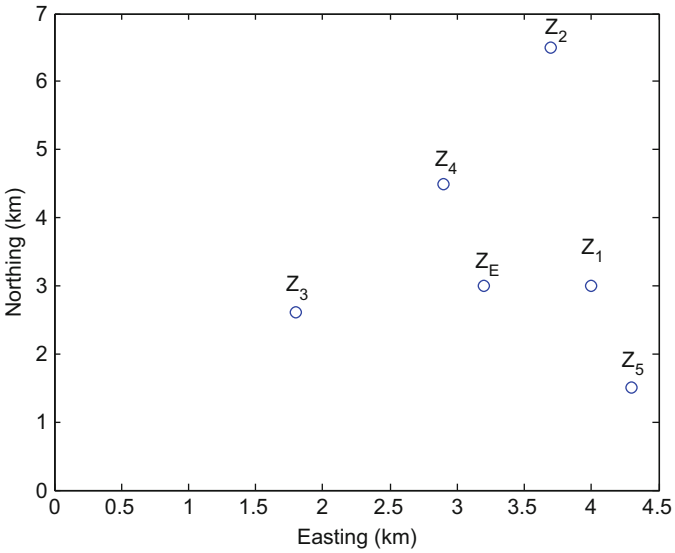


Fig. 6.17 Sample geographical locations

The distances between each pair are given as in the following matrix:

$$D = \begin{bmatrix} 0 & & & & \\ 3.512834 & 0 & & & \\ 2.236068 & 4.338202 & 0 & & \\ 1.860108 & 2.154066 & 2.19545 & 0 & \\ 1.529706 & 5.035871 & 2.731300 & 3.310589 & 0 \end{bmatrix}$$

The distances between the estimation point, *ZE*, and others are given as

	1	**2**	**3**	**4**	**5**
E	0.8	3.535534	1.456022	1.529706	1.860107524

After calculations, the half-square difference matrix appears as follows:

$$\gamma(d) = \begin{bmatrix} 0 & & & & \\ 332112.5 & 0 & & & \\ 313632.0 & 264.5 & 0 & & \\ 327240.5 & 18.0 & 144.5 & 0 & \\ 308112.5 & 450.0 & 24.5 & 288.0 & 0 \end{bmatrix}$$

For large sample sizes, the structure of the SV is determined as a linear function as

$$\gamma(d) = 0.3 + 3.8d$$

and accordingly the SV matrix turns out to be as

$$\Gamma = \begin{bmatrix} 0 & 13.68 & 8.78 & 7.37 & 6.11 & 1 \\ 13.68 & 0 & 16.78 & 8.47 & 19.93 & 1 \\ 8.78 & 16.78 & 08.86 & 10.68 & 1 & \\ 7.37 & 8.47 & 8.86 & 0 & 12.88 & 1 \\ 6.11 & 19.93 & 10.68 & 12.88 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

The inverse of this matrix becomes

$$\Gamma^{-1} = \begin{bmatrix} -0.1489 & 0.0122 & 0.0159 & 0.0465 & 0.0742 & -0.1032 \\ 0.0122 & -0.0603 & -0.0016 & 0.0524 & -0.0027 & 0.4696 \\ 0.0159 & -0.0016 & -0.0788 & 0.0393 & 0.0252 & 0.2699 \\ 0.0465 & 0.0524 & 0.0393 & -0.1317 & -0.0065 & -0.0512 \\ 0.0742 & -0.0027 & 0.0252 & -0.0065 & -0.0902 & 0.4149 \\ -0.1032 & 0.4696 & 0.2699 & -0.0512 & 0.4149 & -10.9519 \end{bmatrix}$$

Likewise from Eq. 6.61 realizing the distances between the estimation site and other sites as given above, it is possible to obtain

$$B = \begin{bmatrix} 3.34 \\ 13.74 \\ 5.83 \\ 6.11 \\ 7.37 \\ 1 \end{bmatrix}$$

It is possible to find from Eq. 6.56 the lambda values as

$$\lambda = \begin{bmatrix} 0.4916 \\ -0.0275 \\ 0.2676 \\ 0.2004 \\ 0.0679 \\ -0.5255 \end{bmatrix}$$

One can evaluate the estimation value from Eq. 6.57 which leads to $ZE = 91.27$ m. The Kriging estimation variance is the weighted sum of the SV for the distance from the points to the estimation location which can be calculated from Eq. 6.58 leading to $\sigma_E^2 = 4.02\text{m}^2$. This implies that the standard error of estimation is $\sigma_E = 2.00$ m.

It is possible to make such calculations with estimation and its estimation variance at every point. For this purpose, it is useful to make cross-validation by considering each measurement as if it does not exist, and Kriging methodology gives the estimation and error variance. Hence, one can construct two maps for the Kriging estimates as a best guess of the mapped variable configuration and an error map showing the confidence envelope that surrounds this estimation. All these are based on the measurement and estimation sites' configuration and distances between measurements used in the estimation process and on the degree of spatial continuity of the ReV as expressed by the spatial covariance and preferably SV models.

6.8 Universal Kriging (UK)

Unfortunately, the ordinary Kriging cannot be used unless the ReV has a constant mean although it is not known. The successive differences canceal the regional mean in the SV calculations, which means that in the case of symmetric variation, the regional abrupt change is shown in Fig. 6.9 and trends in Fig. 6.11. It is, therefore, necessary to develop another ReV estimation procedure that should take into consideration these systematic mean variations. The universal Kriging considers first-order stationary ReV as consisting of two components, namely, the regional drift and the residuals. The drift is the long-term average (expectation) value of the ReV within the estimation site neighborhood. It slowly varies and it is the nonstationary component of the local region. On the other hand, the residuals are the differences between the measurements and the corresponding drift values. If the drift is removed from the ReV, then the residuals can be modeled by the

ordinary Kriging principles. Hence, in the procedural structure of the universal Kriging, there are three steps:

1. Removal of the drift values from the measurements of the ReV
2. Application of the ordinary Kriging to establish the residuals at non-measurement points
3. Addition of these residual values to the original drift values

The difference between the trend surface as defined in Chap. 3 Sect. 3.10 and the drift covers the partial local area, whereas the trend surface extends over all the ReV variability domain. The underlying drift component can be removed from the original measurements of the concerned ReV through different methodologies. Among these are linear and nonlinear trend surface fitting, 2D Fourier analysis, etc. These methodologies depict the drift component, and its subtraction from the original measurements leaves the residuals with almost zero arithmetic average. Under the statistical theory which includes universal Kriging, a single-valued, continuous, mappable property is called a ReV and is considered to consist of two parts, a drift, or expected value, and a residual, or deviation from the drift. The drift may be modeled by a local polynomial function within a neighborhood that is analogous to a local trend surface. If the drift is removed, the residual surface can be regarded as first-order stationary in a statistical sense. Hence, again the simple or ordinary Kriging models can be used for the residual data set. After the estimation of residuals again the summation with the convenient drift values, the original ReV can be estimated.

Apart from this rather complicated and piecewise application, it is also possible to develop universal Kriging equations including another condition of drift in the derivations with another Lagrange multiplier. Hence, the original measurements can be used directly in the calculations. For example, similar set of equations can be derived for the universal Kriging to Eqs. 6.57, 6.58 and 6.59 and Eq. 6.53 as

$$C = \begin{bmatrix} \gamma(z_1,z_1) & \lambda(z_1,z_2) & \cdots & \gamma(z_1,z_k) & 1 & L_{1,1} & L_{2,1} \\ \gamma(xz_2,z_1) & \gamma(z_2,z_2) & \cdots & \gamma(z_2,z_k) & 1 & L_{1,2} & L_{2,2} \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot \\ \gamma(z_k,z_1) & \gamma(z_k,z_2) & \cdots & \lambda(z_k,z_k) & 1 & L_{1,1} & L_{2,1} \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\ L_{1,1} & L_{1,2} & \cdots & L_{1,k} & 0 & 0 & 0 \\ L_{2,1} & L_{2,2} & \cdots & L_{21,k} & 0 & 0 & 0 \end{bmatrix} \tag{6.62}$$

$$\Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ \lambda_k \\ \mu_0 \\ \mu_1 \\ \mu_3 \end{bmatrix}$$

and finally

$$B = \begin{bmatrix} \gamma(z_E, z_1) \\ \gamma(z_E, z_2) \\ . \\ . \\ \gamma(z_E, z_k) \\ 1 \\ L_{1,k} \\ L_{2,k} \end{bmatrix} \tag{6.63}$$

respectively. Herein zi represents a vector of the coordinates of point I, while L1,i ($L2$,i) is the scalar value representing the location of this point along coordinate axis horizontal (vertical) or 1 (2), for instance, east-west (north-south) direction. The vector of universal Kriging weight Λ is found by Eq. 6.56 except that W and B are given by Eqs. 6.62 and 5.64. The measurement vector that includes k nearest values is given as

$$M = \begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ Z_k \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.64}$$

In this vector, there are $(k + d + 1)$ elements, where the last $d+1$ elements that correspond to the Lagrange multipliers are zero.

Universal Kriging is a procedure that can be used to estimate values of a surface at the nodes of a regular grid from irregularly spaced data points. If the surface is second-order stationary or can be made stationary by some transformation, the spatial autocorrelation will express the degree of dependence between all locations on the surface and most particularly between observations and grid nodes.

Example 6.3 The problem given in Example 6.2 can be extended for the application of Universal Kriging. By making use of Table 6.6, the numerical form of the coefficients matrix can be obtained similar to the previous example but with additional location values as

$$\Gamma = \begin{bmatrix} 0 & 13.68 & 8.78 & 7.37 & 6.11 & 1 & 4.2 & 3.0 \\ 13.68 & 0 & 16.78 & 8.47 & 19.93 & 1 & 3.7 & 6.5 \\ 8.78 & 16.78 & 0 & 8.86 & 10.68 & 1 & 1.8 & 2.6 \\ 7.37 & 8.47 & 8.86 & 0 & 12.88 & 1 & 2.9 & 4.5 \\ 6.11 & 19.93 & 10.68 & 12.88 & 0 & 1 & 4.3 & 1.5 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 4.2 & 3.7 & 1.8 & 2.9 & 4.3 & 0 & 0 & 0 \\ 3.0 & 6.5 & 2.6 & 4.5 & 1.5 & 0 & 0 & 0 \end{bmatrix}$$

The inverse of this 8×8 matrix can be obtained as

$$\Gamma^{-1} = \begin{bmatrix} -0.1248 & 0.0212 & -0.0192 & 0.0342 & 0.0887 & -0.9626 & 0.2474 & 0.0031 \\ 0.0212 & -0.0328 & -0.0222 & 0.0552 & -0.0184 & -0.5517 & 0.0815 & 0.1934 \\ -0.0192 & -0.0222 & -0.0252 & 0.0559 & 0.0108 & 1.7447 & -0.3575 & -0.0652 \\ 0.0342 & 0.0522 & 0.0559 & -0.1246 & -0.0177 & 0.2611 & -0.1285 & 0.0333 \\ 0.0887 & -0.0184 & 0.0108 & -0.0177 & -0.0634 & 0.5085 & 0.1571 & -0.1646 \\ -0.9626 & -0.5517 & 1.7447 & 0.2611 & 0.5085 & 40.1311 & -8.5379 & -5.6915 \\ 0.2474 & 0.0815 & -0.3575 & -0.1285 & 0.1571 & -8.5379 & 2.5455 & -0.0474 \\ 0.0031 & 0.1934 & -0.0652 & 0.0333 & -0.1646 & -6.6915 & -0.0474 & 1.5258 \end{bmatrix}$$

On the other hand, the elements of the right-hand side vector in Eq. 6.39 become as follows:

$$B = \begin{bmatrix} 3.34 \\ 13.74 \\ 5.83 \\ 6.11 \\ 7.37 \\ 1 \\ 3.2 \\ 3.0 \end{bmatrix}$$

The universal Kriging weight vector can be found after the necessary algebraic calculations according to Eq. 6.56 as

$$\Lambda = \begin{bmatrix} 0.4625 \\ -0.0362 \\ 0.3094 \\ 0.2157 \\ 0.0486 \\ 0.4546 \\ -0.2998 \\ 0.0123 \end{bmatrix}$$

Finally, after all these calculations, the estimation value can be obtained from Eq. 6.57 as $ZE = 91.34$ m.

One can notice that for the given example, there is no major difference between the ordinary and universal Kriging methodologies. Ordinary Kriging, however, in common with other weighted-averaging methods, does not extrapolate well beyond the convex hull of the control points. That is, most estimated values will lie on the slopes of the surface, and the highest and lowest points on the surface usually will be defined by estimation (control) points.

6.9 Block Kriging (BK)

All the abovementioned Kriging methods (simple, ordinary, and universal) can be considered as punctual Kriging methodologies because their estimations are obtained at the support of individual sites. In the block Kriging, however, line, areal, or volume supports are considered, and a single estimation value is obtained that is representative for the whole support. This Kriging procedure also starts to work on a set of ReV measurement values, but its final product is valid for a certain domain in 1, 2, or 3 dimensions according to the type of problem. For instance, in ore grade reserve, pollution concentration, etc., estimations 3D shapes are considered. Rainfall maps and the hydrographs are of 2D.

The early versions of Kriging techniques were all concerned with the block Kriging since the ore reserves and grade degrees were the major concentrations for research in the domain of geology. In fact, the most commonly used block Kriging is similar to the ordinary Kriging with the interpretation of known vector, B, elements which are taken as the average of the spatial covariance or SV values within the interest of block domain, A. Actually the elements of vector Bn represent the covariances or SVs between the observations and A, integrated over the domain of 2D or 3D.

The major problem prior to the block Kriging application is the determination of block sizes. Each smaller subarea or sub-volume is represented by the geographical coordinates of its center point, and spatial covariance or SV is determined between

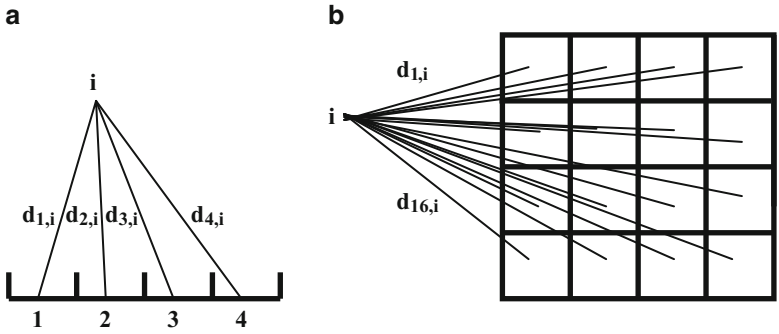


Fig. 6.18 Block pieces (**a**) line, (**b**) area

an observation, *Zi*, and each of the center points within a block. Then all of these covariance or SV is averaged in order to determine the point-to-block spatial covariance, cov(*A, Zi*), or SV, $\gamma(A, Xi)$ between block A and observation. If possible, the subdivisions of blocks should be regular and should be the same. It has been suggested by Isaaks and Srivastava (1989) that $4 \times 4 = 16$ and $4 \times 4 \times 4 = 64$ subdivisions that are adequate for the area and volume blocks. The fundamentals of block Kriging are presented in a simple manner by Olea (1999). It is commonly the counterpart of ordinary Kriging, and a set of normal equations must be solved similar to Eq. 6.59. In this new formulation, the spatial dependence function (SDF) either in the form of covariance or SV (or CSV) does not represent the relationship between two points, but they are the averages of the spatial covariance or SV (or CSV) between the point observations, *Zi*, and all possible points within an area, *A*.

Figure 6.18 shows the line and areal blocks and measurement point *i*. In Fig. 6.18a, the line is drawn into four pieces and the distances between the center of each piece and the measurement point I as d1,i, d2,i, d3,i, and d4,i. Likewise, in Fig. 6.18b, there are 16 subareas in a given block with distances d1,i, d2,i, . . ., d16,i.

Provided that the distances are known, their corresponding SV values can be obtained from a given theoretical SV function. The arithmetic averages of the relevant SV values give the representative SV value for line or areal block. Furthermore, now the block is considered as a point with this average SV value, and then accordingly one of the convenient Kriging techniques as mentioned above can be applied suitably.

6.10 Triple Diagram Model (TDM)

Since gradual (trend) or abrupt (shifts) climatic change questions have gained particular attention in recent years, most of the researches on lake level changes are concerned with meteorological factors of temperature and precipitation data.

Along this line of research, Hubert et al. (1989), Vannitsem and Demaree (1991), and Sneyers (1992) used statistical methods to show that temperature, pressure, and flow series in Africa and Europe have altered several times during the present century. On the other hand, as stated by Slivitzky and Mathier (1993), most of the modeling of levels and flow series on the Great Lakes has assumed stationarity of time series using either Markov or ARIMA (AutoRegressive Moving Average) processes presented by Box and Jenkins (1976). These models may work on lags of one, two, three, or more, but they consider the linear structure in the lake level fluctuations. Since lake level fluctuations do not have stationarity property, classical models such as Markov and ARIMA processes cannot stimulate lake levels reliably. Multivariate models using monthly lake level variable failed to adequately reproduce the statistical properties and persistence of basin supplies (Loucks 1989; Iruine and Eberthardt 1992). On the other hand, spectral analysis of water levels pointed to the possibility of significant trends in lake level hydrological variables (Privalsky 1990). Almost all these scientific studies relied significantly on the presence of an autocorrelation coefficient as an indicator of long-term persistence in lake level time series. However, many researchers have shown that shifts in average lake level might introduce unrealistic and spurious autocorrelations. This is the main reason why the classical stochastic and statistical models often fail to reproduce the statistical properties. However, Mathier et al. (1992) were able to reproduce adequately the statistical properties of a shifting-mean model. In the following sequel, a version of the Kriging methodology is adopted and used for the lake level estimations. For this purpose, the world's largest soda lake, Lake Van on the Anatolian High Plateau in eastern Turkey (38.5 N and 43 E), is adopted for application (Fig. 6.19). Lake Van area has very severe winters with frequent temperatures below 0 °C. Most of the precipitation falls during winter season in

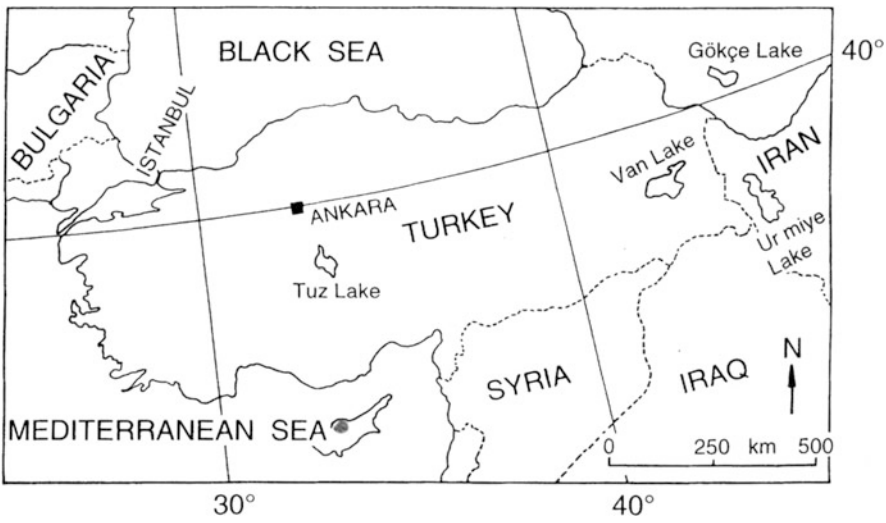


Fig. 6.19 Location map

the form of snow, and toward the end of spring, heavy rainfalls occur. High runoff rates occur in spring during snowmelt, and more than 80 % of annual discharge reaches the lake during this period. The summer period (July to September) is warm and dry with average temperatures of 20 °C. Diurnal temperature variations are about 20 °C.

Human beings can visualize at the maximum 3D variations. The best configuration and interpretation of such variations can be achieved in 3D Cartesian coordinate systems through contour maps. Generally, maps are regarded as the variation of a variable by location variables that are either longitudes and latitudes or eastings and northings (Isaaks and Srivastava 1989; Cressie 1993; Kitanidis 1997). Hence, it is possible to estimate the concerned (mapped) variable value for a given pair of location variables. Since one wants to predict the current lake level from previous records, it is suggested that two previous records replace the two location variables. In this manner, it is possible to map the current values of a variable based on two previous values of the same variable. The first step in any Kriging methodology prior to mapping is to determine the sample SV which guides for the theoretical model that will be employed in the classical Kriging modeling. For this purpose, the scatter of SV values versus distance is obtained for lag-one, lag-two, and lag-three. In order to depict the general trend of the scatter diagram, the distance range is divided into nine intervals, and the average of the SV values that fall within each interval is considered as the representative SV value within the midpoint distance of that interval as suggested by Myers et al. (1982). Different theoretical SV models such as linear, power, spherical, and Gaussian types are tried for the best fit, and at the end, the Gaussian SV is seen to have the best match with the sample SV trend (see Fig. 6.20). The Gaussian model is the most suitable in all lags, and the properties of fitted Gaussian SV model are presented in Table 6.7.

Such a mapping technique is referred to as triple diagram method (TDM) (Şen 2008). Such maps are based on three consecutive lake levels. TDMs help to make interpretations in spite of extremely scattered points. Although Davis (1986) has suggested for mapping the application of various simple regional techniques such as inverse distance, inverse distance square, etc. which consider the geometrical configuration of the scatter points only without the use of a third variable, herein, preparation of the TDM is based on classical Kriging technique.

The construction of a TDM requires three variables, two of which are referred to as independent variables (predictors), and they constitute the basic scatter diagram. The third is the dependent variable, which has its measured values attached to each scatter point. The equal-value lines are constructed by the Kriging methodology concepts explained in earlier sections of this chapter.

Lake Van water level records are used for the implementation of the Kriging methodology so as to obtain triple diagrams that give the common behavior of three variables, which are taken consequently from the historical time series data. The first two variables represent the two past lake levels and the third one indicates the present lake levels. Hence, the model has three parts, namely, observations (recorded time series) as input, triple diagram as response, and the output as prediction. It is possible to consider lags between the successive data at one, two,

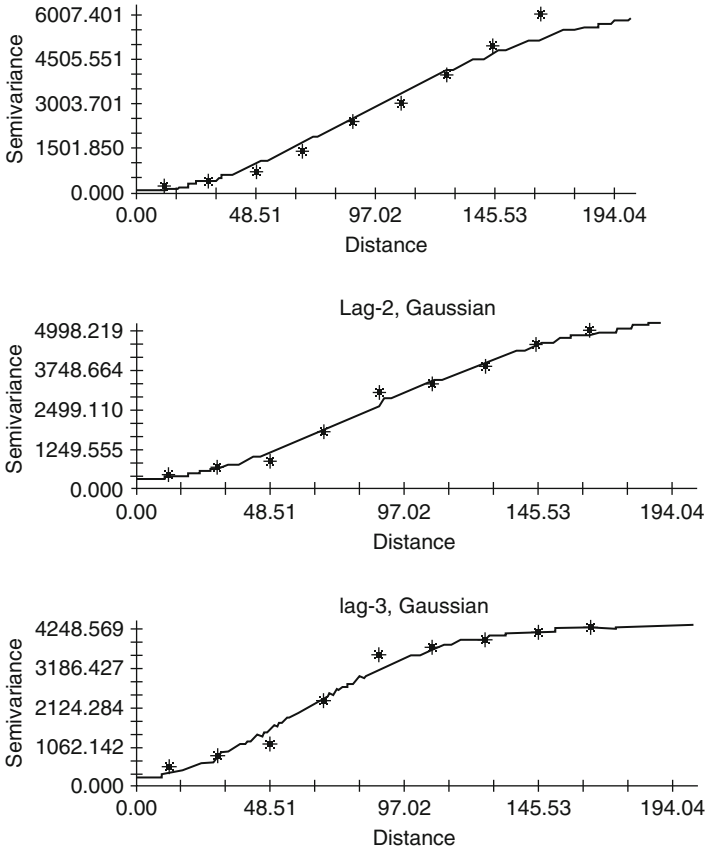


Fig. 6.20 Empirical and theoretical SV for three lags

Table 6.7 Theoretical Gaussian SV parameters

Lag	Nugget (cm^2)	Sill (cm^2)	Range (cm)	Correlation coefficient
1	70.0	6250	213.20	0.977
2	270.0	5555.0	586.80	0.990
3	250.0	4327.0	136.0	0.977

three, etc. intervals. Such an approach is very similar to a second-order Markov process, which can be expressed as

$$H_i = \alpha H_{i-1} + \beta H_{i-2} + \varepsilon_i \tag{6.65}$$

where Hi, $Hi-1$, and $Hi-2$ are the three consecutive lake levels; α and β are model parameters; and finally, εi is the error term. The application of such a model requires prior to any prediction procedure the parameter estimations from the

available data. Furthermore, its application is possible under the light of a set of assumptions, which includes linearity, normality (Gaussian distribution of the residuals, i.e., εi's), variance constancy, ergodicity, and independence of residuals. The triple diagram replaces Eq. 6.65 without any restriction in the form of map. Such a map presents the appearance of natural relationship between three consecutive time values of the same variable.

In order to apply the triple diagram approach, it is necessary to divide the data into training and testing parts. Herein, the last 24 months (2 years) are left for the test (prediction), whereas all other data values are employed for training, which is the TDM as in Fig. 6.21.

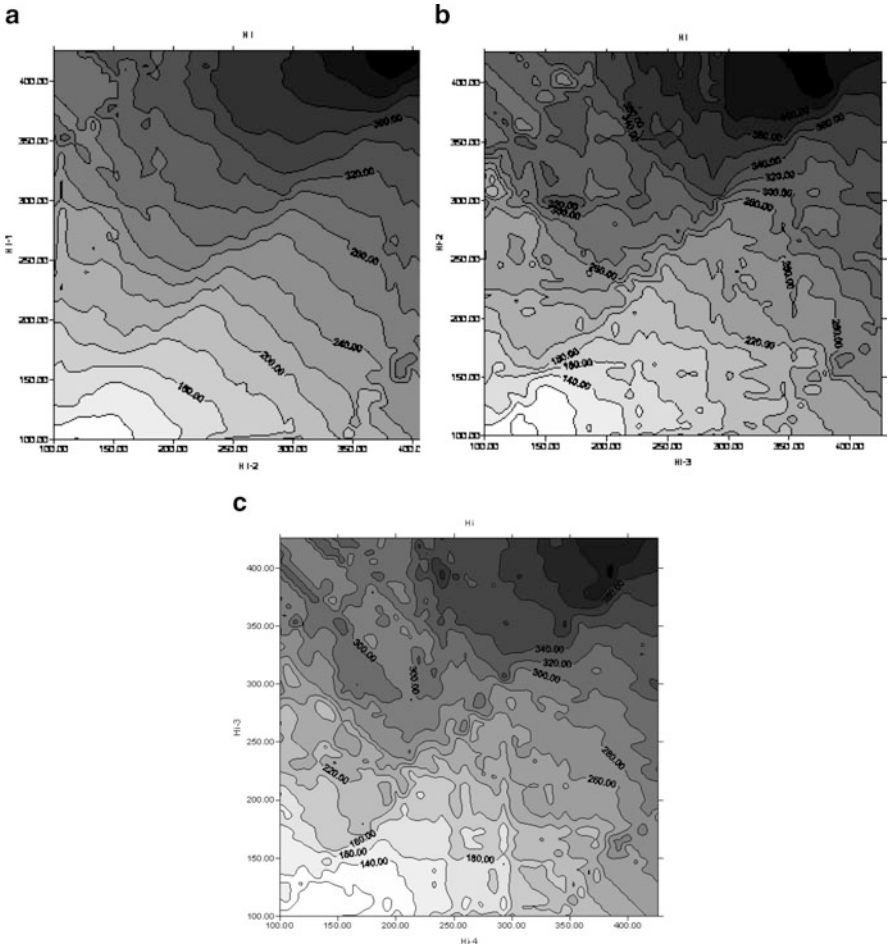


Fig. 6.21 Lake level TDMs (**a**) lag-one, (**b**) lag-two, (**c**) lag-three

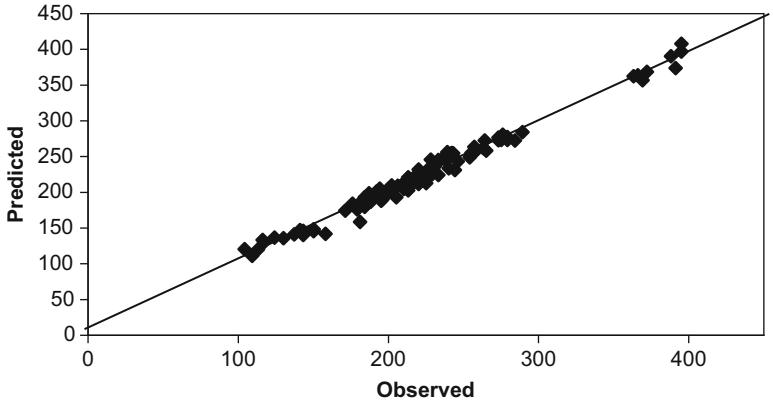


Fig. 6.22 Lag-one model verification

Prior to any prediction, it is possible to draw the following interpretations from these figures:

1. In the case of lag-one, there is a strong relationship between H_{i-1} and H_{i-2} with increasing contour values of H_i along almost 45° line (see Fig. 6.21a). The small H_i values are concentrated at small H_{i-1} and H_{i-2} values; this implies the clustering of small values of the three consecutive lake levels. Similarly, high lake level values of the three consecutive levels also constitute high values cluster. This means that small values follow small values and high values follow high values, which indicates positive correlations. Local variations in the contour lines appear at either low (high) H_{i-1} or high (low) H_{i-2} values. Consequently, better predictions can be expected within a certain band around the 45° line (Fig. 6.22). It is possible to deduce the following set of logical rules from Fig. 6.21b:

   IF H_{i-1} is low and H_{i-2} is low, THEN H_i is low.
   IF H_{i-1} is medium low and H_{i-2} is medium, THEN H_i is medium.
   IF H_{i-1} is high and H_{i-2} is high, THEN H_i is high.

   These rules can be used for fuzzy logic inference system as suggested by Zadeh (1968).
2. In Fig. 6.21b, the variations in the contour lines become very distinctive and rather haphazard compared to Fig. 6.21a. This implies that, with the increment in the lag value, present time lake level prediction will have more relative error. There is also a distinctive 45^0 line but with comparatively narrower band of certainty around it.
3. Finally, at lag-three case (Fig. 6.21c), the contour pattern takes even more haphazard variation. This implies increase in the relative error of predictions.

Şen et al. (1999, 2000) identified suitable models and estimates for lake level fluctuations and their parameters for trend, periodic, and stochastic parts. A

Table 6.8 Lag-one lake level prediction (cm)

H_{i-2}	H_{i-1}	H_i	Prediction	Relative error (%)
119	112	110	109.32	0.62
107	111	114	118.40	3.72
125	130	125	134.87	7.32
130	125	118	128.43	8.12
125	118	105	118.60	11.47
120	138	142	145.50	2.41
138	142	138	139.53	1.10
142	138	131	134.16	2.35
138	131	117	131.64	11.12
127	141	151	144.08	4.58
141	151	151	146.85	2.75
151	151	144	144.46	0.32
137	140	144	138.76	3.64
140	144	159	140.22	11.81
144	159	182	157.03	13.72
199	202	195	203.62	4.23
202	195	185	191.33	3.31
195	185	177	182.29	2.90
189	193	202	202.01	0.00
193	202	221	209.94	5.00
202	221	245	229.48	6.34
262	254	244	252.90	3.52
254	244	239	243.16	1.71
244	239	241	231.89	3.78
			Average	4.83

second-order Markov model is found suitable for the stochastic part. As explained before, TDM of lake levels can replace the second-order Markov process. In this manner, it is not necessary to use first- and second-order autocorrelation coefficients, in order to take into account more persistence. In order to make predictions for the last 24 months that are not used in the triple diagram constructions in Fig. 6.21, it is necessary to enter $Hi-1$ and $Hi-1$ for each month on vertical and horizontal axes, respectively. The prediction value of Hi can be either read from this map approximately or calculated by using Kriging prediction equations. The prediction results are shown in Table 6.8 with corresponding relative error amounts. Individual errors are slightly greater than 10 %, but the overall prediction relative error percentage is about 4.83 %.

Figure 6.3 indicates the observed and predicted Hi values. It is obvious that they follow each other very closely and on the average observed and predicted lake level series have almost the same statistical parameters.

The triple diagram model depicts even the increasing trend, which is not possible directly with the second-order Markov process. During the prediction procedure, there is no special treatment of trend, but even so, it is modeled successfully.

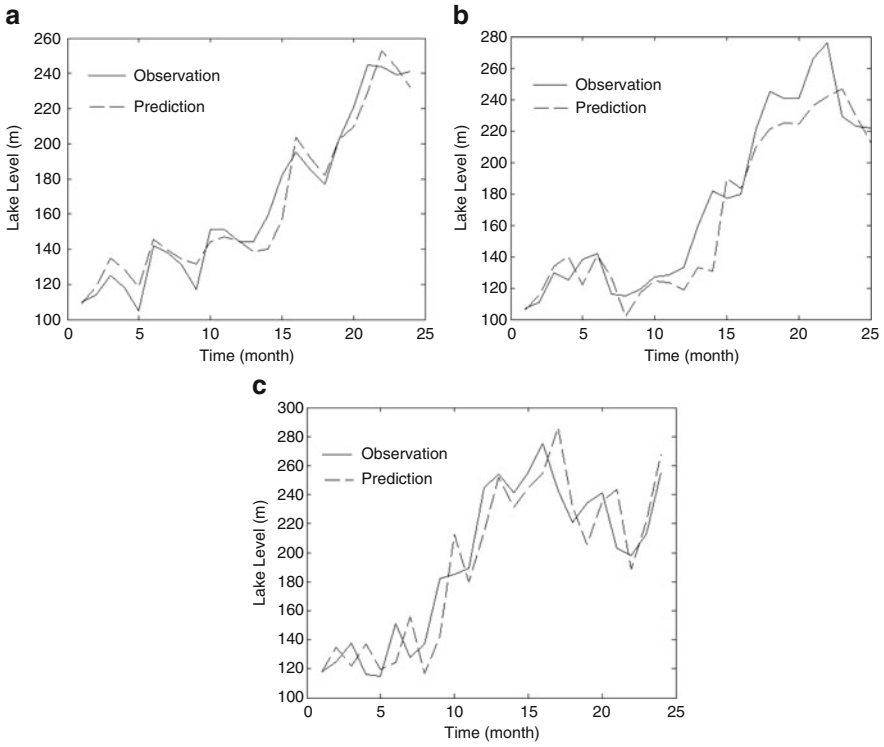


Fig. 6.23 Observed and predicted lake levels (**a**) lag-one, (**b**) lag-two, (**c**) lag-three

However, in any stochastic or statistical modeling, it is first necessary to make trend analysis and separate it from the original data. In order to further show the verification of the triple diagram approach for lake level predictions in Figs. 6.22 and 6.23 the test data are plotted along with the predictions. It is obvious that observation and prediction traces follow each other closely, which indicates the reliability of the triple diagram method.

6.11 Regional Rainfall Pattern Description

The mean annual and seasonal rainfall records in the southwest of Saudi Arabia are adapted from reports published by the Hydrology Division, Ministry of Agriculture and Water in Saudi Arabia, and Al-Jerash (1985). Rainfall records at 63 stations from 1971 to 1990 are selected for the Kriging application (Fig. 6.24). These stations are chosen based on four criteria (Subyani 2004, 2005):

1. They represent the best spatial coverage of the region.
2. They maximize the same monthly rainfall records.

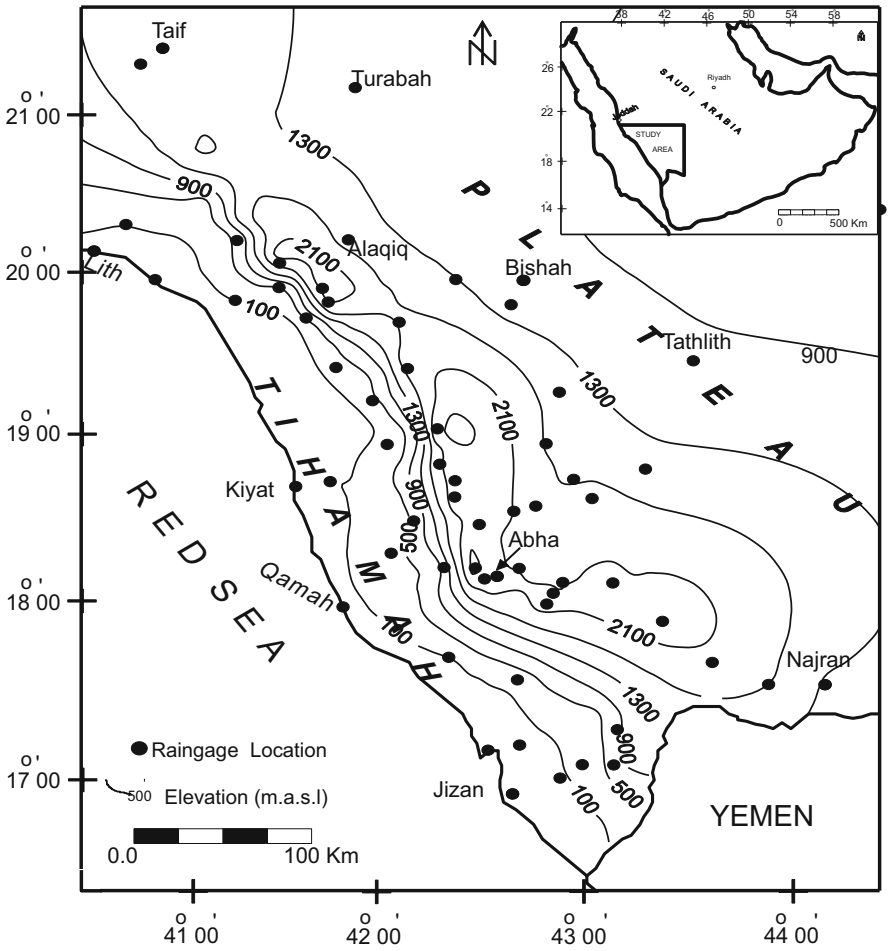


Fig. 6.24 Topographic map in southwest Saudi Arabia

3. They have continuous monthly rainfall.
4. They reflect the wide variety of environments within the study area literally from coast (Tihamah) and mountains and to leeside of the mountains (desert).

Descriptive statistics for the 63 stations are listed in Table 6.9 for these three regions. All together 25 stations are located within the coast, 24 within mountains, and 14 stations within the leeside.

Problems in the data, such as non-normality, trend, and outliers, should be fixed before developing any kind of model. Normality of the sample data distribution is known to improve the results from Kriging. Transformation is very important to make the data more symmetric, linear, and constant in variance. Since annual and

Table 6.9 Data set grouped into three regions

Region	Number of stations	Winter (mm)	Spring (mm)	Summer (mm)	Fall (mm)	Annual (mm)
Coast (Tihamah)	25	141[a]	199	238	152	684
		13[b]	6	5	3	55
		53.5[c]	51.4	58.6	53.7	217.2
		33.4[d]	51.3	58.2	40.5	160.3
Mountains	24	169	264	148	69	527
		20	80	16	10	150
		78.5	155.3	57.1	30.8	321.8
		47	53.5	35.4	18.8	120.4
Leeside (desert)	14	28	99	30	37	150
		7	29	3	3	44
		17.4	64.4	11.9	9.6	103.3
		6.6	17.8	8.22	8.3	27.7

[a]Sample maximum
[b]Sample minimum
[c]Sample mean
[d]Sample standard deviation

seasonal rainfall data are considered, it is pragmatic to find one transformation which works reasonably well for all. The Box-Cox transformation is widely used and can be easily managed, so that the skewness of transformed data $Z(x,t)$ becomes close to zero (Salas 1993).

However, rainfall histogram in arid regions, as stated by Hevesi et al. (1992), behaves as lognormal distribution. Hence, the transformation as $Y = \ln Z(x)$ is applied for determining approximately normal annual and seasonal data. It is accepted as normally distributed, if the computed Kolmogorov-Smirnov statistic (Dmax) is less than the corresponding critical value. The critical value for the 5 % level of significance is $D0.05 = 0.171$, which is greater than Dmax of transformed data. Thus the null hypothesis of the transformed data normality cannot be rejected at 0.05 level of significant. Further investigation can be done by visually observing the normal probability plots, and most of the data lie on a straight line for the transformed rainfall values. In addition, the skewness coefficients are reduced close to zero (Subyani 2005). Table 6.10 shows the statistics and normality test for original and transformed annual and seasonal data.

The back-transformed value, i.e., exp $Y(x)$, is a biased predictor. However, the unbiased expression for the Kriging estimates $Z(x)$ is given by Aboufirassi and Mariño (1984), Gilbert (1985), and Deutsch and Journel (1992). As

$$Z^*(x) = \exp\left\{Y^*(x) + \sigma_y^2/2\right\} \tag{6.66}$$

Table 6.10 Descriptive statistics and normality test for annual and seasonal data

Season	Mean (mm)	Median (mm)	St. dev. (mm)	Skewness	CV (%)	K−S (*Dm*)
Annual	232	178	151	0.8	0.65	0.18
Winter	55.0	41.0	43.1	1.15	0.78	0.20
Spring	94.0	81.0	68.0	0.72	0.72	0.14
Summer	47.6	32.0	47.3	1.84	0.99	0.17
Fall	35.2	21.0	33.2	1.44	0.95	0.19
Ln-annual	5.22	5.18	0.7	−0.7	0.13	0.10
Ln-winter	3.71	3.71	0.79	−0.051	0.21	0.07
Ln-spring	4.21	4.39	0.92	−0.65	0.22	0.11
Ln-summer	3.38	3.46	1.06	−0.24	0.31	0.08
Ln-fall	3.1	3.04	1.02	−0.15	0.33	0.08

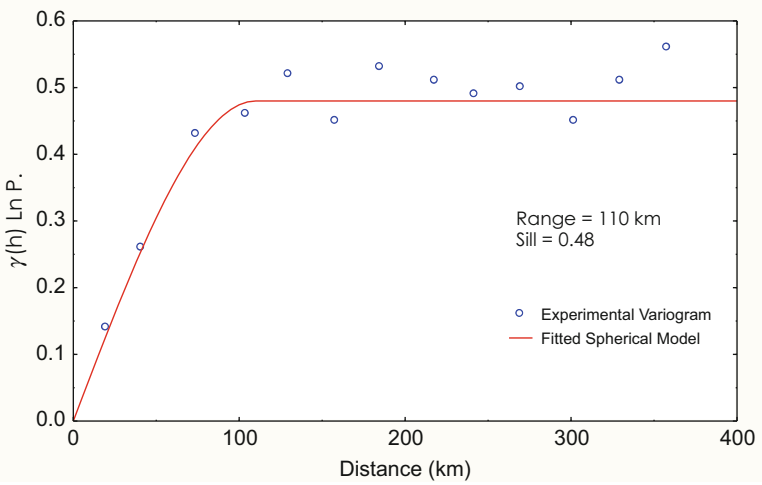


Fig. 6.25 Experimental and fitted SV model for Ln-annual rainfall

where $Z^*(x)$ is the original data in millimeters and $Y^*(x)$ is the natural logarithm and σ_y^2 is the lognormal Kriging variance. The estimation variance is given as

$$\sigma_{Z^*}^2 = (Z^*)^2\left[\exp\left(\sigma_y^2\right) - 1\right] \tag{6.67}$$

These two last expressions are used for constructing the rainfall isohyetals and their variances. Figure 6.25 shows the sample and fitted SV for the natural log of average annual rainfall (LnAAR). An isotropic spherical model with no nugget but with a sill equal to the sample variance of 0.48 and a range of 110 km is selected as the best representation of the spatial structure.

Table 6.11 SV cross-validation

Season	Model	N	MEE	RMSE
Ln-annual	Spherical	60	0.067	0.94
Ln-winter	Exponential	60	0.027	0.84
Ln-spring	Exponential	61	0.08	0.71
Ln-summer	Spherical	62	0.07	0.91
Ln-fall	Spherical	62	0.04	1.08

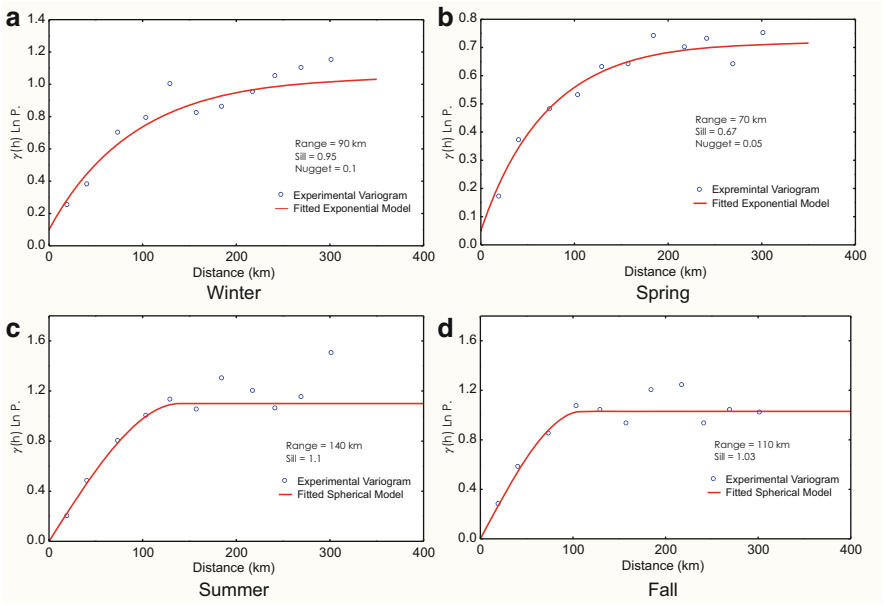


Fig. 6.26 Experimental and fitted SV models for Ln-seasonal rainfall

Cross-validation is applied to check whether the spherical model is adequate to describe the spatial correlation of the annual rainfall. For diagnostic check, the mean estimation error (MEE) and the root-mean-square errors (RMSE) for LnAAR are 0.067 and 0.94, respectively, which suggest the validity of the spherical model. Similar tests are performed for seasonal values and the results are shown in Table 6.11.

Figure 6.26 presents the sample and theoretical SVs and their parameters for winter, spring, summer, and fall seasons. In winter, the sample SV model is fitted with the small nugget as 0.05. The sill is 0.67 and the practical range of dependency, i.e., radius of influence, is 70 km (Fig. 6.26a).

In spring sample, SV behaved as an exponential model with nugget of 0.1. The sill is 0.95 with a practical range of dependency equal to 90 km (Fig. 6.26b). In the summer and fall, the sample SV behaves as a spherical model. The sill for these two seasons is around 1.00, and the ranges of correlation are within 140 km and 110 m, respectively (see Fig. 6.26c, d).

Kriging estimates are computed for the average annual and seasonal rainfall amounts. The sample SV represents the continuity structure quite well. The cross-validation supports selection of models and their parameters. For Kriging estimation and variance, the back-transformed values are applied.

The kriged isohyets for annual rainfall show a rapid increase in average from the Red Sea shoreline up to the mountains and a gradual decrease to the north and east parts of the study area (Fig. 6.27a).

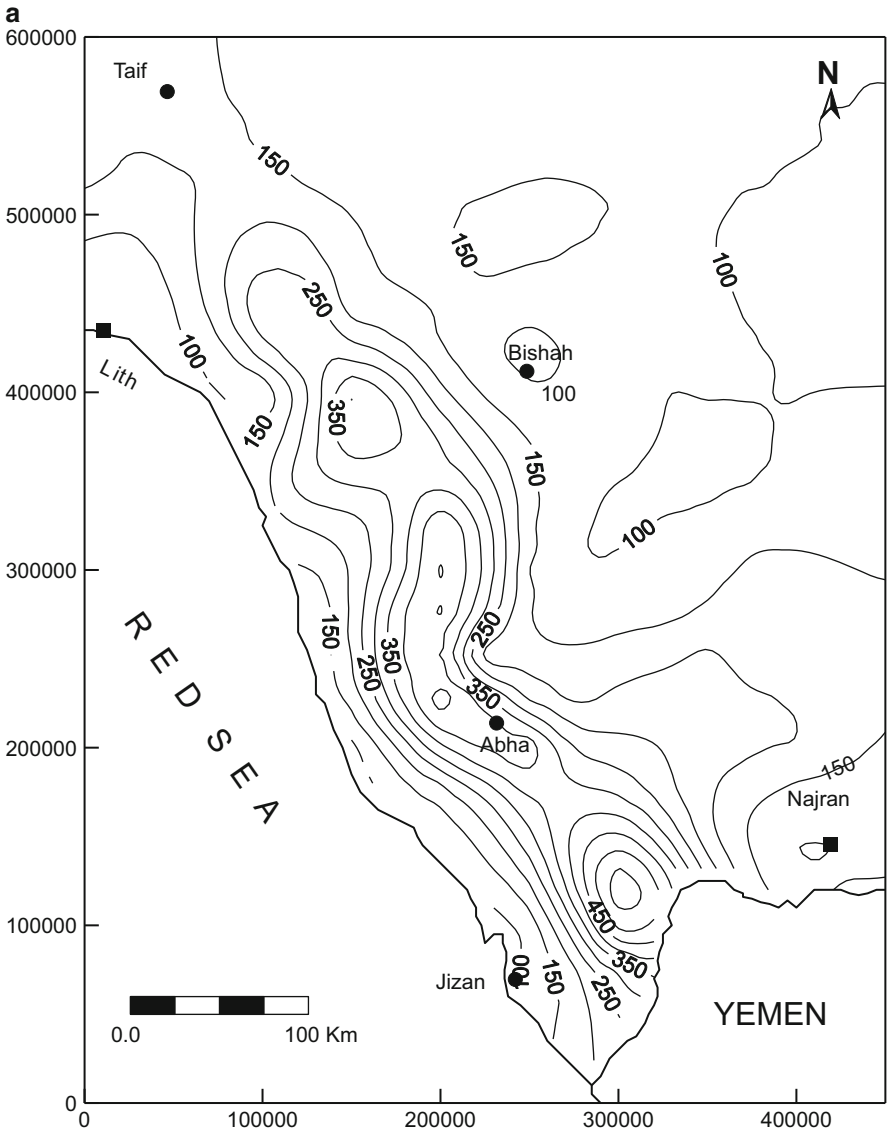


Fig. 6.27 Isohyetal map of Kriging for annual rainfall (mm)

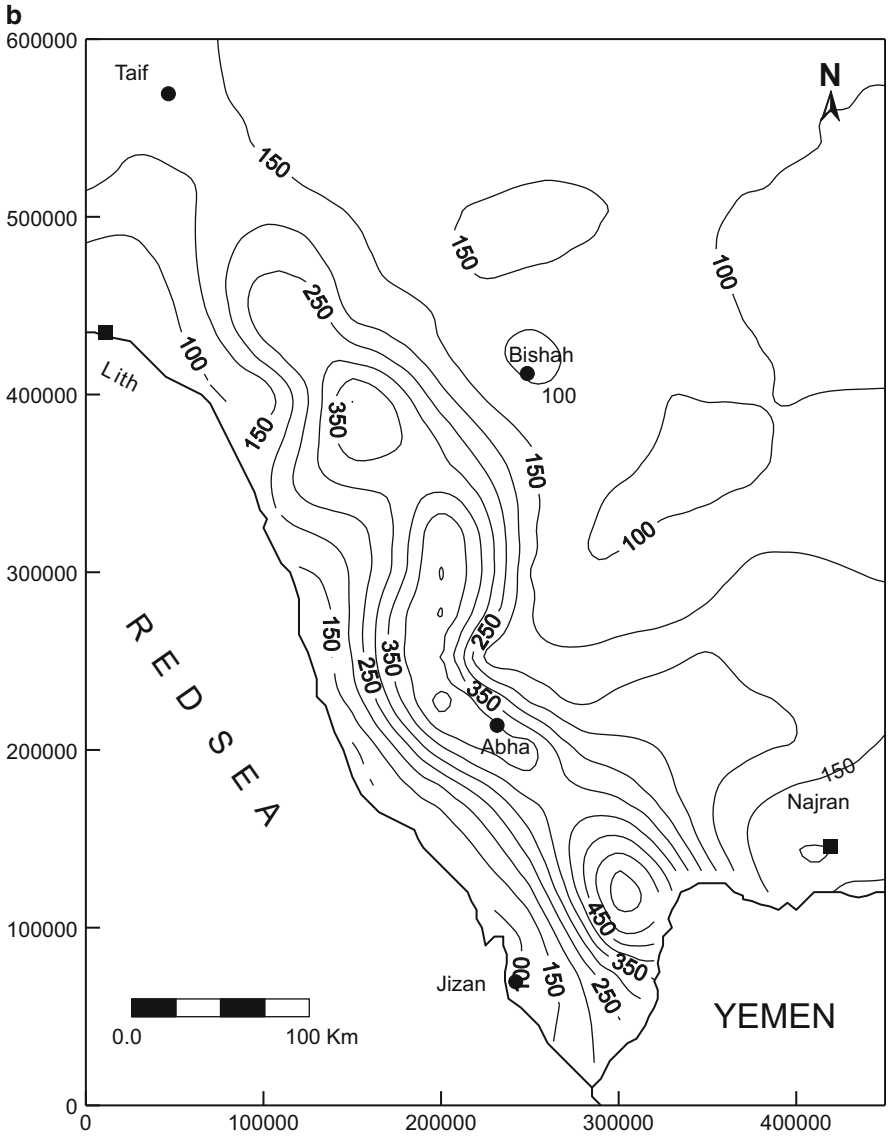


Fig. 6.27 (continued)

Orographic effects are produced toward the mountain area with the maximum kriged estimate exceeding 350 mm/year. In the east and northeast parts of the study area, Kriging estimates are around 100 mm/year. In the northern part with a moderate elevation reaching more than 1000 m, Kriging estimates also exceed

100 mm/year. This figure reflects the topographic variation similar to Fig. 6.24 with annual rainfall that generally increases with elevation.

Kriging variances indicate similar behavior to the average annual rainfall estimates. Small values near the clusters of stations in the mountain area (Fig. 6.27b) indicate high estimation accuracy whereas large values in the north, east, and northeast, hence low estimation accuracy areas owing to the scarcity of sample locations. Generally, high estimation variances appear at areas of lacking data.

During December–February months most of the time rainfall occurrences are associated with moist and cold northerly Mediterranean origin air, which is coupled with the local Red Sea convergence zone effect leading to orographic rainfall occurrences. Figure 6.28a shows that the Kriging estimates exceed 120 mm in the middle and northern section of the mountainous areas.

However, they do not exceed 30 mm in the south of Tihamah, because there is no Mediterranean or Red Sea effect, and the elevation is not high enough for orographic rainfall occurrence. On the other hand, the southern part of the study area receives less than 100 mm of rainfall due to the absence of monsoons. The plateau area (east and northeast parts of the study area) receives less than 20 mm, because it is located in the shadow (leeside) area. Kriging estimation variance map has a similar trend throughout the study area as shown in Fig. 6.28b.

In spring (March–May), the whole region comes under the influence of southeast monsoon air stream flow, the Red Sea convergence zone, and Mediterranean depression, which distribute the rainfall in all regions. The Kriging estimates give more detailed information about the rainfall distribution as shown in Fig. 6.29a.

Rainfall in this figure increases gradually from the Red Sea coast (40 mm) to the mountain where the highest amount of rainfall falls (more than 160 mm) and decreases to the plateau area, which receives about 100 mm. Generally, the southwest region of the Arabian Peninsula receives the highest amount of rainfall during the spring season compared to other seasons. This high amount of rainfall is a result of increasing African-Mediterranean interaction effect, where rainfall occurs orographically in the mountains, and southeast monsoon effect where the plateau and eastern slope receive more rainfall than the Red Sea coast. Kriging variances show an increase in estimation accuracy as shown in Fig. 6.29b.

In summer (June–August), the southwest monsoon flow from the Indian Ocean and the Arabian Sea is the predominant factor, which increases the rainfall along the Scarp mountains and low elevation areas in the south of the study area. Kriging estimates for summer season exceed 120 mm in the mountains and 160 mm in the foothills near the Yemen border at the southwestern corner of the Arabian Peninsula (Fig. 6.30a).

Rainfall decreases toward the northern part of the study area due to its distance from the monsoon effect, even though this area has a high elevation. Moreover, the Kriging variances show no change in estimation accuracy in the foothills and mountains, but there are changes in the plateau area as shown in Fig. 6.30b.

In fall (September–November), the local diurnal circulation and the southern air stream weaken. In other words, it is a transition period from summer to winter, and, in general, the area receives little amount of rainfall. The Kriging estimation in the foothills and the mountains in the southern part of the study area shows that they receive higher amount of rainfall than the northern areas, similar to the fall monsoon flow effects as shown in Fig. 6.31a. The Kriging variances show an

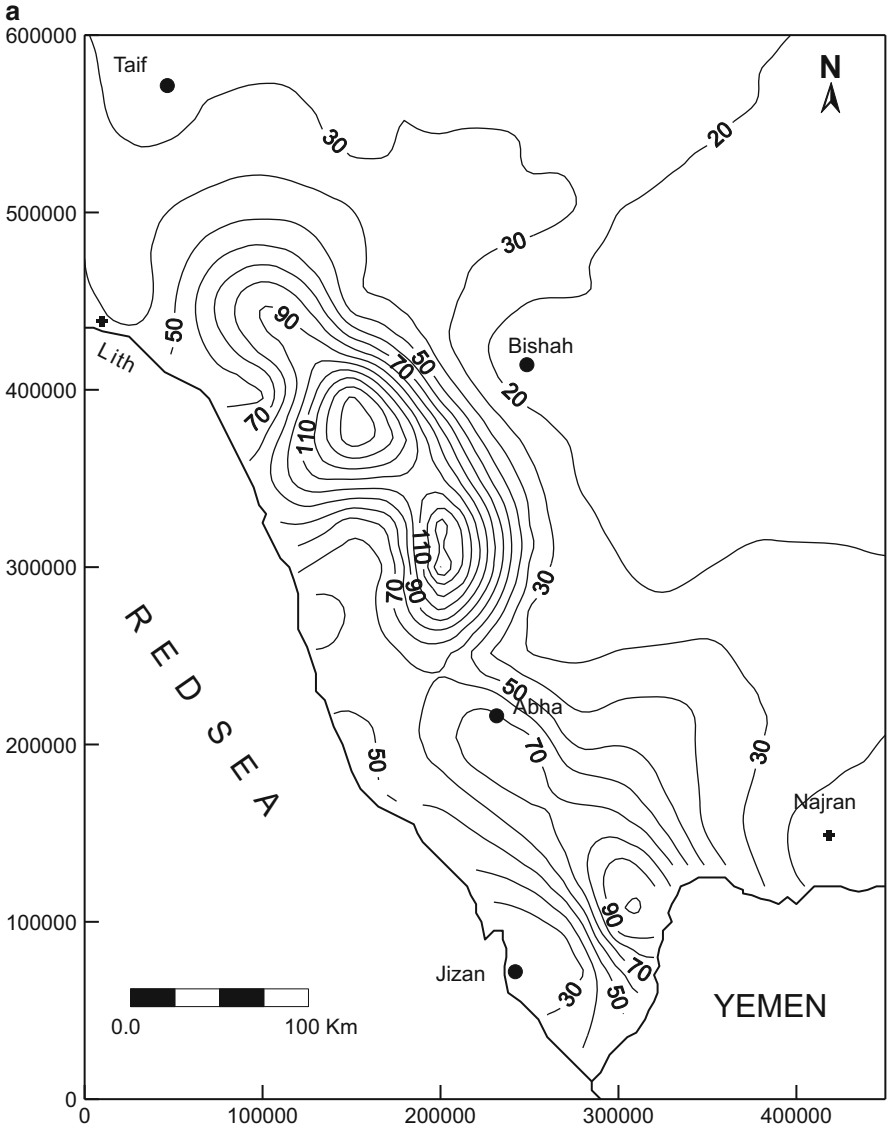


Fig. 6.28 Isohyetal map of Kriging for winter rainfall (mm)

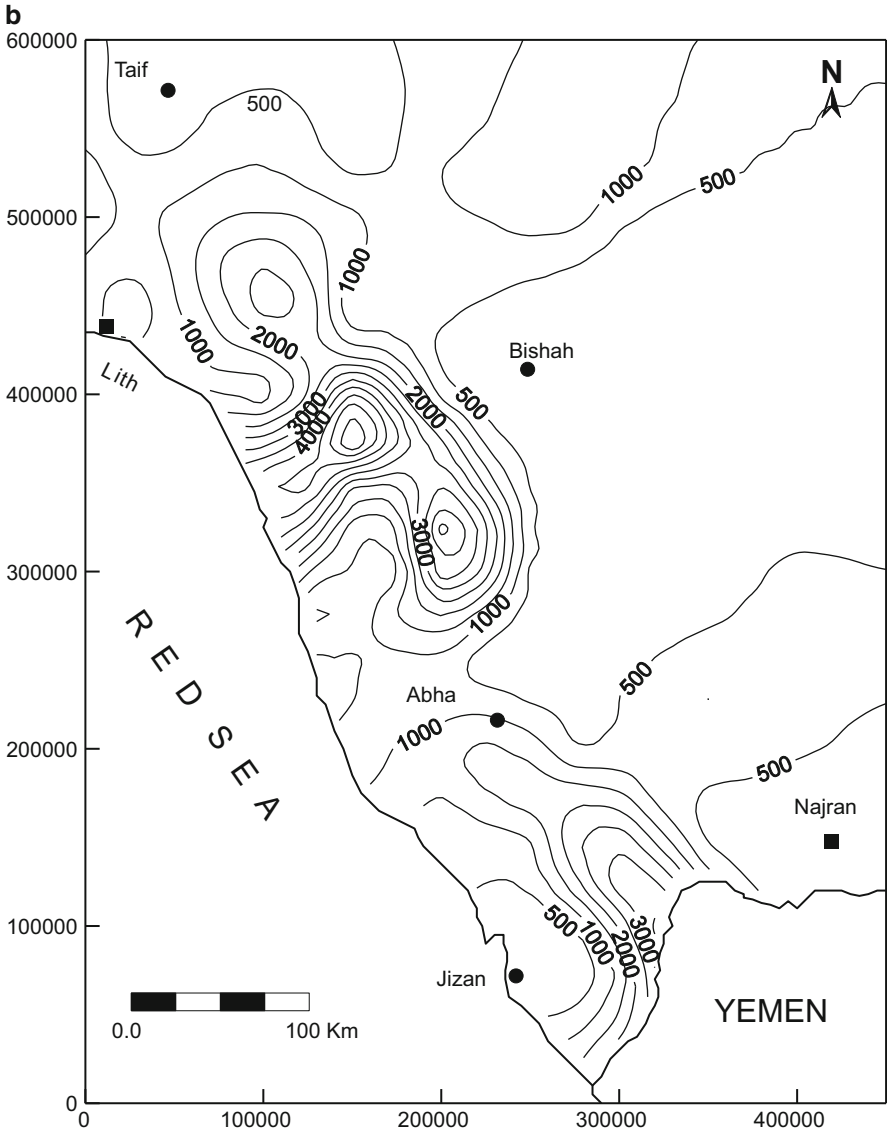


Fig. 6.28 (continued)

increase in estimation accuracy in the northern part of the study area, whereas there is no clear change in the southern part as shown in Fig. 6.31b.

Generally, rainfall is predominant in the northern mountain areas during winter as a result of the Mediterranean effect, and it is widespread in all regions during spring because of the local diurnal circulation effects. Orographic conditions are clear in winter and spring seasons. This orographic factor is also clear for the

appearance of the nugget effect in the exponential models in both winter and spring seasons. During summer, rainfall moves toward the south due to the monsoon flow effect with its southwesterly wind. However, during fall, as a transition season, the area becomes under the influence of monsoon as well as the local diurnal circulations. Figure 6.32 illustrates these spatio-seasonal variations of rainfall in the southwest of Saudi Arabia.

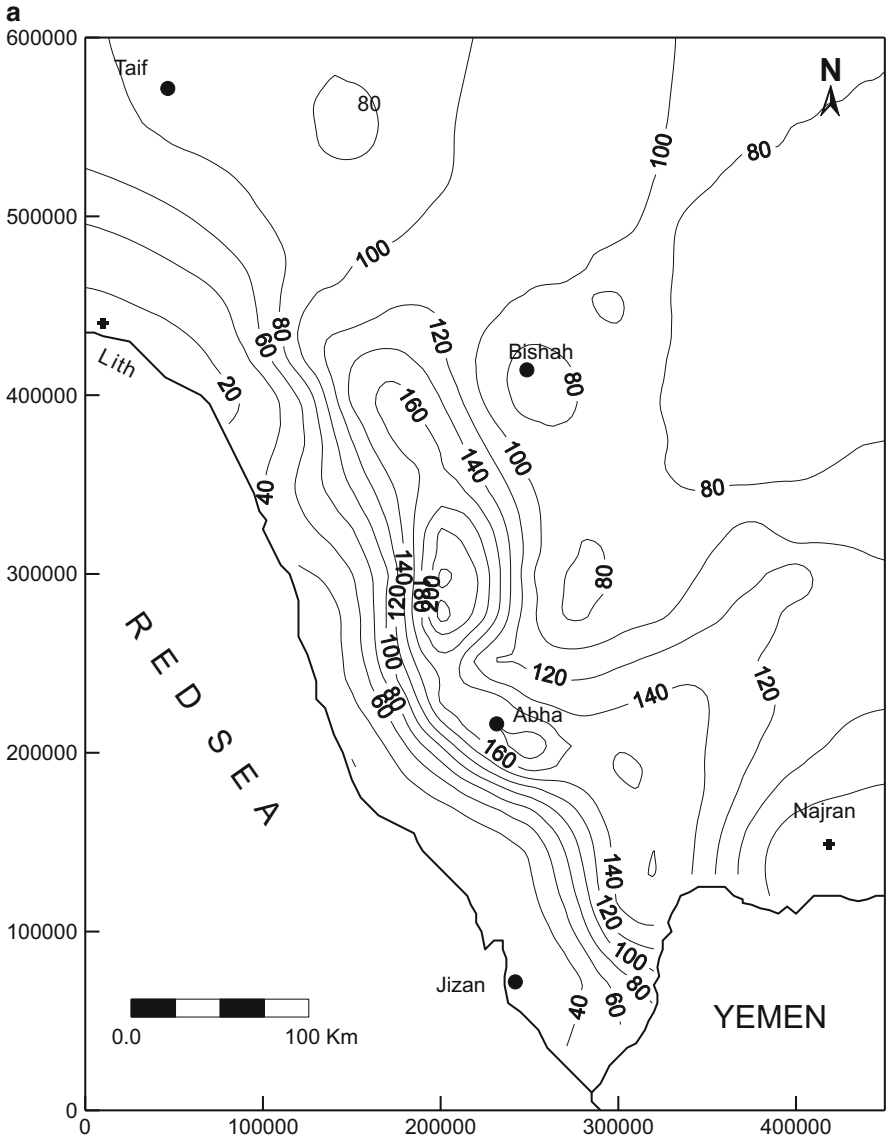


Fig. 6.29 Isohyetal map of Kriging for spring rainfall (mm)

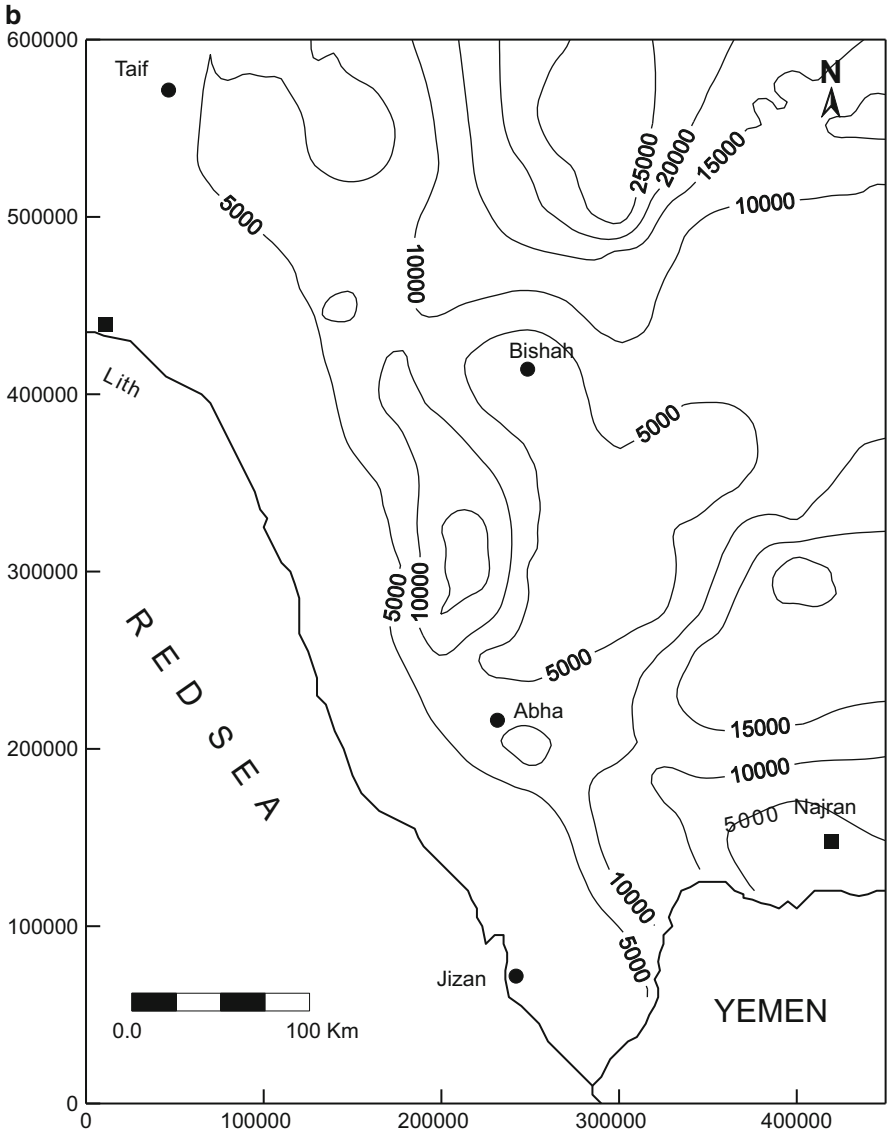


Fig. 6.29 (continued)

The Kriging estimation variances are also investigated concerning the spatial and temporal variation of rainfall in the study area for these four seasons. In space variation, the small value or high estimation accuracy of Kriging variances occurs in the mountainous areas in all seasons. Toward the east, north, and northeast of the study area, there is a consistent increase in variances implying low estimation

accuracy. In time variation, Kriging variance increases from winter to fall. These variations in space and time are due to several factors such as:

1. Clusters of stations in the mountain areas reflect higher estimation accuracy.

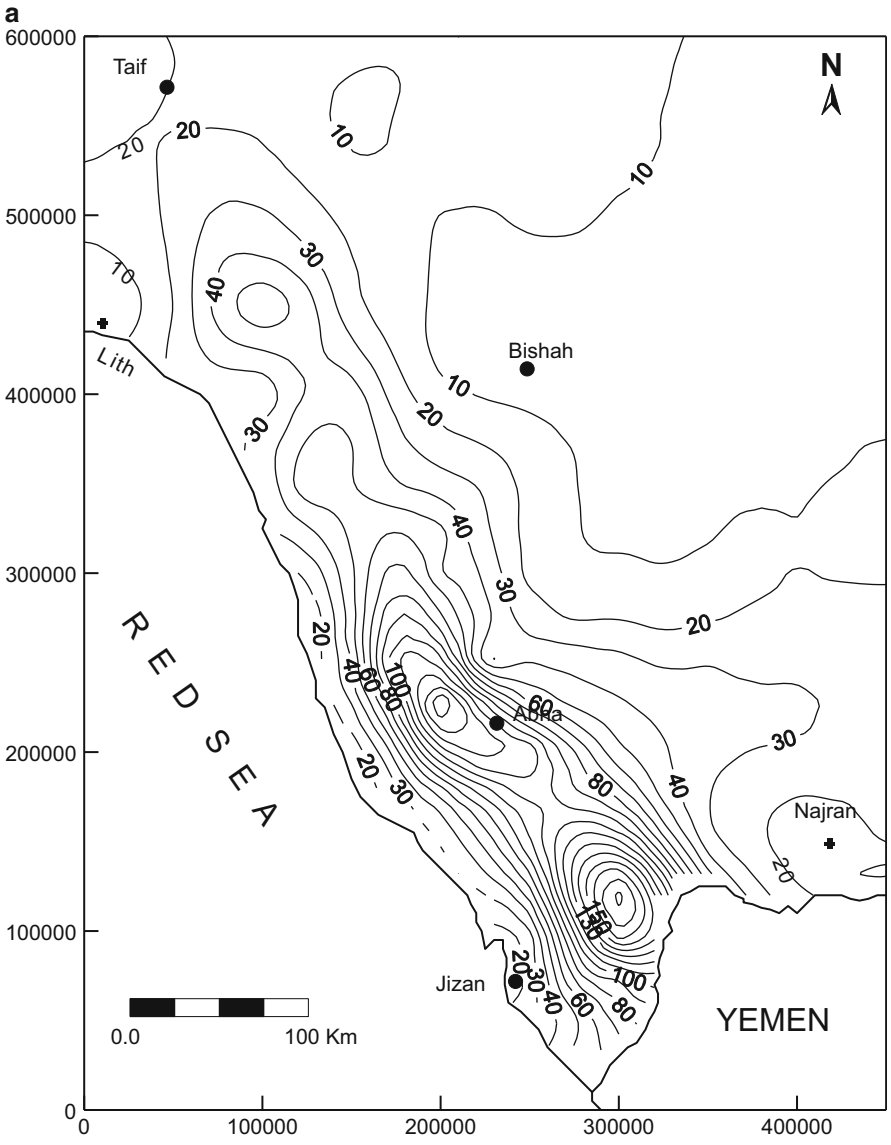


Fig. 6.30 Isohyetal map of Kriging for summer rainfall (mm)

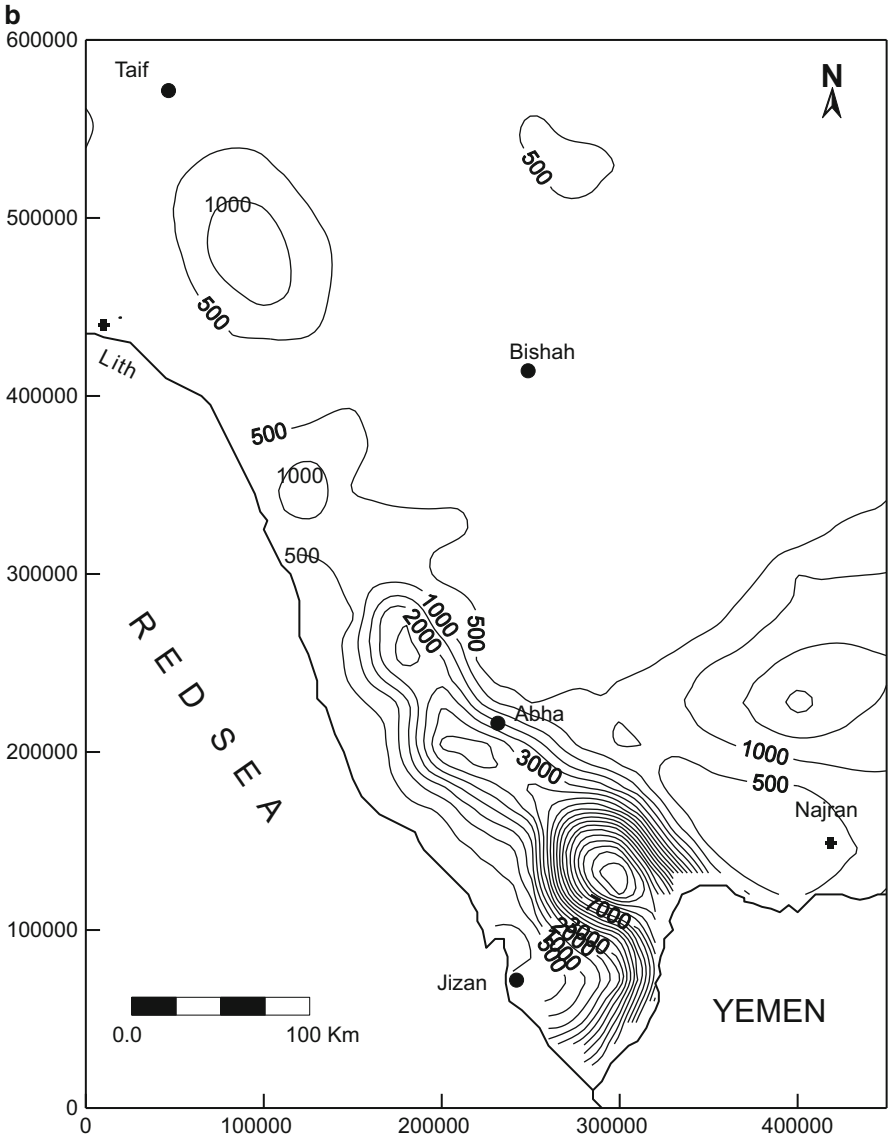


Fig. 6.30 (continued)

2. Scarcity of stations in north, east, and northeast areas reflects lower estimation accuracy.
3. High variability in summer and fall rainfall gives rise to low accuracy estimation compared with the somewhat lower variability in rainfall during winter and spring seasons.

The description of the rainfall variability in space and/or in time is among the fundamental requirements for a wide variety of human activities and water resources project design, management, and operation. Geostatistical methods are applied to develop new maps for the prediction of rainfall over different seasons. The assigned objectives of this study are to predict the magnitude and variation of

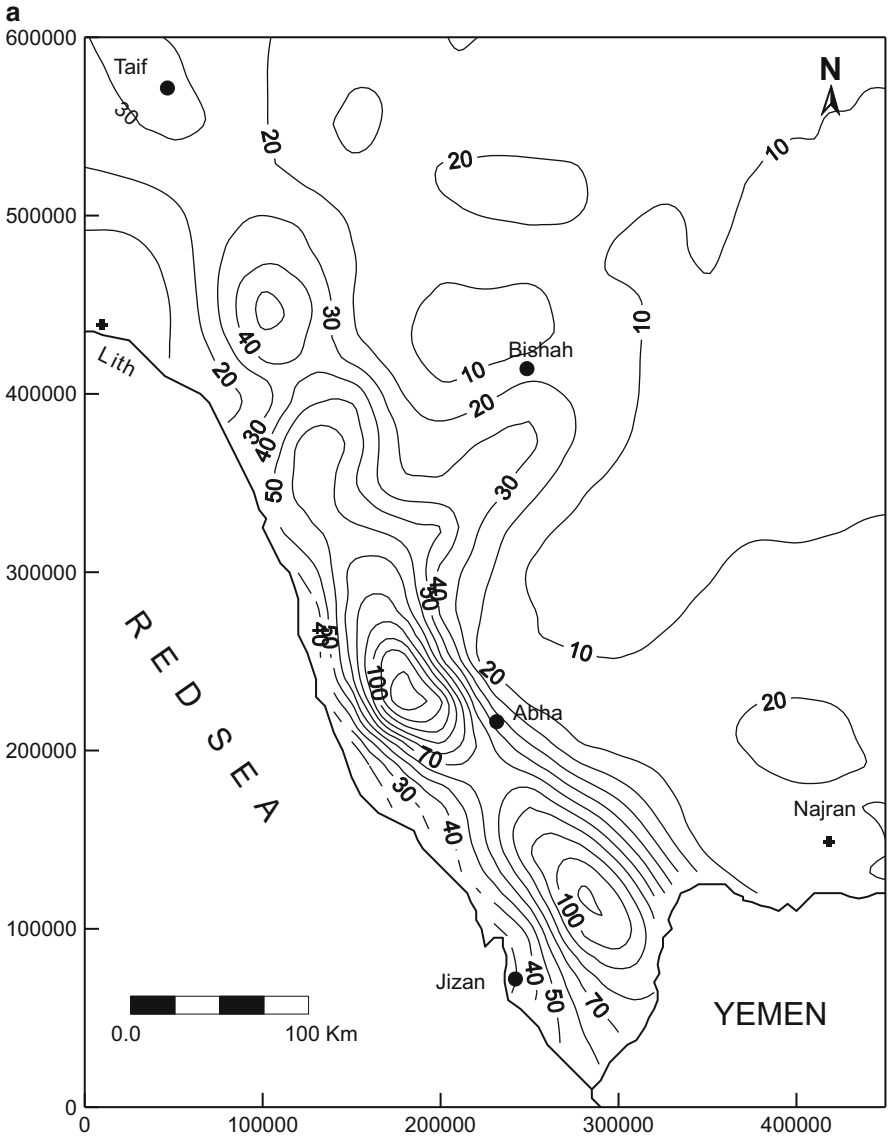


Fig. 6.31 Isohyetal map of Kriging for fall rainfall (mm)

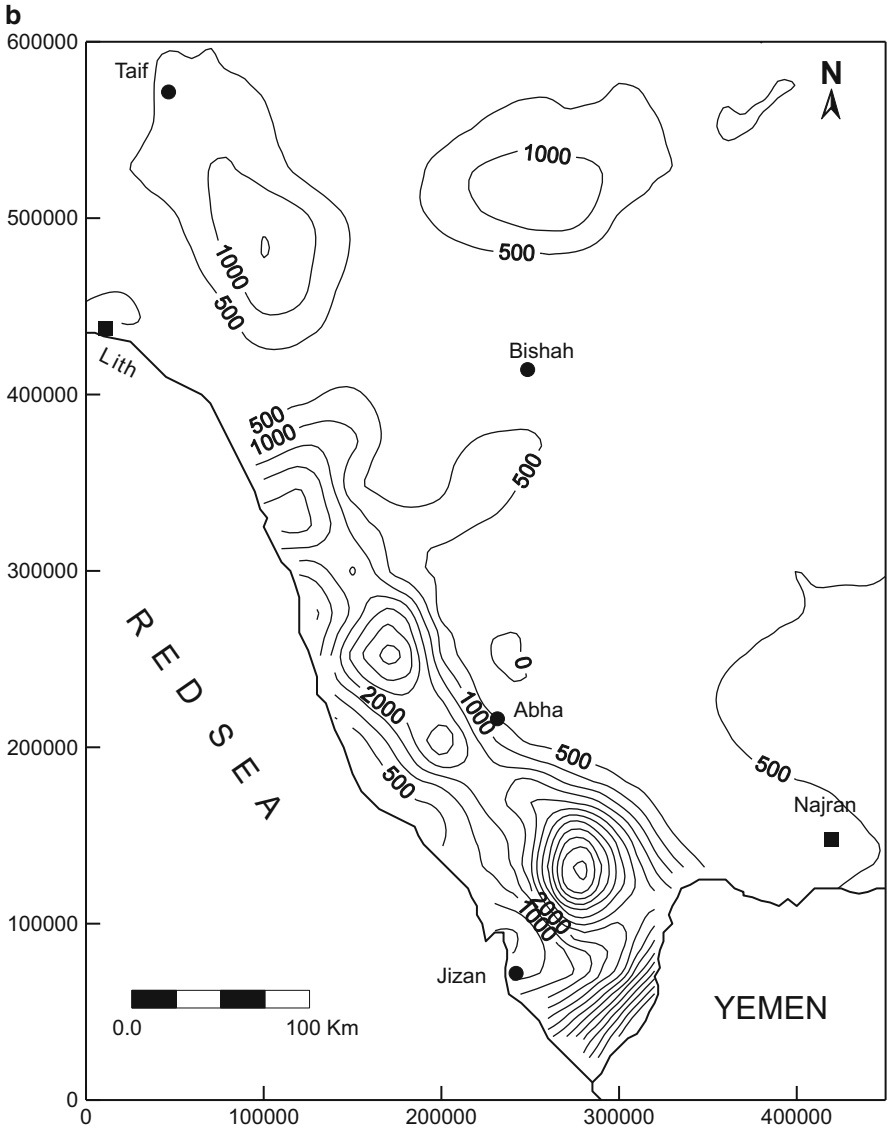


Fig. 6.31 (continued)

the rainfall in space as well as during different time periods. These techniques are applied to rainfall data gathered from meteorology station network covering the southwest region of the Arabian Peninsula. Rainfall in this area is characterized by high variation in spatial and temporal distributions.

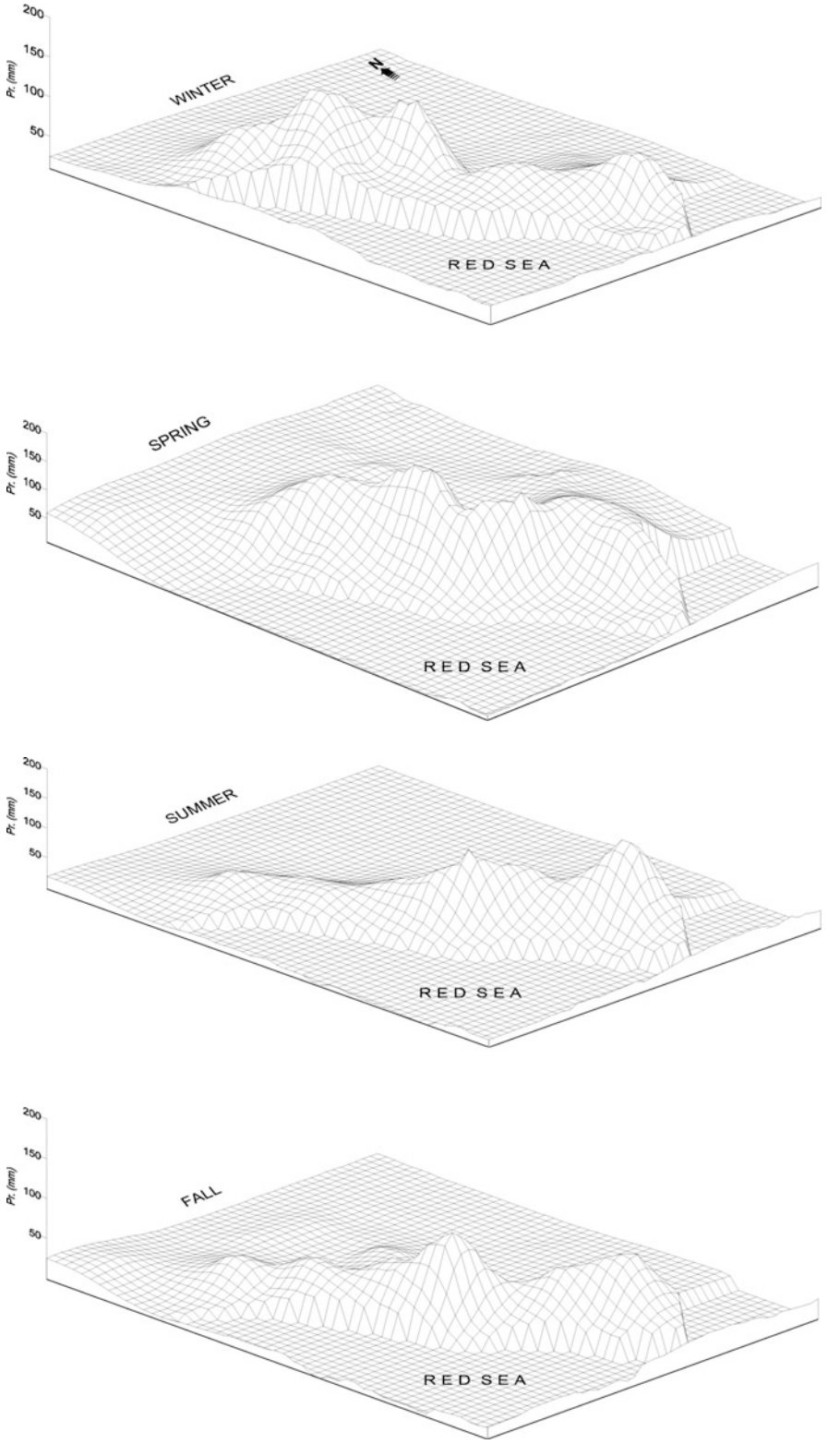


Fig. 6.32 Spatiotemporal Kriging maps for rainfall in southwest Saudi Arabia

References

Aboufirassi M, Marino MA (1984) A geostatistically based approach to the identification of aquifer transmissivities in Yolo Basin, California. Math Geol 16(26):125–137
Al-Jerash M (1985) Climatic subdivisions in Saudi Arabia: an application of principle component analysis. J Clim 5:307–323
Barnes SL (1964) A technique for maximizing details in numerical weather map analysis. J Appl Meteorol 3:396–409
Box GFP, Jenkins GM (1976) Time series analysis forecasting and control. Holden Day, San Francisco, 560 pp
Buell CE (1958) The correlation between wind and height on a isobaric surface II: summer data. J Meteorol 15(12):502–512
Carr J, Glass C (1984) Estimation of earthquake response spectra using kriging. In G. Verly et al (eds) Geostatistics for natural resources characterization. Proceedings of the 2nd NATO Advanced Study Institute of Geostatistics for Natural Resources Characterization: Part 2, D. Reidel, Dordrecht Holland, 745–752
Carr J, Glass C (1985) Treatment of earthquake ground-motion using regionalized variables. J Math Geol 17:221–241
Carr J, Bailey R, Deng E (1985) Use of indicator variogram for an enhanced spatial analysis. J Math Geol 18:197–213
Clark I (1979) Practical geostatistics. Applied Science Publishers, London, 281 p
Cressie NAC (1993) Statistics for spatial data, revisedth edn. Wiley, New York, 900 pp
Cressman GP (1959) An operational objective analysis system. Mon Weather Rev 87(10):367–374
David M (1977) Geostatistical ore reserve estimation. Elsevier Scientific Publishing Company, 364 pp
Davis JC (1986) Statistics and data analysis in geology. Wiley, New York, p 646
Delhomme JP (1978) Kriging in the hydrosciences. Adv Water Resour 1:251
Deutsch CV, Journel AG (1992) GSLIB geostatistical software library and user's guide. Oxford University Press, New York, TIC: 224174
Dixon R (1969) Orthogonal polynomial as a basis for objective analysis. Meteorological Office, Scientific Paper No. 30, Her Majesty's Stationery Office, London, 20 pp
Dixon R (1976) An objective analysis using orthogonal polynomials. In: Proceedings of the JOC Study Group Conference on four-dimensional Data Assimilation, Paris, GARP Program on Numerical Experimentation, Report No. 11, pp 73–85
Eddy A (1964) The objective analysis of horizontal wind divergence fields. Q J R Meteorol Soc 90:424–440
Eliassen A (1954) Provisional report on spatial covariance and autocorrelation of the pressure field. Institute of Weather and Climate Research Academy of Science Oslo Rep. 5
Fisher RA (1935) The design of experiments. Oliver and Boyd, Edinburg
Flattery TW (1970) Spectral models for global analysis and forecasting, Proceedings of the Sixth AWS Technical Exchange Conference, U. S. Naval Academy, 21–21 September, Air Weather Service Technical Report 242, pp 42–54
Fritsch JM (1971) Objective analysis of a two-dimensional data field by the cubic spline technique. Mon Weather Rev 99(5):1122–1143
Gandin LS (1963) Objective analysis of meteorological fields. Leningrad, Gidromet; Translated from Russian, Jerusalem, Israel Program for Scientific Translations, 1965, 242 pp
Gandin LS (1965) Objective analysis, lectures on numerical short-range weather prediction, WMO, Regional Training Seminar, pp 633–677
Gilbert S (1985) Introduction to applied mathematics. Wellesley-Cambridge Press, Wellesley, 02181
Glass CE (1978) Application of regionalized variables to micro-zonation. In: Proceedings 2nd International Conference on Micro-zonation for Safer Construction, vol 1, 509–512

Habib Z (1999) Optimum interpolation modeling supported by the cumulative semivariogram for spatial-temporal meteorological variables. Unpublished Ph.D. thesis, Istanbul Technical University, 156 pp

Hevesi JA, Istok JD, Flint AL (1992) Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I. Structural Analysis. J Appl Meteorol 31:661–676

Hubert P, Carbomnel JD, Chaouche A (1989) Segmentation des series hydrometeoraloques: application a des series de precipitation et de debits de L'Afrique de L'ouest. J Hydrol 110:349–367

Iruine KN, Eberthardt AK (1992) Multiplicative seasonal ARIMA models for lake Erie and Lake Ontario water levels. Water Res Bull 28(3):385–396

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, Oxford, 561 pp

Journel AG (1983) Non-parametric estimation of spatial distribution. Math Geol 15(13):445–468

Journel AG, Huijbregts CI (1978) Mining geostatistics. Academic, London, 710 p

Kalman RE (1960) A new approach to linear filtering and prediction problem. J Basic Eng 82, Series D(1):35–45

Kitanidis PK (1997) Introduction to geostatistics: applications in hydrogeology. Cambridge University Press, 249 pp

Krige DG (1951) A statistical approach to some basic mine evaluation problems on the Witwateround. J Chim Min Soc South Africa 52:119–139

Ledimet F, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observation: theoretical aspects. Tellus 38A(2):97–110

Lorenc AC (1981) A global three-dimensional multivariate statistical interpolation scheme. Mon Weather Rev 109:701–721

Loucks FD (1989) Modeling the great lakes hydrologic-hydraulic system. Ph.D. thesis, University of Wisconsin, Madison

Marsily (1986)

Martinez CA (1996) Multivariate geostatistical analysis of evapo-transpiration and precipitation in mountainous terrain. J Hydrol 174:19–35

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

Matheron G (1971) The theory of regionalized variables and its applications. Ecole de Mines, Fontainbleau

Mathier L, Fagherazzi L, Rasson JC, Bobee B (1992) Great Lakes net basin supply simulation by a stochastic approach. *INRS-Eau Rapp Scient(fique* 362, INRS-Fau, Sainte-Foy, 95 pp

Meleschko VP, Prigodich AE (1964) Objective analysis of humidity and temperature, Trudy Sympoziuma Pochislenyum Melodam Prognoza Pagody. Gidrometeoizdat, Moscow

Myers DE, Begovich CL, Butz TR, Kane VE (1982) Variogram models for regional groundwater chemical data. Math Geol 14:629–644

Olea RA (1975) Optimum mapping techniques using regionalized variable theory: series on spatial analysis no. 3. Kansas Geological Survey, Lawrence, 137 p

Olea RA (1999) Geostatistics for engineers and earth scientists. Kluwer Academic Publishers, Boston, 303 pp

Panofsky HA (1949) Objective weather map analysis. J Meteor 6:386–392

Privalsky V (1990) Statistical analysis and predictability of Lake Erie water level variations, In 'Proc Great Lakes Water Level Forecasting and Statistics Symposium', Eds. H.C. Hurtmann and M.J. Donalhue, Great Lake Commission, Ann Arbor, Michigan, pp 255–264

Salas M (1993) Analysis and modeling of hydrologic time series. In: Maindment DA (ed) Handbook of hydrology. McGraw-Hill, New York, pp 19.1–19.72

Sasaki Y (1958) An objective analysis based on variational method. J Meteorol Soc Jpn 36 (3):77–88

Schlatter TW (1988) Past and present trends in the objective analysis of meteorological data for now-casting and numerical forecasting, Eight Conference on Numerical Weather Prediction. American Meteorological Society. pp. j9–j25

Schlatter TW, Branstator GW (1987) Experiments with a three-dimensional statistical objective analysis scheme using FGGE data. Mon Weather Rev 115:272–296
Şen Z (1980) Adaptive Fourier analysis of periodic-stochastic hydrologic series. J Hydrol 46:239–249
Şen Z (1983) Predictive hydrologic models in hydrology. Nord Hydrol 14:19–32
Şen Z (1997) Objective analysis of cumulative semivariogram technique and its application in Turkey. J Appl Meteorol 36:1712–1720
Şen Z (2008) Wadi hydrology. CRC Lewis Publishers, Baco Raton, 386 pp
Şen Z, Habib Z (2001) Monthly spatial rainfall correlation functions and interpretations for Turkey. Hydrol Sci J 46(5):829–829
Şen Z, Kadıoğlu M, Batur E (1999) Cluster regression model and level fluctuation features of Van Lake, Turkey. Ann Geophys 17:273–279
Şen Z, Kadıoğlu M, Batur E (2000) Stochastic modeling of the Van Lake monthly fluctuations in Turkey. Theor Appl Clim 65:99–110
Seo DJ (1998) Real-time estimation of rainfall fields using radar rainfall and rain gage data. J Hydrol 208(1–2):37–52
Slivitzky M, Mathier L (1993) Climatic changes during the 20th century on the Laurentian Great Lakes and their impacts on hydrologic regime. NATO Adv Study Inst, Deauaille, France
Sneyers R (1992) On the use of statistical analysis for the objective determination of climate change. Meteorol Z 1:247–256
Student (1907) On the error of counting with a hemocytometer. Biometrika 5:351–360
Subyani AM (2004) Geostatistical study of annual and seasonal mean rainfall patterns in south-west Saudi Arabia. Hydrol Sci J 49(5):803–817
Subyani AM (2005) Hydrochemical identification and salinity problem of groundwater in Wadi Yalamlam basin, Western Saudi Arabia. J Arid Environ 60(1):53–66
Vannitseni S, Demaree G (1991) Detection et modelisation des secheresses an Sahel-proposition d'une nouvelle methodologie. Hydra! Continent 6(2):155–171
Yates F (1938) The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. Biometrika 30:444–466
Zadeh LA (1968) Fuzzy algorithms. Inf Control 12:94–102

Chapter 7
Spatial Simulation

Abstract Earth sciences phenomena are heterogeneous, anisotropic, non-uniform, and random especially in their spatial behaviors, and their exhaustive sampling is not possible, and the only scientific way for the regional and global assessment of these phenomena is through suitable mathematical modeling techniques. The main purpose is to deduce from limited and small samples the general characteristics and finally the generating mechanism of the phenomenon. Development in computer sciences enhanced time-consuming simulation studies, which developed into a feasible and practical tool to reveal nonlinear and nonequilibrium evolution of an earth sciences phenomenon based on the fundamental laws governing the nature. Nonlinearity and causality of a natural phenomenon are demonstrated to be consistently explained by simulations based on fundamental laws. Simulation methodology is timely born with practically usable starts of digital computers during 1950s. Simulation studies are essential because the existing analytical methods are not sufficiently powerful to prove the validity of the fundamental laws for naturally occurring phenomena, but simulation could do this effectively. Spatial pattern simulations in 1-, 2-, and 3D space are explained through autoregressive models, which are based on the spatial correlation function. Rock masses are fractured into different sizes, and they are intact from each other due to fracturing. Different simulation techniques are presented for dependent (persistent) and independent fracture patterns, which are exemplified in the field by scanline measurements. Along this line, rock quality designation (RQD) quantification is furnished through various simulation models, and final products are presented in terms of various charts that can be used in practical applications. Porous medium is very significant for water, oil, and gas storages, and therefore, its spatial behavior is significant for effective assessments, planning, operation, and management. Autorun spatial modeling technique is developed and applied for the simulation of porous medium. Finally, the application of the cumulative semivariogram technique is presented for intact length simulation in rock masses.

Keywords Autoregressive models • Autorun • Simulation • Fracture network • Rock quality designation • Intact length • Persistence • Markov process • ARIMA process • Multi-directional simulation

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5_7

7.1 General

Spatial simulation models of geological phenomenon such as ore grades, groundwater level elevations, porosity, chemical compositions of different lithological units, and fracture spacing are bound to be increasingly important, due to their ability to model the underlying generating mechanism of these phenomena. The first approximations in quantifying the geological phenomena were rather deterministic, and they did not take into account any factor of chance in their descriptions. In fact, the variability in these phenomena was known to geologists for many decades, but due to the deterministic training, the solutions sought were also deterministic. For instance, at most the arithmetical average of the data concerned is calculated, and then this value is treated as the best estimator of the phenomena. This way of calculation gave rise to overestimations, which consequently made the geologists aware of the fact that the variability within the data should definitely be taken into consideration (Krige 1951). This awareness led geologists to consider the frequency distribution functions of any spatial variable. Hence, a new trend of statistical methods in analyzing the geological data is started (Krumbein 1970; Agterberg 1975). It is a mere application of the statistical methods to process geologic data.

Later, it is recognized that in addition to variability inherent in the data, the interdependence either within the data themselves or with other variables is very important due to the continuity of spatial variables. In order to account for the dependence, either the correlation techniques (Matern 1960; Switzer 1965) or SVs are employed (Matheron 1965).

As a consequence of the aforementioned developments, the geologists started to think about a convenient model to simulate the spatial variables so as to be able to control them in the case of any change as well as in assessing the risks associated with the data. Of course, developments of digital computers had a great impact on this trend since without computers any simulation study is very tedious, needs great patience, and rather time-consuming, i.e., not practical.

Although simulation studies in other branches of science like engineering and economics have started long time ago, the simulation of geological phenomena is postponed. This is due to the fact that in other disciplines, simulation is needed only along one axis, for instance, time axis. Therefore, the simulation of 1D variable is extensively available in the literature (Box and Jenkins 1970).

General simulation models are possible for generation of anisotropic as well as isotropic synthetic patterns in 1D, 2D, or 3D, which have significance for the purpose of modeling geologic properties such as ore grades, reservoir porosity, mineral distribution, fracture spacing, aperture, orientation, etc. General procedures for such simulations by the autoregressive processes (including Markov processes) are given for the model parameter estimation and synthetic pattern generation. The model works on the square net basis and generates sequential pattern first along any desired direction for 1D simulation, and then 2D patterns are constructed with reference to two orthogonal 1D sequences. Applications to synthetic 2D pattern are

shown for isotropic cases with different model parameters. The extension of model to 3D space is readily available.

Because any geological phenomenon occurs in space, in general, it has to be simulated in 3D space. The first of spatial simulation models is proposed by Cliff and Ord (1973) who took into consideration an overall spatial correlation coefficient, a set of physical constants, and a random component all within a linear model concept. Later, Journel (1974) proposed a turning-band method whereby 2D or 3D geological variables are projected onto equally spaced lines, and then 1D simulations along these lines are transformed into 3D variables. On the other hand, Sharp and Aroian (1985) have presented herringbone method of simulation for 2D and 3D geological variables. However, as explained by Şen (1989a, b), there are some practical problems with their model.

7.2 3D Autoregressive Model

The general form of the simulation model for 3D is similar to autoregressive models (Markov processes), which can be given as

$$x_{i,j,k} = \alpha x_{i-1,j,k} + \beta x_{i,j-1,k} + \gamma x_{i,j,k-1} + \varepsilon_{i,j,k} \tag{7.1}$$

where $x_{i,j,k}$ is the spatial variable at a point with coordinates i, j, and k with respect to a reference point as in Fig. 7.1.

Herein, α, β, and γ are the model parameters and $\varepsilon_{i,j,k,}$ is a random component with zero mean and variance, σ_ε^2, and it independent from $x_{i\text{-}1,j,k}$, $x_{i,j\text{-}1,k}$, and $x_{i,j,k\text{-}1}$. This model has four parameters, namely, α, β, γ, and $\sigma_{\varepsilon 2}$, to be determined from an available set of data. In Eq. 7.1, the geologic variable, $x_{i,j,k}$, is a standard variable with zero mean and unit variance. This does not cause any loss of generality in

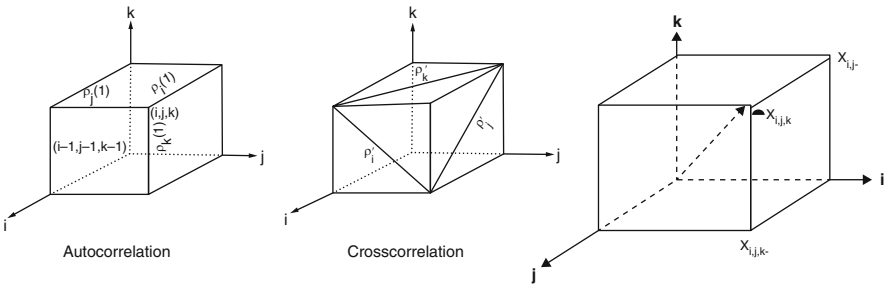


Fig. 7.1 3D space

simulation studies since after the simulation of standard variables, the final form of them can be found as

$$X_{i,j,k} = \mu_{i,j,k} + \sigma_{i,j,k} \times_{i,j,k} \tag{7.2}$$

in which $X_{i,j,k}$ is the spatial variable at position (i, j, k) with mean, $\mu_{i,j,k}$, and standard deviation, $\sigma_{i,j,k}$. Standardization procedure renders the original variable into a second-order stationary variable (Sen 1979a, b). Any simulation model has three major stages in arriving at its final goal. These are:

1. Given a set of observations $x_{i,j,k}$ $(i, j, k = 1, 2, \ldots, n)$, how to express model parameters as functions of data? This is referred to as parameter estimation stage.
2. How to check the validity of model for given data? This is the identification stage, which is achieved by calculating the residuals as

$$\varepsilon_{i,j,k} = x_{i,j,k} - \alpha x_{i-1,j,k} - \beta x_{i,j-1,k} - \gamma x_{i,j,k-1} \tag{7.3}$$

   If the model is valid, then the sequence of $\varepsilon_{i,j,k}$ should have an independent (completely independent) structure. This stage is referred to as the diagnostic checking for model suitability.
3. Usage of the suitable model to generate equally likely synthetic sets of data which are statistically indistinguishable from the original data, i.e., they should have the same distribution functions and average statistical parameters. This corresponds to the simulation of the underlying geological phenomenon. In real-time estimations, this is equivalent to prediction stage.

7.2.1 Parameter Estimation

The model parameters must be estimated from available data such that the simulated variables yield statistically indistinguishable parameters in the long run. One of the basic requirements in any simulation is to derive meaningful relationships between the model and the statistical parameters of the available data. The number of model parameters should be equal to the number of statistical parameters, and accordingly Eq. 7.1 allows preservation of four statistical parameters. The expectations (arithmetic averages) provide obsolete information because of the standardization procedure as $E(x_{i,j,k}) = E(x_{i-1,j,k}) = E(x_{i,j-1,k}) = E(x_{i,j,k-1}) = 0$, and $E(\varepsilon_{i,j,k}) = 0$. Further low-order statistical parameters are the spatial variance and three lag-one autocorrelation coefficients each along i, j, and k axes. For the

variance, the multiplication of both sides in Eq. 7.1 by $x_{i,j,k}$ and then taking expectations leads to

$$E\left(x_{i,j,k}^2\right) = \alpha E\left(x_{i-1,j,k}x_{i,j,k}\right) + \beta E\left(x_{i,j-1,k}x_{i,j,k}\right) + \gamma E\left(x_{i,j,k-1}x_{i,j,k}\right) + E\left(\varepsilon_{i,j,k}x_{i,j,k}\right) \tag{7.4}$$

where $E\left(x_{i,j,k}^2\right) = 1$, $E\left(x_{i-1,j,k}x_{i,j,k}\right) = \rho_i(1)$, $E\left(x_{i,j-1,k}x_{i,j,k}\right) = \rho_j(1)$ and $E\left(x_{i,j,k-1}x_{i,j,k}\right) = \rho_k(1)$. Because $\varepsilon_{i,j,k}$ is dependent on $x_{i,j,k}$ only, one can write $E\left(\varepsilon_{i,j,k}x_{i,j,k}\right) = E\left(\varepsilon_{i,j,k}^2\right) = \sigma_\varepsilon^2$; the substitution of these last expressions into Eq. 7.4 yields

$$\alpha\rho_i(1) + \beta\rho_j(1) + \gamma\rho_k(1) + \sigma_\varepsilon^2 = 1 \tag{7.5}$$

in which $\rho_i(1)$, $\rho_j(1)$, and $\rho_k(1)$ are the lag-one autocorrelation coefficients along the i, j, and k axes, respectively. It is obvious that Eq. 7.4 reduces to Sharp and Aroian (1985) simulation expression if $\alpha = \beta = \gamma = \varphi''$ and with their notations $\rho_i(1) = \rho_{100}, \rho_j(1) = \rho_{010}, \rho_k(1) = \rho_{001}$ and $\sigma_\varepsilon^2 = \sigma_\varepsilon^2$. Multiplication of both sides in Eq. 7.1 by $x_{i-1,j,k}$, $x_{i,j-1,k}$, and $x_{i,j,k-1}$, respectively, and taking expectations leads to three additional equations as

$$\rho_i(1) = \alpha + \beta\rho_k' + \gamma\rho_j' \tag{7.6}$$

$$\rho_j(1) = \alpha\rho_k' + \beta + \gamma\rho_j' \tag{7.7}$$

$$\rho_k(1) = \alpha\rho_j' + \beta\rho_i' + \gamma \tag{7.8}$$

in which $E\left(x_{i-1,j,k}x_{i,j-1,k}\right) = \rho_k'$, $E\left(x_{i-1,j,k}x_{i,j,k-1}\right) = \rho_j'$ and $E\left(x_{i,j,k-1}x_{i,j-1,k}\right) = \rho_i'$, where ρ_k', ρ_j' and ρ_i' indicate the lag-one cross-correlation coefficients on the $k = j = i =$ constant planes, respectively. Furthermore, because $\varepsilon_{i,j,k}$ is independent of $x_{i\text{-}i,j,k}$, $x_{i,j\text{-}1,k}$, and $x_{i,j,k\text{-}1}$, hence, $E\left(\varepsilon_{i,j,k}x_{i-1,j,k}\right) = 0$, $E\left(\varepsilon_{i,j,k}x_{i,j-1,k}\right) = 0$ and $E\left(\varepsilon_{i,j,k}x_{i,j,k-1}\right) = 0$. A sketch has been presented in Fig. 7.2 to show the auto- and cross correlations.

The four unknown model parameters can be obtained as a result of simultaneous solution of Eqs. 7.5, 7.6, 7.7, and 7.8. It is possible to arrange these equations in matrix form explicitly as

$$\begin{bmatrix} \rho_i(1) & \rho_j(1) & \rho_k(1) & 1 \\ 1 & \rho_k' & \rho_j' & 0 \\ \rho_k' & 1 & \rho_i' & 0 \\ \rho_j' & \rho_i' & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \lambda \\ o_\varepsilon^2 \end{bmatrix} = \begin{bmatrix} 1 \\ \rho_i(1) \\ \rho_j(1) \\ \rho_k(1) \end{bmatrix} \tag{7.9}$$

or implicitly

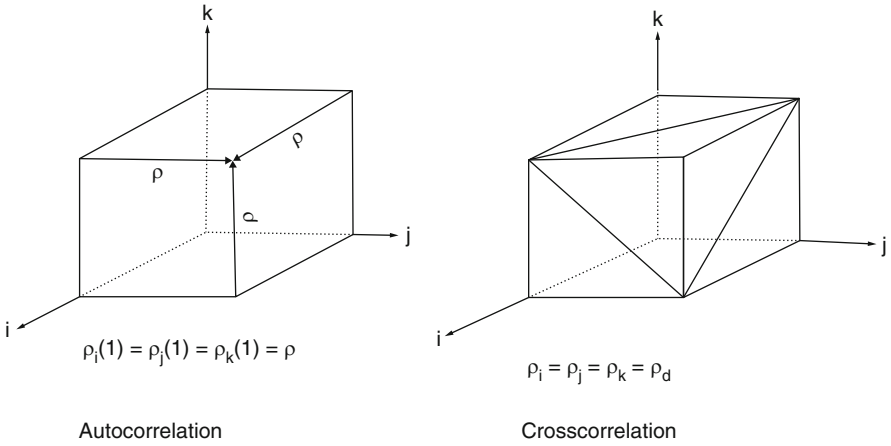


Fig. 7.2 Auto- and cross correlation representations in anisotropy model

$$AX = C \tag{7.10}$$

where A is (4×4) matrix of coefficients with its elements estimated from a given set of data; hence, it is known and dependent on the available data only; X is the unknown vector and includes the model parameters only; and, finally, C is the vector of lag-one autocorrelations. The solution of Eq. 7.8 requires matrix inversion of A after which the solution can be written as

$$X = A^{-1}C \tag{7.11}$$

It is necessary that A is a positive semi-definite matrix. Alternatively, a point worth to notice is that Eqs. 7.6, 7.7, and 7.8 include three unknowns (α, β, and γ) and, after their simultaneous solution, the substitution of these model parameters into Eq. 7.5 yields ${\sigma_\varepsilon}^2$. The matrix notation of these three equations is

$$\begin{bmatrix} 1 & \rho_k & \rho_k \\ \rho'_k & 1 & \rho'_i \\ \rho'_j & \rho'_i & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \rho_i(1) \\ \rho_j(1) \\ \rho_k(1) \end{bmatrix} \tag{7.12}$$

in which the coefficient matrix is symmetric and positive semi-definite. Application of Cramer's rule leads to the following results:

$$\alpha = \frac{\rho_i(1) + \left(\rho'_k\rho'_i - \rho'_j\right)\rho_k(1) + \left(\rho'_j\rho'_i - \rho'_k\right)\rho_j(1) - \rho'^2_j\rho_i(1)}{1 - \left(\rho'^2_i + \rho'^2_j + \rho'^2_k\right) + 2\rho'_i\rho'_j\rho'_k} \tag{7.13}$$

$$\beta = \frac{\rho_j(1) + \left(\rho'_i\rho'_j - \rho'_k\right)\rho_i(1) + \left(\rho'_j\rho'_k - \rho'_i\right)\rho_k(1) - \rho'^2_j\rho_j(1)}{1 - \left(\rho'^2_i + \rho'^2_j + \rho'^2_k\right) + 2\rho'_i\rho'_j\rho'^2_k} \tag{7.14}$$

and

$$\gamma = \frac{\rho_k(1) + \left(\rho'_j\rho'_k - \rho'_i\right)\rho_j(1) + \left(\rho'_k\rho'_i - \rho'_j\right)\rho_i(1) - \rho'^2_k\rho_k(1)}{1 - \left(\rho'^2_i + \rho'^2_j + \rho'^2_k\right) + 2\rho'_i\rho'_j\rho'_k} \tag{7.15}$$

It is clear from Eqs. 7.13, 7.14, and 7.15 that the model parameters (α, β, and γ) are functions of the correlation structure of the geological phenomenon concerned. There are three autocorrelation and three cross-correlation coefficients to be estimated from the data. Hence, the total number of statistical parameters including the average value and the variance to be extracted from the data is equal to eight provided that the geological phenomenon considered is homogeneous.

7.2.2 2D Uniform Model Parameters

In a homogeneous geological phenomenon, the mean and standard deviation values of the samples are independent of spatial variations. Such a phenomenon is said to be second-order stationary. In addition, if it is also isotropic, then the autocorrelations and cross correlations are equal to each other among themselves. This is tantamount to saying that $\rho_i(1) = \rho_j(1) = \rho_k(1) = \rho$ and $\rho'_i = \rho'_j = \rho'_k = \rho_d$ where ρ and ρ_d are isotropic serial and diagonal correlation coefficients. In other words, these correlations do not depend on axial directions as shown in Fig. 7.3. Furthermore, in an isotropic medium, the autocorrelations along any direction is dependent only on the variability along this direction. Hence, the dimensions of cubic blocks in spatial simulation become significant. For the sake of simplicity with no loss of generality, herein, the dimensions are assumed as units. Hence, the isotropic

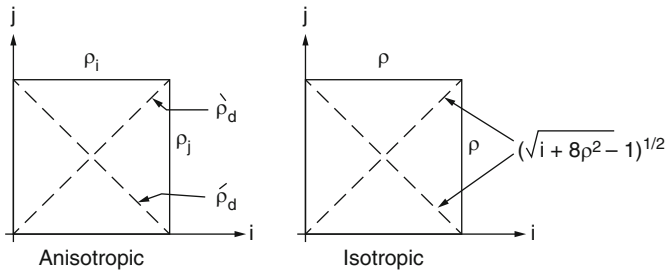


Fig. 7.3 Auto- and cross correlation representations in isotropic model

Fig. 7.4 2D isotropic model parameter

autocorrelation, ρ, in Fig. 7.3 should be interpreted as the correlation between the adjacent corner values of the phenomenon on a square mesh.

In 2D simulation this value will reduce into a square unit as shown in Fig. 7.4.

The model parameter estimations are found from Eqs. 7.13 and 7.14 by substituting the isotropic medium auto- and cross correlations, i.e., $\rho'_k = \rho_k(1) = 0$, $\rho_i 1 = \rho_j(1) = \rho$ and $\rho'_i = \rho'_j = \rho_d$

$$\alpha = \frac{\rho}{1 + \rho_d} \tag{7.16}$$

and β has identical value to α. With this the 2D model mathematical expression becomes

$$x_{i,j} = \alpha\left(x_{i-1,j} + x_{i,j-1}\right) + \varepsilon_{i,j} \tag{7.17}$$

The main problem in a 2D isotropic simulation is to satisfy the equality of cross diagonal correlations. This has been treated by Şen (1989a, b), and, with the current notations, the diagonal correlation coefficient should have the following form:

$$\rho_d = \frac{\mathscr{L}1 + 8\rho^2 - 1}{2} \tag{7.18}$$

The variance of the independent random variable, $\varepsilon_{i,j}$, can be found from Eq. 7.9 with necessary substitutions as

$$\sigma_{\varepsilon}^2 = 1.0 - \frac{2\rho^2}{(1+\rho_d)} \tag{7.19}$$

Different interpretations of Eqs. 7.16, 7.17, 7.18, and 7.19 indicate already known facts in the simulation literature as follows:

1. If the diagonal correlation is unity ($\rho_d = 1$), then Eqs. 7.16 and 7.19 reduce to the 1D case for which $\alpha = \rho$ and $\sigma_{\varepsilon}^2 = 1 - \rho^2$. These correspond to the properties of lag-one Markov process.
2. For zero autocorrelation ($\rho = 0$), it is possible to see that $\alpha = \beta = 0$, $\rho_d = 0$, and $\sigma_{\varepsilon}^2 = 1$. These properties imply independent random variable with zero mean and unit variance.

For the square net with dimensions ($n \times n$), the generation procedure requires first the generation of n normally distributed independent random variables $\varepsilon_{1,j}$ ($j = 1, 2, \ldots, n$) with zero mean and variance equal to ($1 - \rho^2$). Subsequently, these variables are converted into an autocorrelated row sequence by means of the first-order Markov process as

$$x_{1,j} = \alpha x_{1,j-1} + \varepsilon_{1,j} \; (j = 1, 2, \ldots, n) \tag{7.20}$$

The first column sequence is also generated in the same way except that the initial ($n-1$) normally distributed independent variables with zero mean and variance equal to ($1-\rho^2$) are generated. It is important that the first row value, $\varepsilon_{1,1}$, is used as the initial value for column Markov model which is expressed as

$$x_{i,1} = \alpha x_{i-1,1} + \varepsilon_{i,1} (i = 2, 3, \ldots, n) \tag{7.21}$$

in which $x_{1,1}$ is taken equal to $\varepsilon_{1,1}$. Then the remainder rows (or columns) are generated through the use of 2D model as

$$x_{i,j} = \alpha\left(x_{i-1,j} + x_{i,j-1}\right) + \varepsilon_{i,j} \; (i, \; j = 2, 3, \ldots, n) \tag{7.22}$$

Prior to the generation of related 2D isotropic variables, it is necessary to generate an array of $(n-1) \times (n-1)$ independent normal variables, $\varepsilon_{i,k}$, ($i, j = 2, 3, \ldots, n$) with zero mean but variance ${\sigma_{\varepsilon}}^2$ given in Eq. 7.19. These independent variables are converted into autoregressive variables through Eq. 7.22 by taking into consideration the two previously generated 1D sequences. The simulation in 2D can be completed either row by row or column-wise.

The generation procedure is applied with different autocorrelation coefficients, and the sample plots are presented in Fig. 7.5. The points in these figures indicate generated variables that have values greater than the mean value, which is taken as

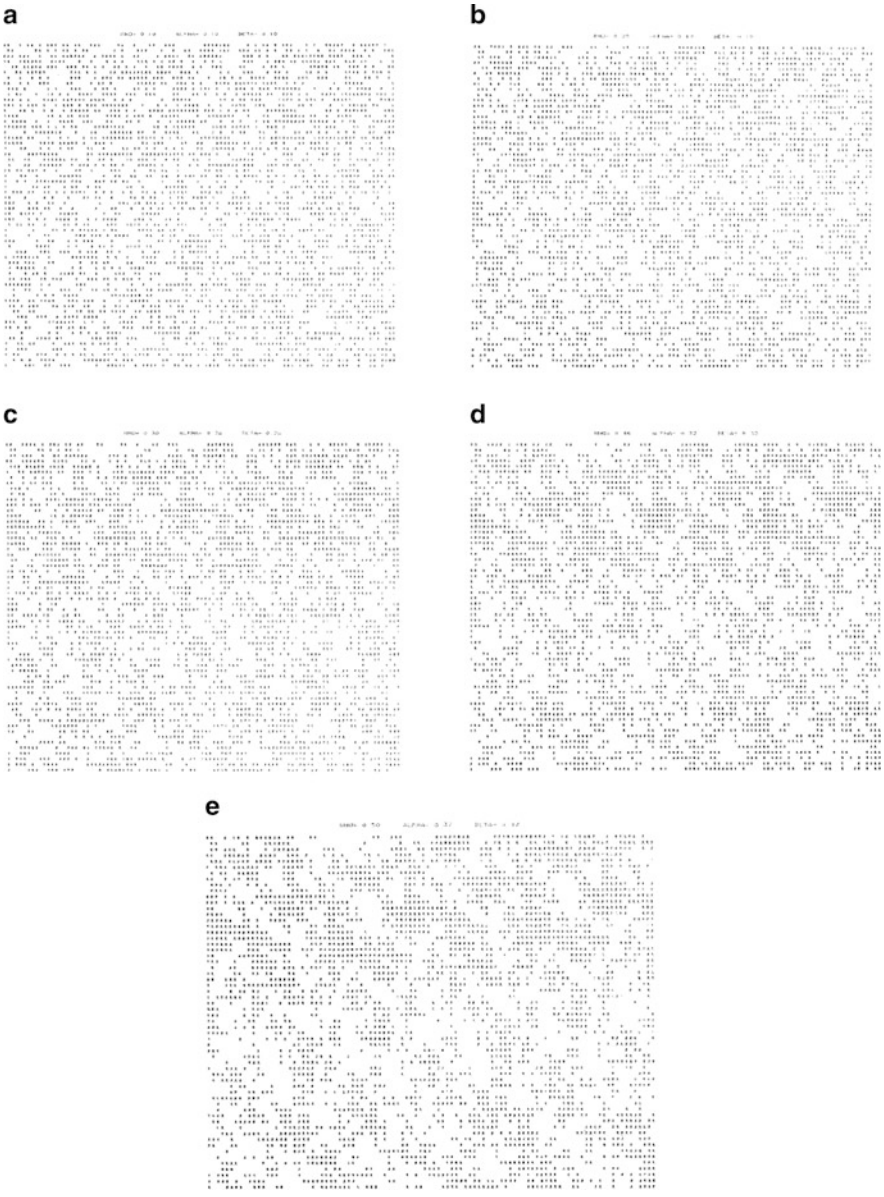


Fig. 7.5 2D isotropic pattern for (**a**) $\rho = 0.1$; (**b**) $\rho = 0.2$; (**c**) $\rho = 0.3$; (**d**) $\rho = 0.4$; (**e**) $\rho = 0.5$

zero. Correspondingly, blanks are the locations of variables less than the mean value. The visual inspection of these figures indicates that they are isotropic which means to say that there is no preferential direction or stratification along any direction.

7.2.3 Extension to 3D

The simulation procedure extension into 3D space is achieved by the execution of the following three stages:

1. 1D variables are generated along orthogonal i, j, and k directions according to lag-one Markov process. The most important point at this stage is that these three orthogonal direction sequences have a common original random variable which is $\varepsilon_{1,1,1}$.
2. Three layers are generated from the three unidirectional sequences in the previous stage according to the 2D generation scheme through similar equations to Eq. 7.22.
3. 3D isotropic simulation model parameters can be found from Eqs. 7.13, 7.14, and 7.15 by considering that $\rho_i(1) = \rho_j(1) = \rho_k(1) = \rho$ and $\rho_i' = \rho_j' = \rho_k' = \rho_d$ which leads to the same value of model parameters $\alpha = \beta = \gamma$ as

$$\alpha = \frac{\rho(\rho_d - 1)}{2\rho_d' - \rho_d - 1} \tag{7.23}$$

Hence, the 3D isotropic simulation model takes the form

$$x_{i,j,k} = \alpha\left(x_{i-1,j,k} + x_{i,j-1,k} + x_{i,j,k-1}\right) + \varepsilon_{i,j,k} \tag{7.24}$$

The substitution of the parameters into Eq. 7.5 leads after some algebra to the variance of 3D independent variables as

$$\sigma_\varepsilon^2 = 1 - \frac{3\rho^2(\rho_d - 1)}{2\rho_d^2 - \rho_d - 1} \tag{7.25}$$

Hence, to complete 3D simulation first, normal independent array $(n-1) \times (n-1)$ is generated with zero mean and variance σ_ε^2 as given in Eq. 7.25. These variables are combined together with the layer autocorrelated variables in the previous stage through Eq. 7.24 parallel to any one of the (i, j), (i, k), and (j, k) layer, and the procedure is repeated likewise until the desired domain is covered.

7.3 Rock Quality Designation (RQD) Simulation

A number of workers have devised various approaches that attempt to standardize and quantify descriptions of rock masses which are dissected by sets of discontinuities. The very first work was due to Terzaghi (1946) who was mainly concerned with the rock defects and loads on tunnel supports. Later, Deere (1963), Coates (1964), Barton et al. (1974), Bieniawski (1974), Goodman (1976), Şen and Kazi

(1984), Şen (1984, 1989a, b), Ege (1987), and many others proposed either statistical or empirical or analytical methodologies or modifications of existing ones in order to describe the discontinuities quantitatively. All of the aforementioned studies sought average values for discontinuities, and their probabilistic descriptions were not mentioned at all. It is the main purpose of this section to seek the PDF for some of the basic rock quality descriptions.

7.3.1 Independent Intact Lengths

Three important geometric properties of discontinuities that are of common interest to engineering geologists are their density (spacing frequency), size (trace length area), and orientation (strike dip direction cosines). The measure that is adopted herein is the spacing between two successive discontinuities, which is referred to as intact length. The intact lengths along any scanline may be measured by the separation of the intersections of discontinuities with sample lines in different directions on the same rock outcrop as shown schematically in Fig. 7.6.

Although there exist several descriptions of rock masses proposed by the International Society for Rock Mechanics (1978), the discontinuity spacing, i.e., intact lengths, deserves special attention and detailed study for effective assessment of rock quality. A simple mean for indicating the rock quality is proposed by Deere (1964) as rock quality designation (RQD). It is, in fact, the ratio of total intact lengths each of which is greater than a predesignated threshold value to the total length of scanline. Due to its relative simplicity, RQD has been used extensively in

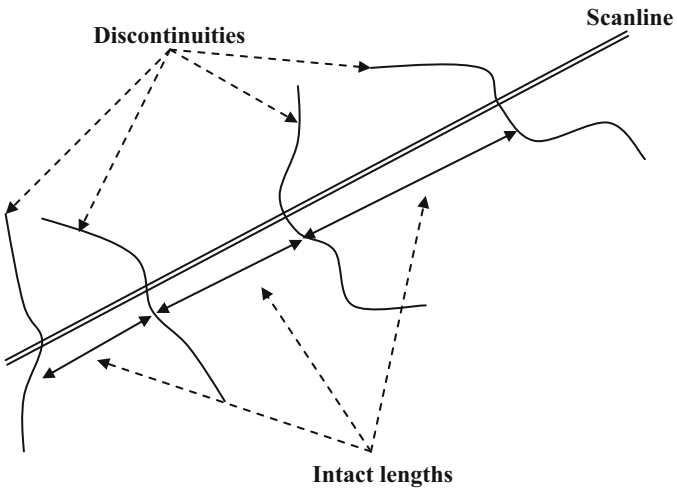


Fig. 7.6 Scanline survey of discontinuities

the rock classification for engineering purposes. Directly RQD or its slight modifications have been employed in a variety of engineering applications. For instance, Piteau (1970) has used RQD for rock slope stability, Barton et al. (1974) in the design of tunnel support, Louis and Pernot (1972) in dam foundation analysis of permeability and Cording and Mahar (1978) in underground chamber design in rocks, and Bieniawski (1974) and Kulhawy (1978) in estimating strength of rock materials.

Most of the researchers so far in the literature have concentrated in the scanline measurement evaluations analytically (Hudson and Priest 1979; Priest and Hudson 1976, 1981; Şen and Kazi 1984; Şen 1984; Kazi and Şen 1985) or empirically (Cruden 1977; Wallis and King 1980) or simulation on digital computers by Goodman and Smith (1980) and Şen (1990a). In order to alleviate some drawbacks in RQD, Şen (1990a) has proposed the concept of rock quality percentage (RQP) and rock quality risk (RQR).

Although the quality of rock based on these measurements can be done by visual inspection, any quantitative method is always better in unifying different opinions about the same rock mass. Therefore, the method of rock quality designation (RQD) was adopted by Deere (1964, 1968) and is expressed as

$$\mathrm{RQD} = \frac{100}{L}\sum_{i=1}^{n} L_i \tag{7.26}$$

where the summation term includes all intact lengths L_i ($i = 1, 2, \ldots, n$) of sound rock between discontinuities that are 0.1 m (0.3 ft) or more apart, and L is the total length of the scanline considered. It is obvious from the defintion of RQD that its values range between 0 and 100. Furthermore, Deere classified any rock based on the RQD value as in Table 7.1.

Initial analytical expressions for RQD based on negative exponentially distributed intact lengths are presented by Priest and Hudson (1976) for very long scanline. The effects of finite length scanline on RQD value are described by Şen and Kazi (1984). Further analytical studies are presented by Şen (1984) for different intact probability distributions such as uniform normal lognormal and gamma PDFs. However, to the author's knowledge, there has appeared no work on RQD simulation based on the PDFs.

Table 7.1 RQD classification

RQD value	Classification
RQD <25	Very poor
25 < RQD <50	Poor
50 < RQD <75	Fair
75 < RQD <90	Good
90 < RQD <100	Excellent

Although relatively uncomplcatd and fast procedure, RQD has limited applications in the rock mass quality description, which is due to the following drawbacks:

1. It does not reveal anything about the intact length properties which are shorter than 0.1 m.
2. Two or more scanline measurements with intact lengths (i.e., equally spaced discontinuities) longer than 0.1 m would have all RQD of 100 %.
3. It provides directional rock qualities but fails to give a global quality description for the whole rock mass. For instance, RQDs of 67 and 92 percentages along the north-south and west-east directions, respectively, describe the same rock mass as fair and excellent quality. Hence, a dilemma arises as to the global rock quality description.
4. As already noticed by Şen (1990a), RQD does not provide answers to questions such as in what percentages does each rock quality (excellent good fair poor and very poor) occur within the same rock mass or what is the probability (risk) that the RQD value will be less than any given RQD value.

The only way to avoid these problems is to obtain the PDF of RQD itself on the basis of given intact length PDF. Such a task is very difficult by analytical means, and therefore Monte Carlo simulation technique is adopted in this paper for reaching the desired goal.

The value of RQD along any scanline is calculated from Eq. 7.26 for the intact lengths, which are longer than a certain threshold value t. In any rock mass, the intact lengths x are random variables, which have a certain PDF. In addition the number n of fractures is also an integer-valued random variable. Consequently, in a statistical context, RQD can be regarded as the random summation of random variables that are truncated at the threshold value. The PDF of intact lengths can be obtained empirically in the field from intact length measurements along very long scanline. However, determining the PDF of the RQD requires many scanlines with different directions and at different places which is quite tedious and not practical. Therefore, the question arises on how to determine the RQD distribution function from the basic PDF of the intact lengths. It has been already shown by Şen (1990a) that an analytical derivation of the RQD distribution function is almost impossible, and, therefore, the only way to obtain it is by numerical methods using Monte Carlo techniques.

Simulation of stochastic variables is rather similar to the numerical solution methods in mathematics. That is to say, provided that the underlying properties of a phenomenon are known, then simulation gives a way of reaching the desired goal numerically. The desired goal here is the RQD distribution. For such a simulation study, the following steps must be considered:

1. Determine the underlying PDFs of the intact lengths within a rock mass. In previous studies, this PDF has been taken to be either negative exponential or lognormal PDF, which have the mathematical forms:

$$f(x) = \lambda e^{-\lambda x} \qquad (0 \le x \le \infty) \tag{7.27}$$

and

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma_{Lnx}} \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_{Lnx}}\mathrm{Ln}\left(\frac{x}{\overline{m}_x}\right)\right]\right\} \qquad (0 < x < \infty), \tag{7.28}$$

respectively, where λ is the average number of discontinuities, $\overline{m}_x$ is the mean of the intact lengths, and σ_{Lnx} is the standard deviation of logarithmic intact lengths. Equation 7.27 has been used in most evaluations of RQD. The logarithmic distribution has not been used as often. The latter has more flexibility in representing the intact lengths because it takes into account the standard deviation of intact lengths independently of their mean value.

In the simulation study herein, first uniformly distributed random variables u with a range from zero to one are generated, and then they are transformed into negative exponentially distributed intact lengths with parameter λ by

$$s = -\frac{2.3}{\lambda}\log u = -\frac{1}{\lambda}\mathrm{Ln}\, u \tag{7.29}$$

However, for the logarithmic distributions, first the uniformly distributed pairs of variables (u_1 and u_2) are transformed into a normally distributed random variable pairs (s_1 and s_2) with a procedure already presented by Hammersely and Handscomb (1964) as

$$s_1 = \sqrt{-2\log u_1}\,\cos\,(2\pi u_2) \tag{7.30}$$

$$s_{21} = \sqrt{-2\log u_1}\,\mathrm{sin}\,(2\pi u_2) \tag{7.31}$$

Herein, u and u are uniformly distributed random variables within the range of 0–1. The values of s are then transformed into logarithmically distributed intact lengths x using.

$$x = e^{-x} \tag{7.32}$$

in which s represent s_1 or s_2. In this study, 1,000,000 identically distributed intact lengths were generated separately with negative exponential or lognormal PDFs.

2. These intact lengths are classified into two groups depending on whether they are greater or smaller than a threshold value t. In the RQD definition as given by Eq. 7.26, the summation of x's over the scanline length is calculated. The scanline length is in fact equal to the summation of the whole intact lengths provided that it starts and stops at discontinuity.
3. The RQDs for a set of predetermined numbers are identified.
4. The relative and cumulative frequency distribution functions for RQD values from (3) are calculated.

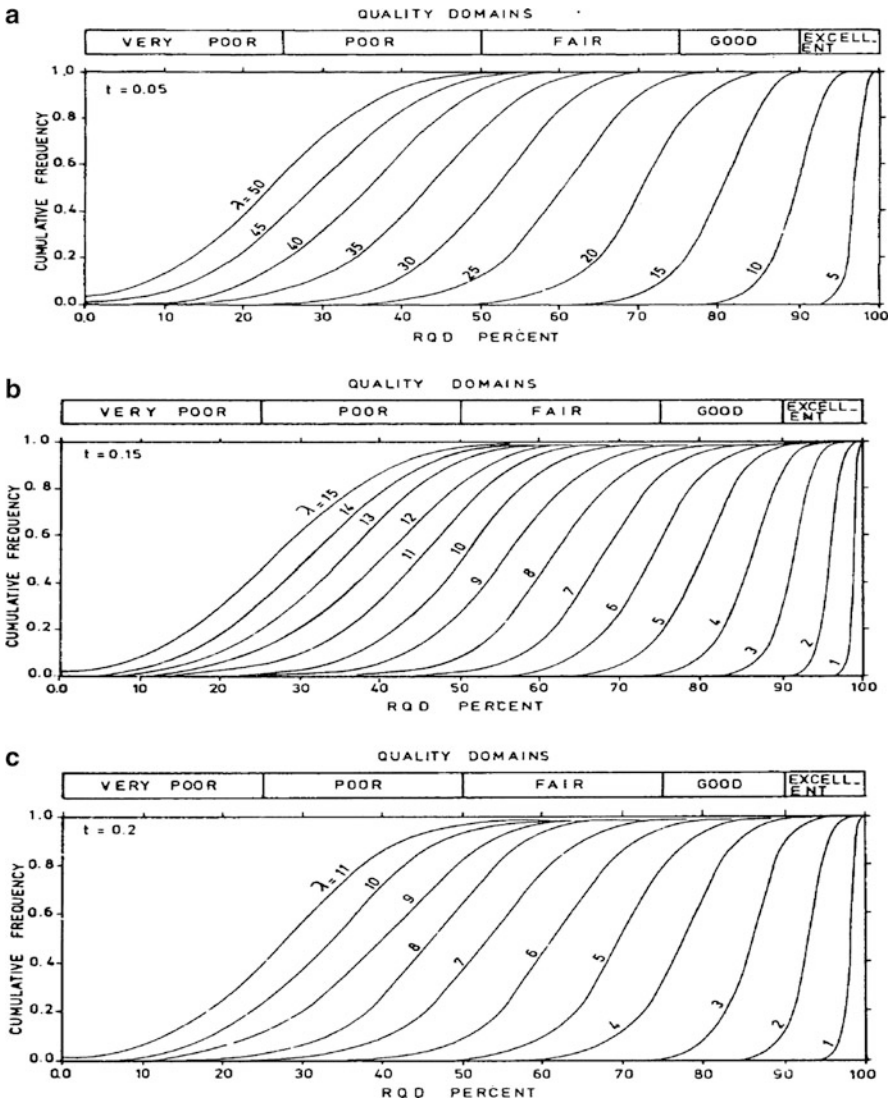


Fig. 7.7 RQD chart for negative exponential PDF with (**a**) $t = 0.05$; (**b**) $t = 0.15$; (**c**) $t = 0.20$

The cumulative PDF of RQD values resulting from negative exponentially distributed intact lengths with a set of average fracture number is presented in Fig. 7.7 for five different threshold values.

On the same figure, classification of RQD values is also shown. There is a similarity between the grain size distribution of granular rocks and these graphs which show the rock quality distribution of the fractured hard rocks. Field experiences show that any fractured rock is heterogeneous, and accordingly more than

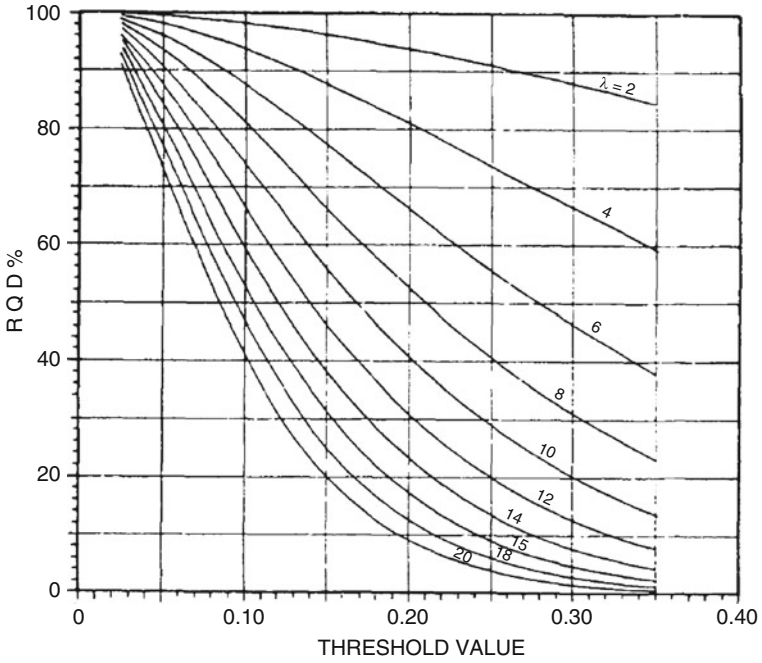


Fig. 7.8 RQD threshold value discontinuity number chart (negative exponential PDF)

one type of rock quality exists within the same rock mass. The percent ages of these different qualities can be found quantitatively from Fig. 7.7. Inspection of this figure leads to the following significant conclusions:

1. The smaller the average number of discontinuities, the smaller is the range which indicates the uniformity of the rock quality. For instance, whatever the threshold value, the rock has excellent quality provided that the average number of discontinuities is equal to unity.
2. As the average number of discontinuities increases, the rock becomes heterogeneous. For instance, in Fig. 7.7, when the threshold value is 0.05, the curve that represents 20 average intact lengths has three different qualities, namely, excellent, good, and fair portions. It is clear from the same curve that the majority of values are in the "good" quality zone, whereas other qualities are less likely to occur.
3. An increase in the average number of discontinuities leads to deteriorating rock qualities as shown in Fig. 7.8.
4. As the threshold value decreases, the rock quality increases (see Fig. 7.8). On the other hand, for a given threshold value, the deterioration rate in the rock quality is higher at small discontinuity numbers. For instance, at 0.20 truncation level, the reduction in the rock quality is almost 12 % between discontinuity numbers 2 and 4 whereas 3 % from 18 to 20.

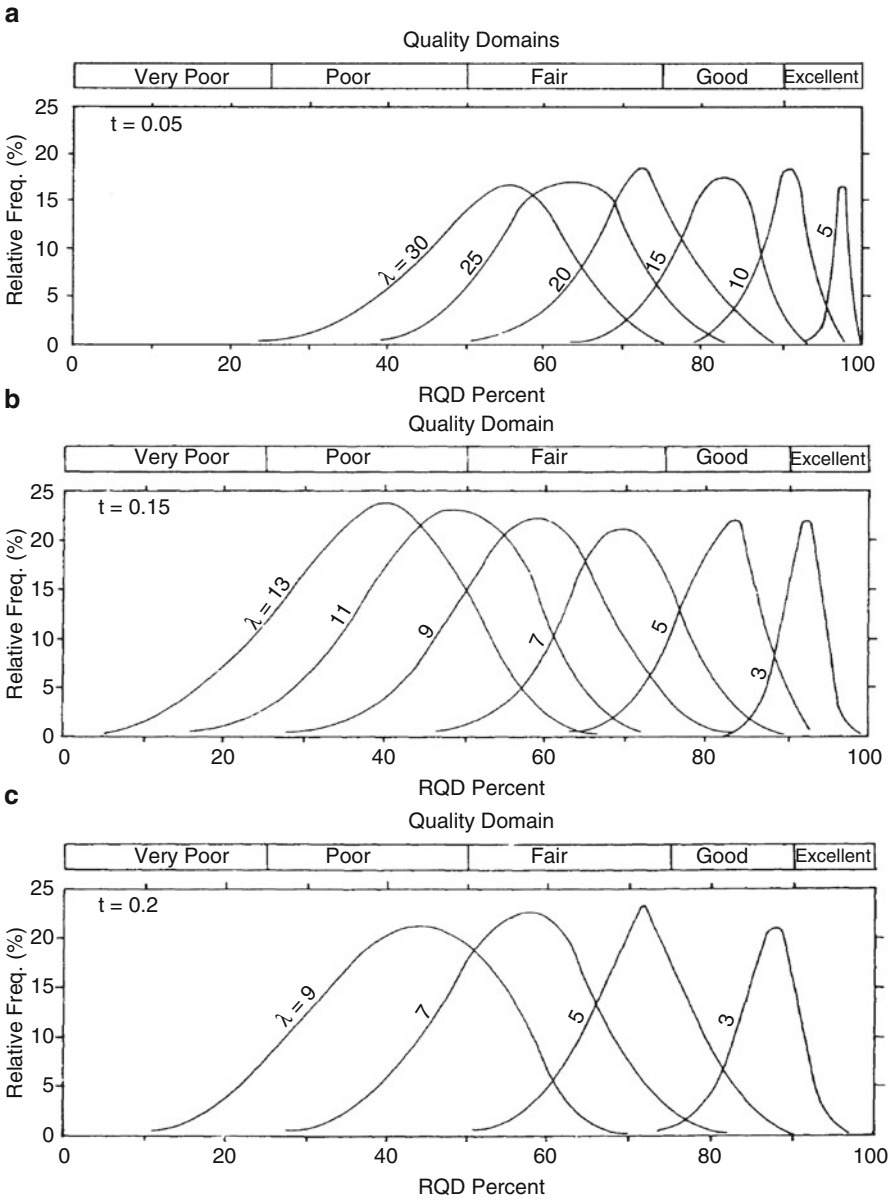


Fig. 7.9 RQD negative exponential PDFs with (**a**) $t = 0.05$; (**b**) $t = 0.15$; (**c**) $t = 0.20$

In additon, the PDF of the RQD for a set of discontnuity numbers at different threshold values appears in the negative exponential distribution forms as in Fig. 7.9.

One of the most striking properties of these PDFs is that irrespective of the discontinuity number and threshold value, they are invariably symmetric. The

positions of the maximum points on any one of these curves along the RQD axis shows the most likely rock quality within the rock mass.

Due to the aforementioned symmetry, the average rock quality description value coincides with the most likely rock quality value. This point indicates that the average RQD value is equal to the maximum likelihood estimation of the averages resulting from these PDFs. This value corresponds with the classical RQD value as defined by Deere. Besides the mean mode and the median values of the rock quality description are equal to each other. This last statement suggests that the PDF of RQD within a rock mass can be approximated by a normal distribution as

$$f(r) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2\right] \tag{7.33}$$

in which r is a dummy variable representing RQD and μ and σ are the population mean and standard deviation of RQD, respectively. For the negative exponential distribution, Şen and Kazi (1984) have shown already that the RQD has the same mean and standard deviation value which are expressible as

$$\mu = \sigma = 100(1 + \lambda t)\, e^{-\lambda t} \tag{7.34}$$

Finally, the cumulative PDFs of RQD for the underlying intact length distribution as the lognormal PDF are given in Fig. 7.10.

All of the conclusions for the negative exponential PDFs are equally valid for these curves. In addition, comparisons of various graphs in Fig. 7.10 indicate different standard deviations, and an increase in the intact length standard deviation improves the rock quality designation. In other words, the less uniform the fracture spacing, the stronger the rock mass. The following significant conclusions can be drawn from this study:

1. Any rock mass might have different rock qualities at the same time in different directions.
2. Rock quality deteriorates with the increase of the average number of discontinuities for any intact length distribution but an increase in the standard deviation of lognormal PDF of intact lengths leads to improvements in the rock quality designations.
3. The PDFs of RQD are unimodal in any case but symmetrical for the negative exponentially distributed intact lengths.
4. Any dominant type of rock quality has almost the same maximum frequency, confined within 0.20–0.25 for threshold levels more than or equal to 0.10 m.

7.3.2 Dependent Intact Lengths

All of the aforementioned studies have a common point in that they give an RQD estimation without consideration of the intact length correlation. However, it is a fact that even on the same outcrop of the rock, there might be correlated intact

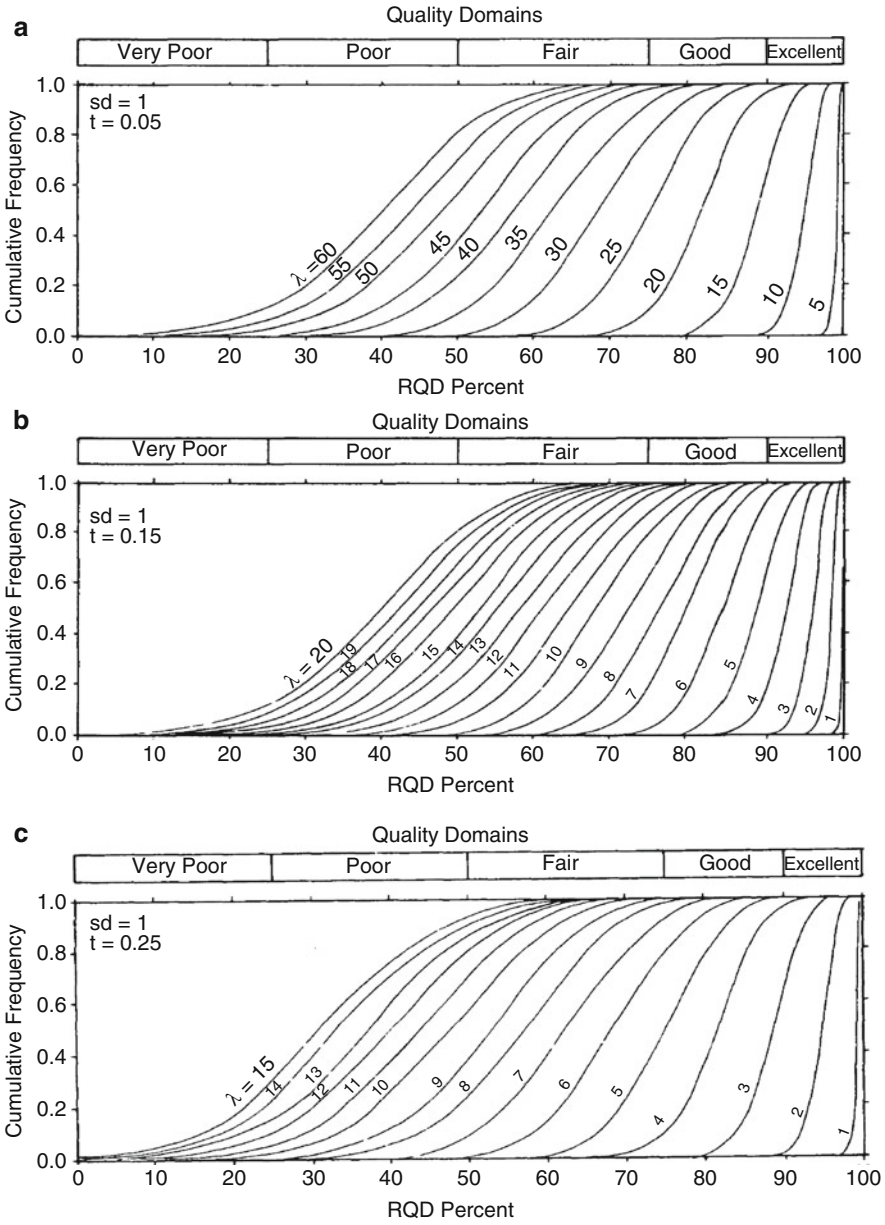


Fig. 7.10 RQD description cart for lognormal PDF (**a**) ($t = 0.05$, $m = 0.0$, $s = 1.0$); (**b**) ($t = 0.15$, $m = 0.0$, $s = 1.0$); (**c**) ($t = 0.25$, $m = 0.0$, $s = 1.0$)

lengths along scanline taken at various orientations (Eissa and Şen 1991). Although RQD calculations, according to Deere's (1964) definition directly from the scanline measurements, implicitly account for the intact length correlation, unfortunately, the analytical formulations do account for this correlation neither implicitly nor

explicitly. Accordingly, for correlated intact lengths, the existing analytical results are in error. It is obvious then that the comparison of analytical and empirical RQD estimations is possible accurately only for the cases of independent intact lengths. Otherwise, such a comparison is meaningless.

An important factor in the analysis of rock quality assessments from discontinuity measurements along a scanline is the correlation of the intact lengths. The autorun model and first-order autorun coefficient are proposed as a method of objectively quantifying intact length correlation structure and discontinuity occurrences within a rock mass (Şen 1978, 1984).

Any straight line through the rock mass encounters random number of discontinuities. An intact length is defined as the length of scanline or drill core between two successive discontinuities. In general, if there are $n+1$ discontinuity, the number of intact lengths is n provided that the start and end of the scanline are at discontinuities. A first step in rock mass classification is to consider two types of intact lengths, namely, those whose lengths are greater than a predesignated threshold value or otherwise (Eq. 7.26).

For the sake of convenience, alternative intact lengths will be grouped into two sets as elements a_i ($i = 1, 2, \ldots, k$) in set A and b_j ($j = 1, 2, \ldots, l$) in set B, where k and l are the number of intact lengths in each set. It is obvious that $k + l = n$ which is the total number of intact lengths. Furthermore, in an alternate sequence, either $k = l-1$ or $l = k-1$; however, practically one may assume with no loss of generality that $k = 1 = n/2$. In short, the intact lengths along a scanline will be an alternative combination of elements from two sets, namely, $A = \{a_1, a_2, \ldots, a_k\}$ and $B = \{b_1, b_2, \ldots, b_l\}$ as shown in Fig. 7.11a. In such a combination, the correlation structure

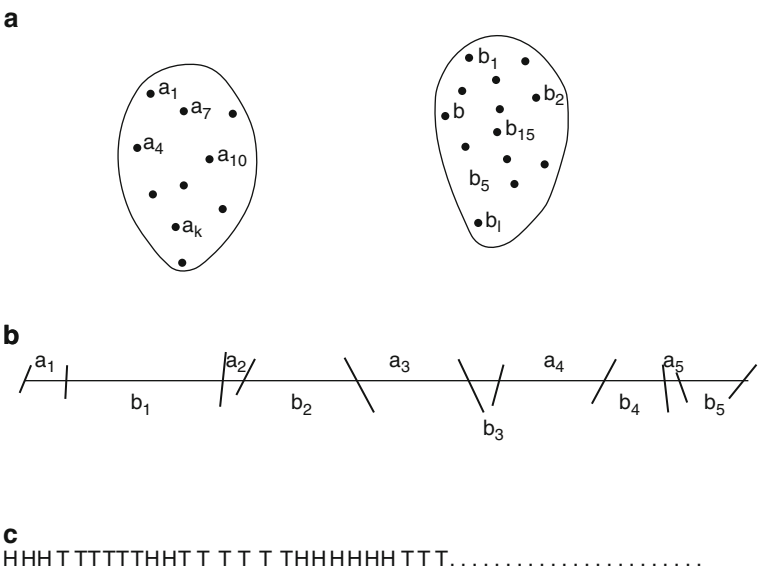


Fig. 7.11 Alternative intact length concepts

of sequence a_1, b_1, a_2, b_2, …. , $a_{n/2}$, $b_{n/2}$ (see Fig. 7.11b) is utmost importance in addition to various statistical descriptions of intact lengths.

Besides, it may well be that the intact lengths in set A have different PDF than B. However, this point lies outside the scope of this paper. Of course, assumption of uncorrelated intact lengths simplifies the analytical derivation of RQD but at a sacrifice of precision. Due to such an assumption, there is no term representing the correlation of intact lengths in any RQD formulations that are available so far in the literature. The major elements that effect the RQD calculations are in detail as follows:

1. The Number of Discontinuities: As mentioned earlier, practically, half of this number will be attached to intact lengths of set A and the second half to set B.
2. Intact Length PDF: It has been assumed in the majority of RQD studies as negative exponential type (Priest and Hudson 1976, 1981). However, the log-normal (Roulean and Gale 1985) or uniform and gamma PDFs (Şen 1984) are also employed to a certain extend.
3. Correlation Structure: Any significant correlation affects not only the discontinuity occurrences but also the intact lengths. This element is ignored invariably in any of the previous RQD studies in the literature. However, only some indirect procedures have been proposed for accounting the intact length correlation (Higgs 1984; Şen 1991).
4. Threshold Value: It is a fixed value as 0.1 m or 4 in. below which the intact lengths are not considered in the RQD calculations.

The first two elements are stochastic variables and in nature they are both serially and crossly correlated. Hence, the probabilistic laws of these stochastic parts lead to meaningful analytical expressions for RQD only after the consideration of correlation structure. For instance, logically any increase in the correlation will imply the occurrence of longer intact lengths along a scanline than the case where the intact lengths are independent. It also implies that the number of discontinuities decreases with increasing correlation. Hence, in general, the existence of relatively longer intact lengths (or lesser number of discontinuities) along a scanline implies improvement in its quality. Consequently, the key in the analytical RQD formulation for correlated intact lengths is the expression of correlation by an objective measure which is adopted herein conveniently as the autorun coefficient.

7.3.2.1 Correlation Measurement

The main question is whether the intact lengths along any direction are serially correlated or not. As was mentioned above, different authors have shown that the intact lengths are distributed according to various distributions among which the normal PDF has the least significant role. On the contrary, the classical correlation techniques (auto- and cross correlation) are valid for normally distributed variables only. Although by suitable transformations intact lengths may be rendered into normally distributed variables, but such transformations distorts the original

correlation function significantly (Şen 1978, 1979a, b). However, the autorun analysis is robust to any PDF or transformation and yields unbiased as well as consistent estimates of correlation. In general, lag-k autorun coefficient is defined by Şen (1978) as follows

$$r_k = \frac{2n_k}{n-k} \tag{7.35}$$

in which n_k is the number of overlapping successive the same type of events lag-k apart and n is the number of unit intact lengths. From the definition, it is obvious that $0< r_k <1$. In the case of purely independent observations, whatever the underlying PDF, Eq. 7.41 becomes equal to 0.5. Therefore, 0.5 shows the fact that the two observations separated by lag-k are independent from each other. On the other hand, if the observations are perfectly correlated, then $r_k = 1.0$.

The autorun coefficient application is very suitable for binary type of data; therefore, prior to its application, the variable concerned such as the intact length must be rendered into a binary form. For such a purpose the analogy suggested by Priest and Hudson (1976) as an unbiased coin tossing sequence of heads and tails will be adopted herein for alternating intact lengths, where a head represents a unit length of intact rock of type A and a tail represents a unit length of type B. With such an analogy, the scanline in Fig. 7.11b can be considered as a sequence of heads and tails (see Fig. 7.11c). The following significant points emerge in such an analogy:

1. The succession of uninterrupted sequence of heads (tails) represents intact lengths of type A (B).
2. Each appearance of alternate successive events, i.e., head-tail or tail-head succession, corresponds to a discontinuity. It is obvious that two successive head-head or tail-tail events represent two units from overall intact lengths. These explanations indicate the suitability of lag-one autorun coefficient, r, in quantifying the intact length correlation structure.
3. The percentages of heads (tails) along a scanline are equal to the probability of type A (B) intact length. Let these probabilities be denoted by p and q, respectively, then obviously $p + q = 1$. In terms of total length, $L_A(L_B)$ for set A (B) intact lengths, the probability can be expressed as $p = L_A/L(q = L_B/L)$.

Assuming uncorrelated intact lengths, Priest and Hudson (1976) have presented the analytical formulation of expected RQD as

$$E(\text{RQD}) = 100(1+\lambda t)\, e^{-\lambda t} \tag{7.36}$$

in which λ is the average number of discontinuities. The discontinuities are also assumed to have Poisson PDF with λ as its sole parameter. Eq. 7.36 should be used only after the confirmation that the intact lengths are independent.

For instance, a scanline composed of 18, 4, 5, and 3 cm intact lengths can be represented on the basis of 1 cm as an intact unit length by analogy to coin tossing as

HHHHHHHHHHHHHHHHHHTTTHHHHHTTT

In such a sequence, the probability of an intact length unit of type A (or H) is $p = 23/30 = 0.77$, and the probability of type B (or T) is $q = 7/30 = 0.23$, with $p + q = 1$. These probability values do not tell us whether the intact lengths are independent or not. However, the first autorun coefficient becomes $r = 19/29 = 0.66$, and since it is bigger than 0.5, there is a positive correlation between intact lengths. This implies, in general, that long intact lengths follow long intact lengths and short ones follow short intact lengths.

7.3.2.2 RQD Formulation and Discussion

Mathematical modeling of a scanline can be achieved by considering either the occurrence of discontinuity numbers or intact lengths, both of which are interdependent random variables. Modeling only one of them is sufficient because it implies the properties of the other. For instance, whatever the PDF, there is only one unique relationship between the expected intact length, $E(x)$, and the expected number of discontinuities, $E(n)$, along any scanline as

$$E(n) = \frac{L}{E(x)} \tag{7.37}$$

in which L is the scanline length. On the other hand, RQD as appears in Eq. 7.26 is equivalent to the summation of random number of random variables, and first by taking the expectations of both sides and then by considering Eq. 7.37, one can write

$$E(\text{RQD}) = \frac{100}{L} E(n) E(x^*) = 100 \frac{E(x^*)}{E(x)} \tag{7.38}$$

in which $E(x^*)$ is the expectation of intact lengths greater than a threshold value, t. Due to the fact that $E(x^*) < E(x)$, the ratio of expectations in the expression always assumes a value between 0 and 100. The expectations on the right-hand side of Eq. 7.38 can be found provided that the PDF of random variables concerned are known.

It can be shown similar to autorun modeling (Şen 1985) that the PDFs of k successive heads and tails are of geometric distribution types as

$$P(n_h = k) = (1 - r_1) r_1^{k-1} \tag{7.39}$$

and

$$P(n_t = k) = \frac{p}{q}(1 - r_1)\left[1 - \frac{p}{q}(1 - r_1)\right]^{k-1}, \tag{7.40}$$

respectively, where n_h is the number of uninterrupted successive heads and n_t successive tails. It is worthy to notice that for $r = p$ these equations give the independent intact length case as available in the literature (Priest and Hudson 1976). Furthermore, expectations of intact lengths in sets A and B can be obtained from Eqs. 7.39 and 7.40, respectively, as

$$E(n_h) = \frac{1}{1 - r_1} \tag{7.41}$$

and

$$E(n_t) = \frac{q}{p(1 - r_1)} \tag{7.42}$$

As mentioned above, the number of set A intact lengths is one less or more than set B intact lengths. In other words, practically they may be assumed as equals, and therefore, each type of intact length has its probability of occurrence equal to 0.5. With this information, the overall expectation of intact lengths, $E(x)$, without distinction between sets A and B can be seen to be $E(x) = 0.5E(n_h) + 0.5E(n_t)$ which yields by consideration of Eqs. 7.41 and 7.42 to

$$E(x) = \frac{1}{2p(1 - r_1)} \tag{7.43}$$

or from Eq. 7.37 one can find the expected number of discontinuities as

$$E(n) = 2Lp(1 - r_1) \tag{7.44}$$

In order to verify the validity of this expression, extensive computer simulations have been carried out through the Monte Carlo techniques, and the results are shown in Fig. 7.12. It is obvious that a very good agreement has been observed between the simulation and analytical formulation.

Finally, the average number of discontinuities, $\lambda_{r_1} = E(x)/L$, per unit length becomes

$$\lambda_{r_1} = 2p(1 - r_1) \tag{7.45}$$

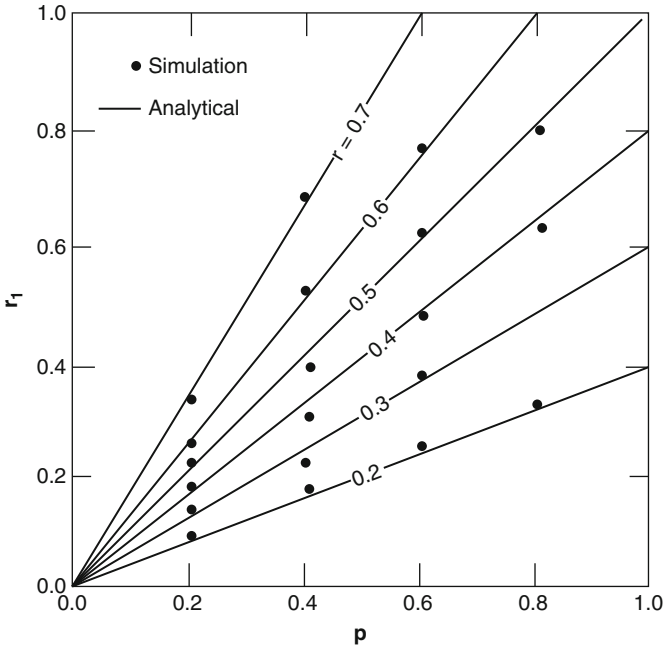


Fig. 7.12 Average number of discontinuity versus scanline length for different autorun coefficients

where the subscript signifies the correlatedness of the intact lengths. The probability, $P(k)$, of k discontinuity occurrences along a scanline of length x at an average rate of λ_{r_1} becomes according to the Poisson process as

$$P(k) = e^{-2xp(1-r_1)} \frac{[2xp(1-r_1)]^{k-1}}{k\acute{y}} \tag{7.46}$$

Since the interest lies in the discontinuity spacing PDF, by considering the distance, d, from one discontinuity to the following, one can write that $P(d < x) = 1 - P(k = 0)$, and, hence, substitution of Eq. 7.45 leads to cumulative PDF as

$$P(d \leq x) = 1 - e^{-2xp(1-r_1)}$$

By taking its derivative with respect to x, the PDF, $f(x)$, of intact lengths can be derived as

$$f(x) = 2p(1-r_1)e^{-2xp(1-r_1)}$$

Subsequently, the expectation of intact lengths that are more than t can be found according to

$$E(x^*) = \int_t^\infty x f(x) dx$$

which leads after the substitution of Eq. 7.46 to

$$E(x^*) = \frac{1 + 2p(1 - r_1)t}{2p(1 - r_1)} e^{-2p(1-r_1)t} \tag{7.47}$$

Finally, the substitution of Eqs. 7.43 and 7.47 into Eq. 7.38 leads to

$$E(\text{RQD}) = 100\,[1 + 2p(1 - r_1)t] e^{-2p(1-r_1)t} \tag{7.48}$$

For independent intact lengths, $r_1 = 0.5$, and the occurrences of unit intact lengths comply with the binomial distribution which leads to geometric intact length distribution with $E(x) = 1/p$ or $\lambda = p = 0.5$, and, therefore, Eq. 7.48 becomes identical to what was suggested by Priest and Hudson (1976) as in Eq. 7.41. Under the light of the aforementioned discussions, one can rewrite Eq. 7.48 as

$$E(\text{RQD}) = 100\,[1 + \lambda(1 - r_1)t] e^{-2\lambda(1-r_1)t} \tag{7.49}$$

The validity of this formula is checked with extensive Monte Carlo simulation technique by using autorun model for generating correlated intact lengths as proposed by Şen (1985). First of all estimates of average intact lengths of sets A and B are calculated as

$$n_A = \frac{1}{m_A} \sum_{i=1}^{k} (n_A)_i \tag{7.50}$$

and

$$n_B = \frac{1}{m_B} \sum_{i=1}^{l} (n_B)_i, \tag{7.51}$$

respectively. Herein, m_A and m_B are the number of intact lengths; $(n_A)_i$ and $(n_B)_i$ are ith intact length in sets A and B, respectively. The geometric PDF parameter which is the first-order autorun coefficient can be estimated from Eq. 6.41 as

$$r_A = (n_A - 1)/n_A \tag{7.52}$$

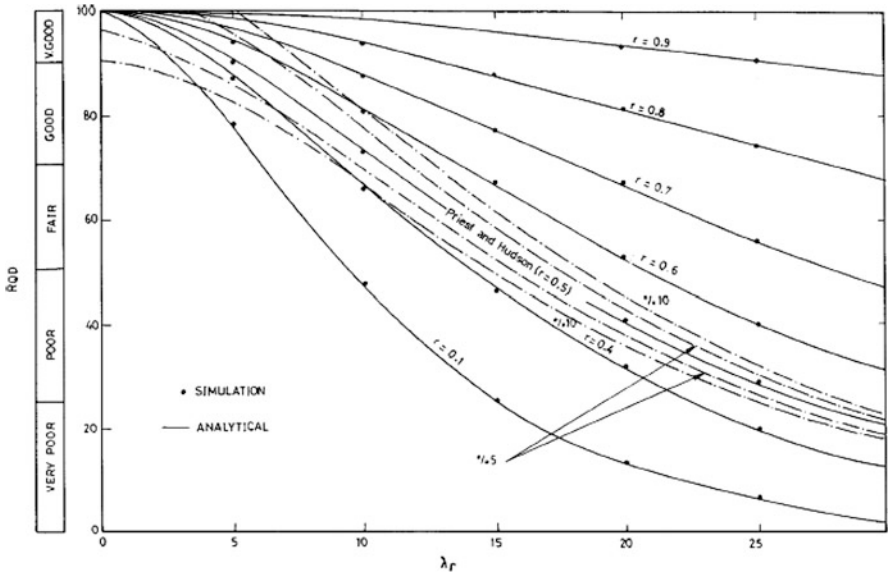


Fig. 7.13 RQD – average number of discontinuity – autorun coefficient chart

Similarly, the geometric PDF parameter, r_B, for set B intact lengths turns out to be

$$r_B = (n_B - 1)/n_B \tag{7.53}$$

Hence, it is possible to generate geometrically distributed and integer-valued alternate lengths, y, with parameters r_A and r_B, respectively, on a digital computer through

$$y = 1 + \log(\varepsilon)/\log(r_1) \tag{7.54}$$

in which $\boldsymbol{\varepsilon}$ is the uniformly distributed random variable between 0 and 1 and r_1 is the geometric distribution parameter which assumes either the value of r_A or r_B as required. The simulation results are presented in Fig. 7.13 together with the analytical solutions, and they show a very good agreement.

The following significant points can be drawn from this figure:

1. The formulation provided by Priest and Hudson (1976) for E(RQD) yields underestimated results if the intact lengths are positively correlated which is the case in most of the natural rocks as will be presented in the application section of this section.
2. Increase in the correlation structure gives rise to increase in the E(RQD) values.
3. Relatively better RQD values are obtained for the same number of discontinuities but correlated intact lengths.

4. The difference between the dependent and independent intact length RQD values is relatively less significant at small λ values than big λ's. In fact, at 5 % relative error level, the correlated intact lengths do not lead to significantly different RQD values provided that $\lambda < 3$ and 10 % error level when $\lambda < 10$. In Fig. 6.3 upper and lower confidence limits at 5 % and 10 % significance levels are shown around the Priest and Hudson independent intact length solution. It is obvious that for small average number of discontinuities, their solution gives confident RQD estimates even though the intact lengths are dependent. However, for big average number of discontinuities, the significance of intact length correlation becomes very pronounced in the RQD estimations.

7.3.2.3 Applications

The applications of the methodology developed herein are carried out for the field data from various parts of the world. The first field data are recorded along the exposed outcrop surfaces of granitic rocks in the western part of the Kingdom of Saudi Arabia. Extensive geological field survey by Otaibi (1990) showed that the area consists of one rock unit which is granite of light pink color on fresh surface and dark brown on weathered surface, medium to coarse grained, and equigranular. This area was selected since it has a good combination of well-exposed bedrock and relatively simple fracture pattern. Three sets of fracture orientations can be seen distinctively in this area. Each one of the fracture set is measured by the scanline perpendicular to the fracture traces. The fracture measurements are carried out at three sites. These sites are selected such that they give rather random characterization of the fracture geometry, i.e., they are quite independent from each other. In order to be able to apply the methodology developed herein, the relevant values are calculated and presented in Table 7.2.

It is obvious that the use of independent intact length RQD formulation does not yield significant deviations from the dependent intact length case. This is due to two major reasons. First of all since the average number of discontinuities is all less than 3 and, therefore, as already explained in the previous section, even if the intact

Table 7.2 Intact length characteristics (Saudi Arabia)

Site	Scanline	λ	$(1/m)$	p	q	r_1 Eq. 6.3	E(RQD) Eq. 6.16	Error (%)
1	x	1.56	0.52	0.48	0.51	98	98	0.0
	y	2.56	0.54	0.46	0.51	97	97	0.0
	z	1.04	0.62	0.38	0.63	99	99	0.0
	x	2.75	0.53	0.47	0.51	97	97	0.0
2	y	1.83	0.59	0.41	0.57	98	99	1.0
	z	1.98	0.63	0.37	0.61	98	99	1.0
	x	1.78	0.48	0.52	0.47	98	98	0.0
3	y	1.80	0.54	0.46	0.52	98	99	1.0
	z	1.56	0.57	0.63	0.55	99	99	0.0

Table 7.3 Intact length characteristics (England)

Scanline	λ	(1/*m*)	*p*	*q*	r_1 Eq. 6.3	*E*(RQD) Eq. 6.16	Error (%)
SL1	6.46	0.46	0.54	0.40	86	81	6
SL2	10.01	0.57	0.43	0.48	74	72	3
SL3	6.52	0.50	0.50	0.43	86	83	4

lengths are strongly correlated, there will not be practically significant difference, i.e., the relative error will be less than 5 %. In addition to this major reason, the autorun coefficients are rather close to 0.5 which also confirms the approximation in the results of *E*(RQD) calculations either by the use of Eqs. 7.36 or 7.47.

The second set of data for the implementation of the methodology are extracted from a previous study by Ryckes (1984) on modes of failure and stability of rock slopes in Tytherington Quarry which is located halfway between Thornbury and Tytherington for about 16 km north of Bristol in the county of Avon, England. The Paleozoic rocks of the lower coal series in the area are affected by many movements in the past which led to faults, folds, and unconformities each of which led to different patterns of intact lengths. Due to these different mechanisms, it is not possible to expect that these intact lengths have independent correlation structure. In order to depict the regional discontinuity pattern in this area, three scanlines were set up at different directions. The bedding plane orientation is almost horizontal for the first scanline which will be referred to SL1. The second scanline SL2 has a direction of 20° toward the southwest, whereas the third, SL3, has an intermediate inclination to the former. The necessary parameters as well as the *E*(RQD) calculations are presented in Table 7.3.

The major difference between Tables 7.2 and 7.3 is that the average discontinuity numbers of scanlines in Table 7.2 are far bigger than Saudi Arabia measurements; however, the autorun coefficients in Table 7.3 are invariably less than 0.5 indicating that there are negatively correlated intact lengths. Consequently, Priest and Hudson (1981, Eq. 6.3), gives overestimation.

7.4 RQD and Correlated Intact Length Simulation

Generally in open rock mechanics, literature persistence implies the areal extent or size of a discontinuity within a plane (Brown 1981). It can be quantified crudely by observing the discontinuity trace lengths on the surface of exposures. However, the type of persistence herein is related to the sequential occurrences of intact lengths which may constitute obvious clusters. Within the contest of this paper, persistence can be defined as the tendency of short intact lengths to follow short intact lengths and long intact lengths to follow long intact lengths. Priest and Hudson (1976) say

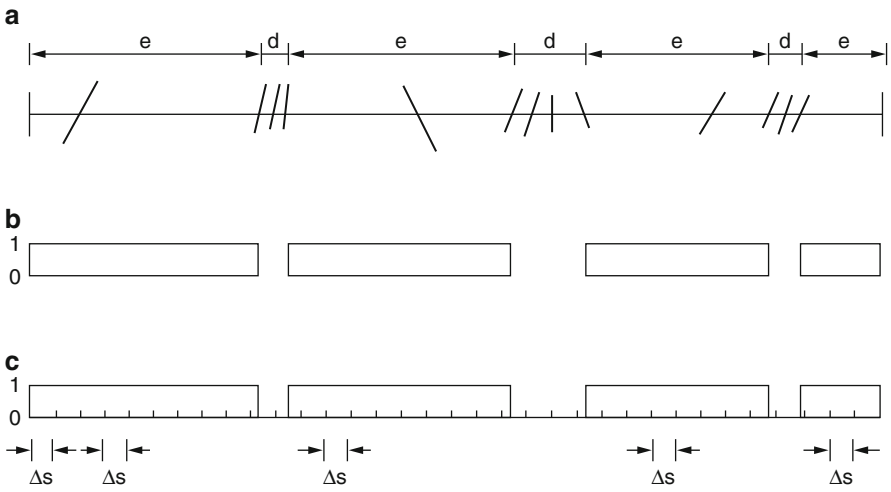


Fig. 7.14 Scanline and models (**a**) effective, (**b**) defective, (**c**) model interval

that a high frequency of low spacing values occurs within clusters and a low frequency of high spacing values occurs between clusters.

The simple way of expressing persistence in a sequence of observations is through either the classical serial correlation coefficient or autorun function Şen (1978). Although the former requires the measurements to be normally distributed, the latter is robust and applicable in any distribution case. However, for normally distributed intact lengths, both give exactly the same result. In fact in such a situation, the autorun coefficient r is convertible to the correlation coefficient ρ by $\rho = \sin\pi(r - 0.5)$. As a consequence only for the normally distributed intact lengths one can use interchangeably the autorun and autocorrelation terminologies.

In the following sequel, the intact length persistence is quantified with the help of autorun coefficient r_k for lag-k defined as in Eq. 7.35. In order to apply this formulation to a scanline, the following steps must be executed on any scanline as presented in Fig. 7.14:

1. By considering the threshold value equal to 0.1 m (or 4 in) as proposed by Deere (1964), the effective intact lengths are assigned uniform value of +1 and defective lengths as 0 (see Fig. 7.14b). In fact the resulting modified scanline has zone of square waves which are separated by zero-valued intervals.
2. Divide the scanline into fixed length of finite intervals Δs as shown in Fig. 7.14c where the intervals of fixed lengths are adopted as +1 cm which is convenient for any practical purpose. The number of such intervals along the scanline is denoted by n.

   The ratio of numbers n within the effective intervals along the scanline to the total interval number n is equivalent to the estimate of RQD, i.e., RQD = 100 (n_1/n).

3. Find the number n of overlapping pairs of 1 and estimate r from Eq. 3.36 for different lags from a sequence of $r_1, r_2, r_3, r_4 \ldots$.

The plot of autorun coefficient versus lag value gives rise to graphs that represent the persistence existing in the original intact sequence. A very significant result at this point is that if practically all of the autorun are equal to zero, then the intact lengths originate from an independent random process, and hence the calculation of RQD as defined by Deere is reliable; otherwise, persistence appears implying clusters in intact lengths and unreliability in RQD calculations.

The applications of the abovementioned persistence procedure to actual data are achieved with considering field measurements from South England near Bristol (Ryckes 1984) and measurements from the western province of Saudi Arabia collected by authors.

The Paleozoic rocks of the lower coal series in the area are affected by various earth movements in the past which led to faults, folds, and unconformities. The location of intact length measurement sites are on the axis of a syncline which inclines toward the south. The bedding plane orientation is almost horizontal for the first scanline (SL1). The second scanline (SL2) has a direction of 20° toward southwest, and the third scanline (SL3) has an intermediate inclination to the former. Further detailed information on these scanline measurements can be found in Ryckes (1984). The fourth scanline (SL4) was an illustrative example adopted from Brady and Brown (1985). However, the fifth scanline (SL5) is a representative scanline measurement in crystalline dioritic rocks in the southwest of Saudi Arabia.

Three scanline measurements from England lead to the autorun functions as shown in Fig. 7.15, and for the illustrative example and the Saudi Arabia scanline measurements, autorun functions are given in Fig. 7.15d, e.

A common property in all these figures is that the first autorun coefficient has the greatest value among other lags, and there appears an exponential or power type of decrease with increasing lag value. It is clear that the classical RQD calculations without considering persistence, i.e., assuming that $r_k = 0$ for all lag-k, are erroneous and should be rectified accordingly. Logically the same RQD value may correspond to different persistence levels or vice versa.

The RQD and persistence values for different scanlines considered herein are shown in Table 7.4.

A striking conflict appears from this table so far, and the rock qualities on RQD and persistence basis are concerned. For the fourth scanline, the same quality conclusion is arrived but other scanlines are in conflict to some extent. Hence, a dilemma arises as to which one is chosen for decision making. The view taken in this research is that persistence-based qualities should be preferred over the RQD-based descriptions since the former takes into consideration not only the mean and variance of intact lengths but also their correlation structure, i.e., clustering effects.

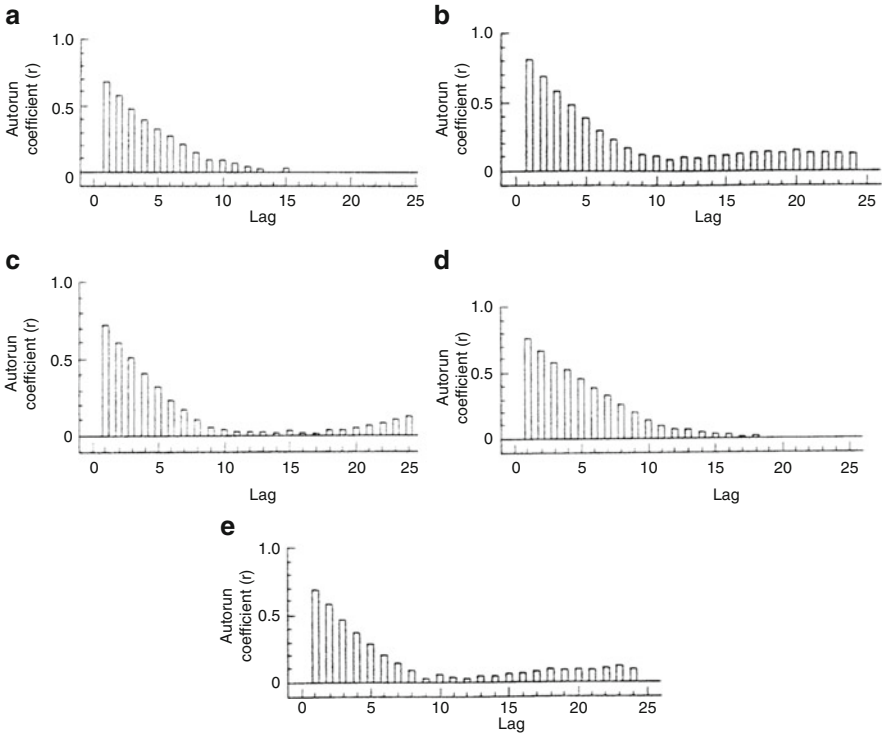


Fig. 7.15 Autorun function for (**a**) SL1, (**b**) SL2, (**c**) SL3, (**d**) SL4, (**e**) SL5

Table 7.4 Scanline rock qualities

Scanline	RQD (%)	Quality	Persistence	Quality
SL1	88	Very good	0.68	Fair
SL2	74	Fair	0.82	Very good
SL3	86	Very good	0.73	Fair
SL4	89	Very good	0.78	Very good
SL5	86	Very good	0.70	Fair

7.4.1 Proposed Models of Persistence

Provided that the autorun functions are available, the question arises as to what type of dependent process represents best the intact length occurrences within a rock mass. A straightforward answer to this question can be given provided that the autocorrelation structures of different theoretical stochastic processes are known beforehand. For this purpose three stochastic processes will be described herein, namely, the independent random process, the lag-one Markov process, and the autoregressive integrated moving-average process (ARIMA) (1, 1).

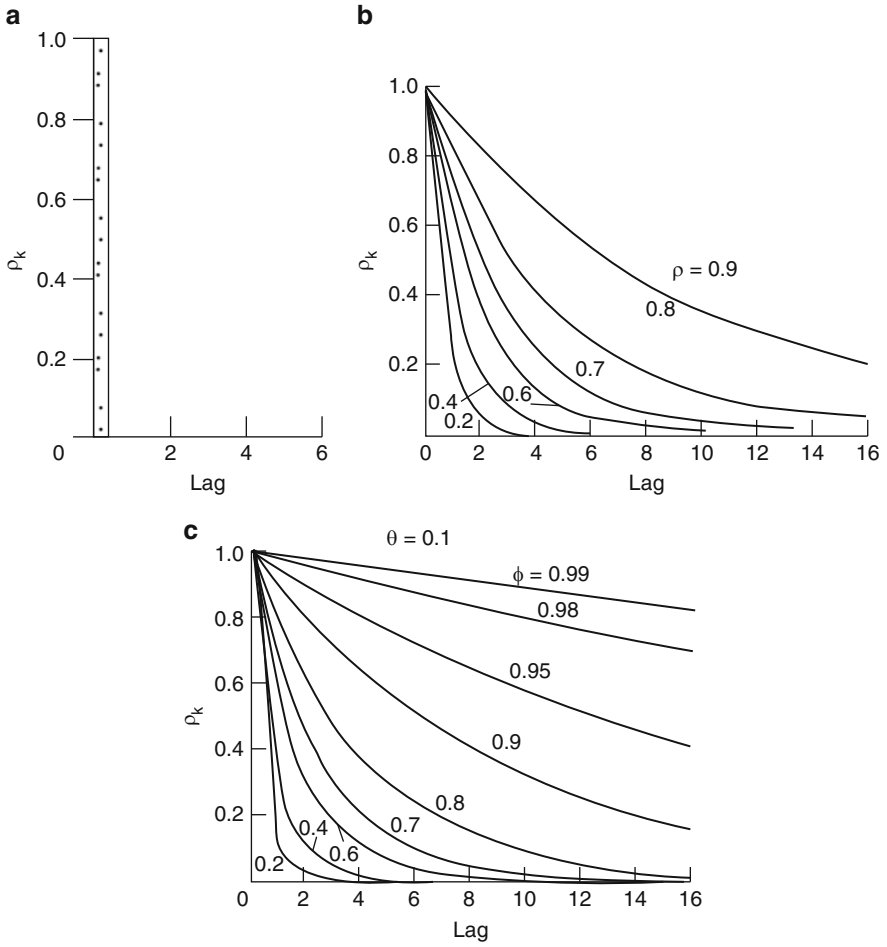


Fig. 7.16 Autocorrelation structure of different processes (**a**) independent process, (**b**) Markov process, and (**c**) ARIMA (1, 1) process

The Independent Process

It does not have any autorun function, i.e., all of the autocorrelation coefficients at each lag except lag-zero are equal to zero as shown in Fig. 7.16a. It is the easiest and simplest stochastic process employed in representing a natural phenomenon. Two parameters, namely, the mean and standard deviation, are enough to describe the phenomenon completely. In fact all of the RQD, rock quality risk (RQR), rock quality percentage (RQP), etc. have been based on the understanding that the phenomenon has an independent structure (Şen 1990a). Since there have been so many studies in the past concerning this process, its repetition is avoided herein.

7.4.1.1 First-Order Markov Process

In general, geologic variables possess dependent structure especially when they are at close intervals from each other. The simplest way of representing any dependence is through autoregressive processes such as the lag-one Markov process (Şen 1974). The mathematical form of this process can be written as

$$l_i = \mu + (l_{i-1} - \mu) + \sigma\sqrt{1-\rho^2}\varepsilon_i \tag{7.55}$$

in which μ, σ, and ρ are the mean standard deviation and the first-order serial correlation coefficient, respectively, of intact lengths l_i and finally ε_i is a normal random shock with zero mean and unit variance. The autocorrelation structure of this process is given as

$$\begin{aligned} \rho_0 &= 1 \\ \rho_i &= \rho^i \end{aligned} \tag{7.56}$$

which has been presented for different ρ values in Fig. 7.16b.

7.4.1.2 ARIMA (1, 1) Process

It is used in order to represent more persistently dependent processes which are noticed during intact length analysis in this study. These processes were prepared first by Box and Jenkins (1970) with the following expression:

$$l_i = \mu + \phi(l_{i-1} - \mu) + \varepsilon_i - \theta\varepsilon_{i-1} \tag{7.57}$$

in which ϕ and θ are model parameters and ε_i is again normal random process with zero mean and unit variance. The autocorrelation structure of this model is given as

$$\begin{aligned} \rho_0 &= 1 \\ \rho_1 &= \frac{(1-\phi\theta)(\phi-\theta)}{\left(1+\theta^2-1\phi\theta\right)} \\ \rho_k &= \phi\rho^{k-1} \quad (\text{for } k \geq 2) \end{aligned} \tag{7.58}$$

which is drawn for the same set of parameters in Fig. 7.16c. Comparisons of graphs in Fig. 7.16 indicate that only ARIMA (1, 1) processes lead to significantly more persistent correlations at large lags than other processes.

7.4.2 Simulation of Intact Lengths

The purpose of this section is to derive the relationship between RQD and the first-order autorun correlation coefficient as a representative of persistence. As already mentioned above, since the autorun functions in Fig. 7.15 indicate persistence structure in the intact length occurrences, they cannot be represented by classical processes such as the negative exponential distribution which is based on the assumption that the intact lengths are generated from an independent process. This implies autocorrelation structure as in Fig. 7.16a. It is obvious from Fig. 7.15 that all of the scanlines considered in this study have autorun functions with significant persistences. In order to be able to simulate these persistent patterns, the intact lengths are considered to be normally distributed so that the autorun functions can be considered as equivalents to the autocorrelation functions which are necessary ingredients for simulation by use of the Markov and ARIMA (1, 1) processes.

Hence the first step in the simulation is to find the best stochastic model that fits the experimental autocorrelation functions. Therefore, the experimental autocorrelation functions that are already given in Fig. 7.15 are compared with a set of theoretical autocorrelation structures as obtained from Eqs. 7.50 and 7.57 for Markov and ARIMA (1, 1) models, respectively, by considering different parameter sets. In fact the identification of model parameters such as ρ, ϕ, and θ is made by visual comparison of experimental and theoretical autocorrelation functions as presented in Fig. 7.17.

It is obvious that for the Markov process, there appears to be only one theoretical autocorrelation function based on the first-order autocorrelation coefficient, whereas ARIMA (1, 1) process has different functions for the same correlation coefficient depending on the values of ϕ and θ. In deciding on the best fit, only the first one third of the experimental autocorrelation must be considered as reliable. On this basis, it appeared that only SL2 had a very good match with the Markov process and the others accord by the ARIMA (1, 1) process but with different set of parameters. Table 7.5 exhibits the model types and parameter values for each scanline.

It is obvious from this table that the least persistent scanline on the basis of autocorrelation function is SL2. However, positive $\phi - 1$ values for the other scanlines indicate that the rock quality is comparatively better than SL1.

After the availability of the model parameters from Table 7.5, it is possible to generate synthetic scanlines by using the appropriate model on digital computers. The benefit from such a simulation is to gain detailed insight into the possibilities of the fracture distribution within the rock mass. In fact simulation results in as many scanlines as required which are statistically indistinguishable from each other.

This is tantamount to saying that along each synthetic scanline, the occurrence of intact lengths will be definitely different but in the long run they will have the same statistical parameters such as the mean and variance as well as the same

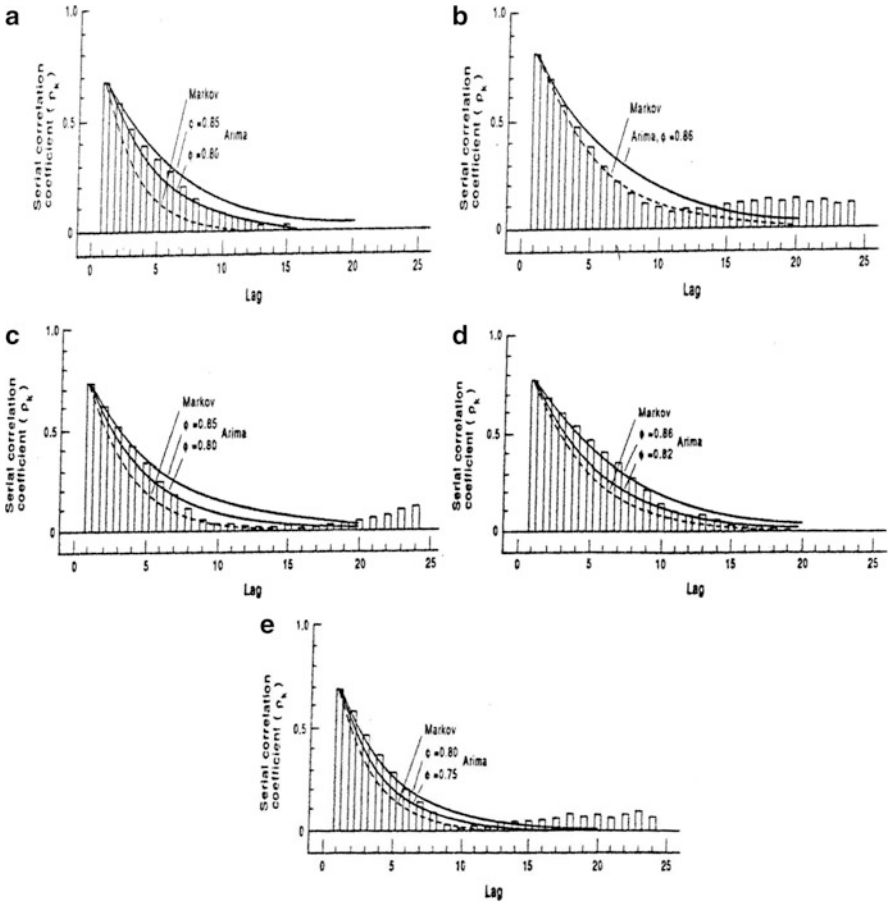


Fig. 7.17 Experimental and theoretical autocorrelation function for (**a**) SL1, (**b**) SL2, (**c**) SL3, (**d**) SL4, (**e**) SL5

Table 7.5 Models and parameters

Scanline no.	Model	Parameters				
		ρ	ϕ	θ	μ	σ
SL1	ARIMA (1, 1)	0.68	0.85	0.333	13.77	14.78
SL2	Markov	0.82	–	–	9.98	10.13
SL3	ARIMA (1, 1)	0.73	0.80	0.152	15.33	14.05
SL4	ARIMA (1, 1)	0.78	0.82	0.103	19.78	17.34
SL5	ARIMA (1, 1)	0.70	0.80	0.200	14.38	9.49

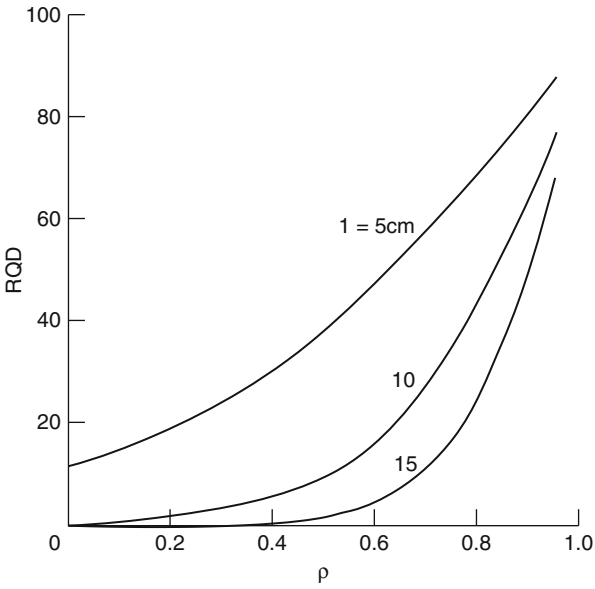


Fig. 7.18 RQD-persistence relation (Markov process)

autocorrelation structure. In order to explore the relationship between the persistence and RQD, the following set of different simulations are undertaken:

1. Irrespective of the scanlines considered in this paper, first of all the relationship between the RQD and persistence parameters (ρ, ϕ, and θ) is sought for standard intact lengths, i.e., the mean and variance of intact lengths are assumed to be zero and one, respectively. Hence, the generated synthetic scanlines are not affected by any other parameter except the persistence. However, these relationships are obtained for different threshold values as 0.05, 0.10, and 0.15 m. The resulting standard curves for the Markov process are presented in Fig. 7.18.
2. It is obvious that in the Markov process case, the RQD is proportionally related to the persistence (in this case ρ) but the relationship is nonlinear. RQD is not sensitive to changes in ρ for small ρ values for instant in the case of threshold 0.10 m when $\rho < 0.5$. However, for $\rho > 0.5$ the sensitivity increases enormously, and therefore the classical RQD calculations may not be reliable especially when $0.5 < \rho < 1.0$ provided that the underlying generating mechanism of intact lengths is of Markovian type. On the other hand, Fig. 6.9 also indicates that the rock quality improves if a low threshold value such as 0.05 m is adopted for the basic RQD definition. On the other hand, the comparison of three curves in the same figure shows that the rock quality improvement from threshold value 0.15 to 0.10 m is less than the transition from 0.10 to 0.05 m.

The same type of simulation but for the ARIMA (1, 1) model leads to standard RQD-persistence relationship as in Figs. 7.19 and 7.20 which present only two samples from an infinite numbers of such relationships.

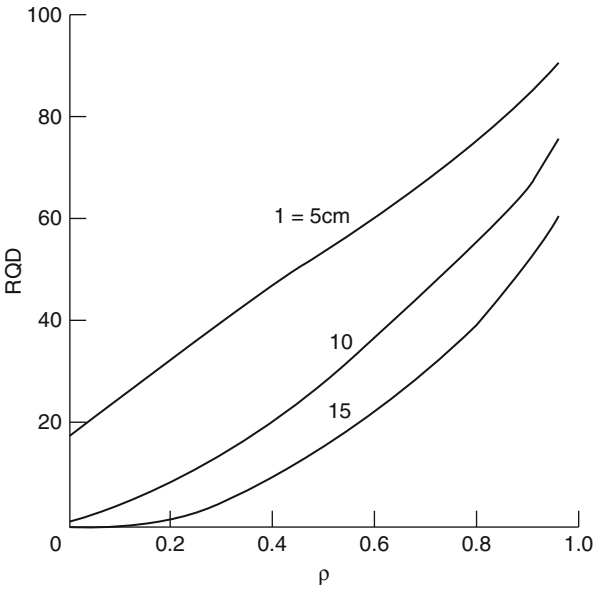


Fig. 7.19 RQD-persistence relation (ARIMA process $\phi = 0.9$)

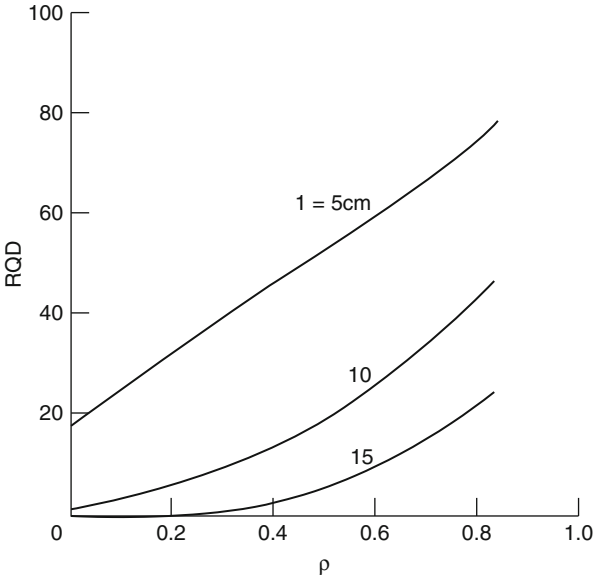


Fig. 7.20 RQD-persistence relation (ARIMA process $\phi = 0.7$)

The first comparison of Figs. 7.19 and 7.20 with Fig. 7.18 shows that the ARIMA (1, 1) process implies better rock quality for the same threshold value. This is indeed the logically expected result from the previous discussions. On the other hand, the sensitivity of RQD to persistence parameter changes decreased significantly for $\rho > 0.5$ but increased for $\rho < 0.5$. In other words, the curvature of the

curves is very small compared with the Markov process to the extent that even in the same cases such threshold values of 0.05 m, the RQD-persistence relationship appears as a straight line. Furthermore, comparison of Fig. 7.19 with Fig. 7.20 indicates that increase in the ϕ value means improvement in the rock quality which is due to the fact that large ϕ values imply longer intact lengths.

2. In order to see the effects of genuine parameters, i.e., mean and variance of the scanlines, a second set of simulation with the same models is performed on digital computers. Some of the representative results for SL1, SL2, and SL3 are presented in Fig. 7.21. A general conclusion from these figures is that the mean and standard deviation of the intact lengths are not sufficient in calculating RQD values. For instance, consideration of Fig. 7.21 leads to the conclusion that the least RQD value as almost 20 will appear provided that the intact lengths along SL1 occur independently ($\rho = 0$). However, in addition to these two basic

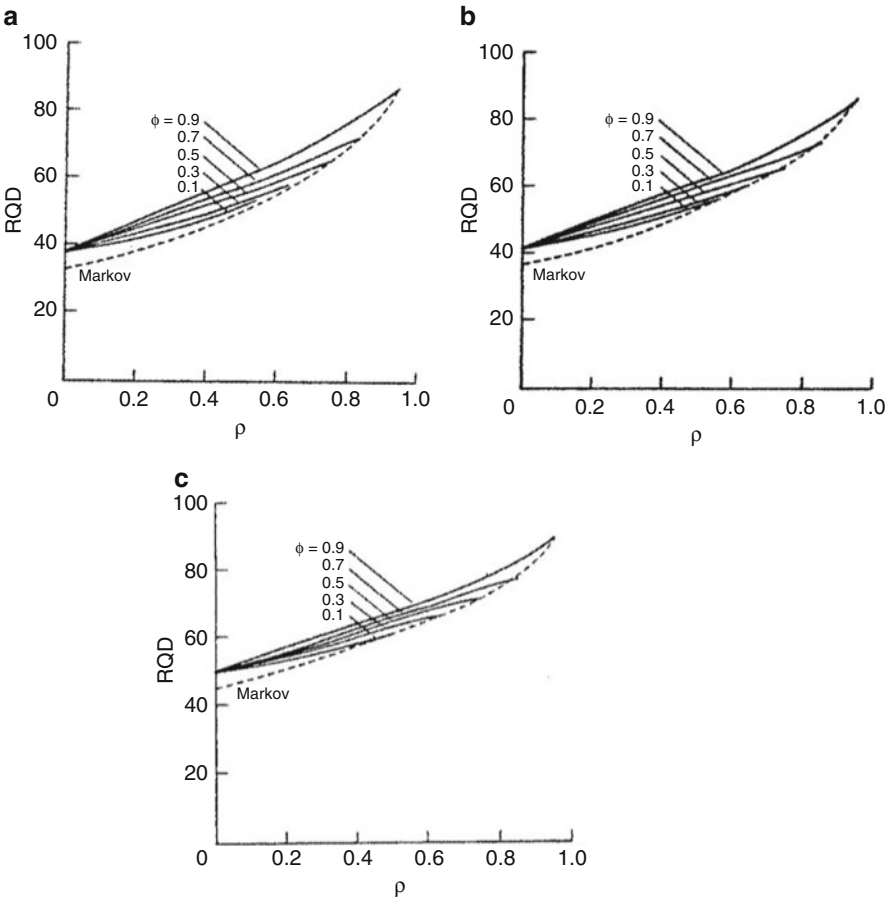


Fig. 7.21 RQD-persistence relation for (**a**) SL1, (**b**) SL2, (**c**) SL3

parameters, the value of ϕ plays a dominant role by causing increment in the RQD value. Similar statements are valid for other scanlines. Besides comparison of Fig. 7.21 with Figs. 7.19 and 7.20 shows that as the mean and standard deviation values increase, the sensitivity of persistence parameter decreases. It is logical to conclude that the longer and less variable are the intact lengths, the more reliable is the classical RQD calculation according to Deere (1964) original definition.

The main theme of this study was to emphasize that in practical studies the RQD calculations are not reliable without the consideration of intact length correlation structure. It is a well-known fact that the classical RQD evaluations are based on the mean value percentage of the intact lengths that are greater than a threshold value only. However, under the light of the simulation study performed in this paper, the following important points are worth to notice:

1. The persistence structure of intact lengths gives rise to additional rock quality increments.
2. There exists a proportional relationship between the RQD and the persistence parameter which is adopted as the first-order autocorrelation coefficient in this study.
3. It is possible to identify the underlying generating mechanism of the intact lengths by comparing the experimental and theoretical autocorrelation functions of stochastic processes such as the Markov or ARIMA (1, 1) processes.

7.5 Autorun Simulation of Porous Material

Any porous medium will have either a solid or a void at each point in space. Quantitatively solid point is represented by -1 and a void point by +1. Figure 7.22 indicates voids as white squares and solids as black patches, and such a spatial distribution of digitized numbers is a bivariate random field or ReV.

Hence, ReV function, $\xi(x)$, is a two-valued random variable in space as (Şen 1990b)

$$\xi(x) = \begin{cases} +1 & \text{if} \quad x \in A \\ -1 & \text{if} \quad x \in A^c \end{cases} \tag{7.59}$$

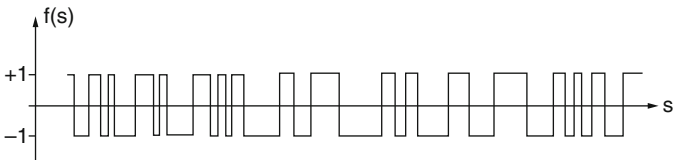


Fig. 7.22 Schematic representation of a line characteristic function

where $\in$ means "belongs to" A and A^c are the sets of voids and solids, respectively, within a porous medium bulk volume, V. These two sets are mutually exclusive and complementary, so it follows that

$$A \cup A^c = V \tag{7.60}$$

and

$$A \cap A^c = \Phi \tag{7.61}$$

where $\cup$ and $\cap$ are union and intersection of sets, respectively, and Φ is the empty set. The expression "random function of coordinate" must be understood in the sense that at each point of the 3D space, the value $\xi(x)$ is a random variable, and, consequently, it cannot be predicted exactly. The values of $\xi(x)$ are subject to a certain PDF. If the PDF is invariant with respect to a shift of the system of points, then the ReV and corresponding porous medium are homogeneous. The same ReV is statistically homogeneous and isotropic when the PDFs are invariant with respect to an arbitrary rotation of the system points (such as a solid body) and to a mirror reflection of the system with respect to an arbitrary plane passing through the origin of the coordinate system. In other words, the statistical moments depend upon the configuration of the grain-void system for which they are formed, but not upon the position of the system in space. In practical terms the porous medium is isotropic if the properties of any point are the same in all directions from that point. The medium is of heterogeneous composition if its nature or properties of isotropy or anisotropy vary from one point to another in the medium. If the porous medium is statistically homogeneous and isotropic, then the moments do not depend on any preferred direction within the medium.

7.5.1 Line Characteristic Function of Porous Medium

Any internal property of porous medium can be quantified by an arbitrary sampling line which passes through a rock in any direction. Each point on this line corresponds to either a void or a solid, and the sample line characteristic function, $f(s)$, can be defined similarly Eq. 7.59 as

$$f(s) = \begin{cases} +1 & \text{if} \quad s \in A \\ -1 & \text{if} \quad s \in A^c \end{cases} \tag{7.62}$$

where s is the distance along the sampling line from any arbitrary origin. The graphical representation of $f(s)$ forms a square wave as it passes alternately from void to grain (Fig. 7.22). This function represents one of the possible realizations of the ensemble of the porous medium. It reflects the size distribution of solids and voids, their orientation, and packing within the sandstone. In practice, $f(s)$ may be

determined by preparing a high-contrast photographic image of a thin section whose pores have been filled with epoxy (Minoura and Coley 1971).

In order to treat $f(s)$ with classical time series techniques, such as autocovariance, spectral, autorun analysis, and so forth, it must be defined as random variable set at n points equally spaced Δs apart along the sampling line. Hence, $n\Delta s$ is equal to the sampling line length. The ordered set of $f(s)$ values at $' = 1, 2, \ldots, \infty$, is called a stochastic process. If the PDF is the same for all I, the stochastic process is said to be weakly stationary. In addition, strict stationarity implies that all of the possible joint PDFs are functions of the distance between pairs. However, in practice, strictly stationary processes are not usually encountered. Non-stationarities in sedimentary rock units may arise from the presence of distinctive layering or gradual grading. In general, sedimentary units are considered as stochastically stationary if the pattern of variation in a property is similar in each sampled area of a bed.

7.5.2 *Autorun Analysis of Sandstone*

The autorun function is especially capable of investigating the sequential properties of random series when only two distinct values exist. Due to its distribution-free behavior, computational simplicity, and robustness to various transformations, it is more flexible than the classical autocorrelation function (Şen 1978). In general, the autorun function is defined for discrete processes as a conditional probability

$$r(k\Delta s) = P[f(k\Delta s) > m / f(\Delta s) > m] \tag{7.63}$$

where k is referred to as the lag and m as the truncation level, taken in most cases as the median value. From the definition it is obvious that values of the autorun function vary between 0 and +1 for any lag-k. Let p be the probability of +1 in a characteristic function (see Eq. 7.62) of infinite length. For a line characteristic function, the truncation level can be taken as any value between +1 and – 1, exclusively. For independent processes, all of the autorun values at different lags are equal to p. However, for any process, the following autorun function properties are valid:

$$\left.\begin{array}{c} r(0) = +1 \\ 0 < r(k\Delta s) < +1 \\ r(\infty) = p \end{array}\right\} \tag{7.64}$$

The small sample estimate of the lag-k autorun coefficient from finite length characteristic functions can be obtained similar to Eq. 7.35 as

$$r(k\Delta s) = \frac{2n_k}{n-k} \tag{7.65}$$

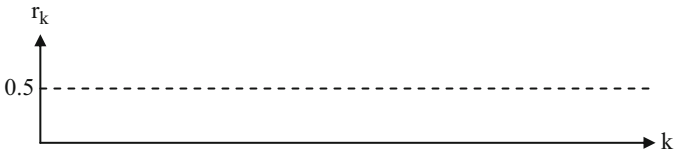


Fig. 7.23 Autorun function of an independent process $[r(k\Delta s) = 0.5,\ p = 0.5]$

where n_k shows the number of overlapping successive + 1 pairs at a distance of $k\Delta s$ apart and n is the number of equally spaced points in the porous medium line characteristic function. In the case of independent process, $r(k\Delta s) = p$ for any lag (see Fig. 7.23). If $l \geq r(k\Delta s) > p$, then on the average, the points lag-k apart fall on void space; otherwise, when $p > r(k\Delta s) \geq p$, they fall on grain space.

The former situation indicates positive dependence and sometimes is referred to as persistence. It implies physically that grain points and void points follow void points. Şen (1978) has shown the analytical relationship between the autorun and autocorrelation coefficients for $p = 0.5$ as

$$r(k\Delta s) = \frac{1}{2} + \frac{1}{\pi}\arcsin\rho_{k\Delta s} \tag{7.66}$$

where $\rho_{k\Delta s}$ is the lag-k autocorrelation coefficient. Assuming a first-order Markov process where $\rho_{k\Delta s} = \rho_{\Delta s}^{k}$, this last expression gives autorun functions as shown in Fig. 7.24. It is obvious that for this case $0 \leq \rho_{k\Delta s} \leq +1$ corresponds to $0.5 \leq \rho_{k\Delta s} \leq +1$. Figure 7.24a–d represent persistence, and the autorun function converges 0.5 asymptotically. These continuous decreases indicate the stationary nature of the underlying sample characteristic function. On the other hand, negative dependence is characterized by autorun functions similar to Fig. 7.24e–h. In Fig. 7.24, r_k implies $r(k\Delta s)$ and $r(\Delta s)$.

The autorun coefficient is robust and is not dependent on any particular distribution function. Furthermore, it is very convenient for binary random variables. The physical interpretation of the autocovariance function may be difficult or impossible to make (Jenkins and Watt 1968). Contrarily, the autorun function has physical meaning as conditional probabilities and is usually simple to interpret. The asymptotic value of the autorun function corresponds to the porosity of the porous medium. However, in practice autorun coefficients for relatively large lags estimated from Eq. 7.65 yield the approximate porosity.

Another very important physical parameter that is directly related to the autorun function is the specific surface, σ. The specific surface is defined as the ratio of the total surface, S', of solids (or voids) to the bulk volume of the porous medium concerned. Hence, generally

$$\sigma = \frac{S'}{V} \tag{7.67}$$

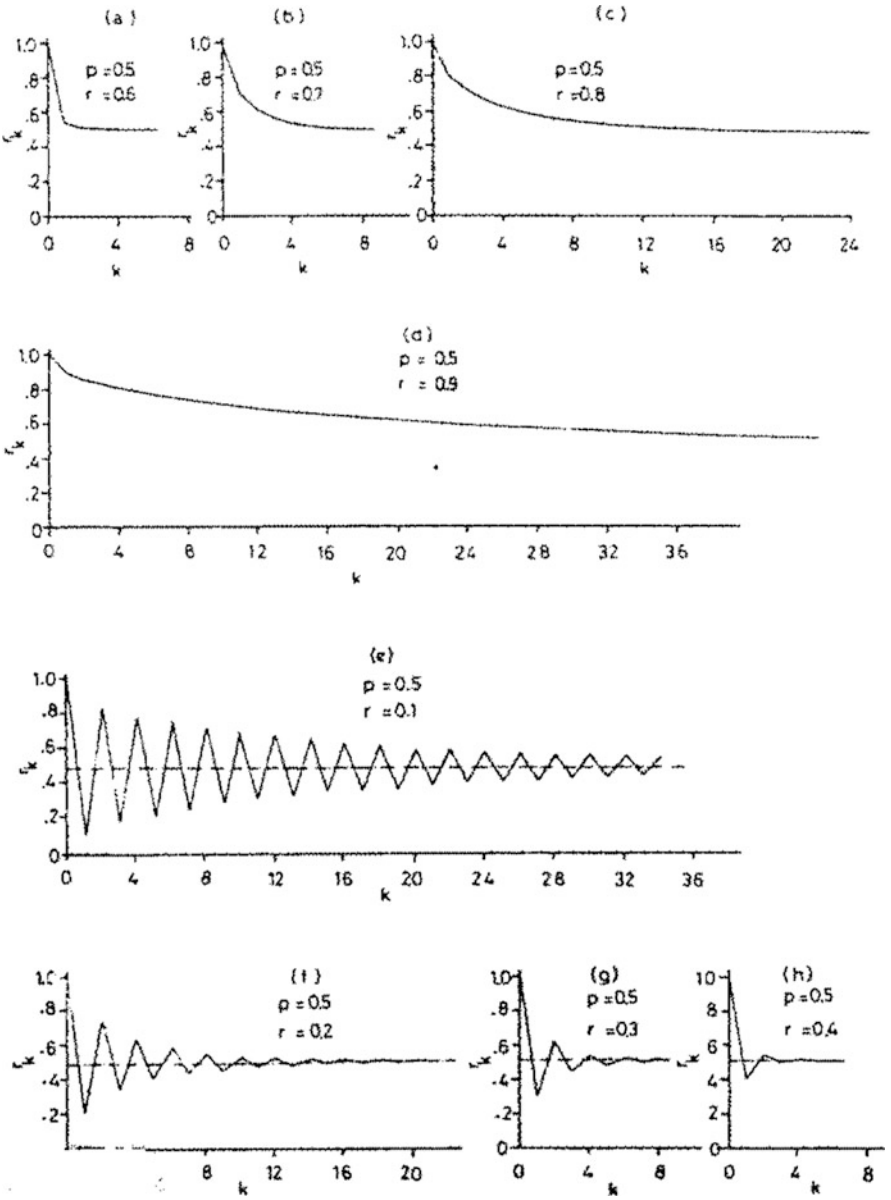


Fig. 7.24 Autorun function of the dependent process for different $r(k\Delta s)$ values and $p = 0.5$

Knowledge of the specific surface is important in cases of adsorption of materials from the fluid flowing through the porous medium. It is also important also in the design of sand and gravel filter columns. In hydrogeology it is related to the specific retention of water in sandstone. The finer the grains, the greater will be the specific retention and specific surface.

The random variable, $\eta(x)$, identifying this surface can be obtained from the basic ReV as

$$\eta(x) = \lim_{\Delta s \to 0} \xi(x)\xi(x + \Delta s) = \begin{cases} +1 & \text{for} \quad x \in S^c \\ -1 & \text{for} \quad x \in S \end{cases} \tag{7.68}$$

where S is the set of points in the porous medium resting on the grain surface and S^c is the complementary set including the remaining points. Hence, – 1 is important in the definition of grain surface. In fact, points on S physically correspond to "crossings" from a grain to a void or vice versa, that is, grain boundaries. Along any sampling line, these crossings occur as random points after each square wave. In the statistical literature, these points are referred to as the zero crossing (Mood 1940). The total number of such crossings in a continuous sequence has been given originally by Rice (1945). An approximate but practical calculation to total cross numbers, T, in any discrete process has been given by Şen (1980) in terms of the first-order autorun coefficient ($k = 1$), as

$$T = 2n[1 - r(\Delta s)]p \tag{7.69}$$

Here T represents points of solid surface along any sampling line provided that the porous medium is isotropic and homogeneous. Hence, the estimate of specific surface $\hat{\sigma}$ along such a line can be determined after dividing Eq. 7.69 by the length of this line $n\Delta s$

$$\hat{\sigma} = 2\frac{1 - r(\Delta s)}{\Delta s}p$$

or more conveniently

$$\hat{\sigma} = 2\frac{r(0) - r(\Delta s)}{\Delta s}p \tag{7.70}$$

It is important to notice at this stage that the ration in Eq. 7.60 is the slope of the autorun function at the origin. A similar relationship has been obtained with the autocorrelation function slope at the origin by Watson (1975). The population (asymptotic) value of the specific surface can be found from Eq. 7.70 as $\Delta s \to 0$ which leads to

$$\sigma = -2r'(0)p \tag{7.71}$$

where $r'(0)$ is the derivative (slope) of the autorun function at the origin. For independent processes $r'(0)) = -\infty$ and for completely dependent processes $r'(0))$ 0. It is therefore expected theoretically that the specific surface is isotropic and homogeneous materials may take any positive value. Since processes in earth sciences fall between the two aforementioned extremes, their specific surface

values fall between zero and infinity. It has been shown by Cramer (1938) that the specific surface of sandstones varies in the range of 150–320 cm^{-1}. The finer the sandstone, the greater the number of crossings on any sampling line, and, therefore, the slope of the autorun function will be greater, leading to greater specific surfaces. It thus becomes very obvious that fine materials will exhibit much greater specific surface than coarse materials. Some fine materials contain an enormous grain surface area per unit volume. The generalization of Eq. 7.71 for any preferred direction, say α, within the anisotropic porous medium is possible after dividing it by the surface area of the sphere with unit radius. Hence, Eq. 7.71 becomes

$$\sigma = -\frac{1}{2\pi}\int r'_{\alpha}(0)p_{\alpha}d\alpha \tag{7.72}$$

where $r'_{\alpha}(0)$ and p_{α} are the derivative of the autorun function at the origin and the porosity along this preferred sampling line, respectively. The numerical calculation of Eq. 7.72 can be achieved by taking various thin sections on different directions.

7.5.3 Autorun Modeling of Porous Media

Since the porous medium has been regarded as a realization of a stochastic process, it is necessary to develop a model for generating pore structure that is more representative than those based deterministically on assemblages of spheres or tubes. However, the spherical beads and capillary tube models are assumed for analytical purposes in solving for fluid flow on the scale of a few pores. Therefore, they are not convenient for generating the stochastic characteristics of the porous medium. A simulation model of the grain-void size distribution can be achieved through the autorun technique. Although such a model does not give, on the average, new information about the medium, it helps to generate all of the possible line characteristic function realizations of the medium. A simulation model of the grain-void size distribution can be achieved through the autorun technique. Although such a model does not give, on the average, new information about the medium, it helps to generate all of the possible line characteristics of the medium. The first simulation model in this direction has been proposed by Roach (1968) for independent processes.

However, in general the porous medium composition of voids and solids has dependent structure. This is tantamount to having clusters of voids and/or solids; that is as a general tendency voids follow voids and solids follow solids. The first autorun coefficient, $r(\Delta s)$, provides a criterion to decide whether the porous medium composition has dependent composition; that is, occurrence of any void or solid does not affect others. In cases of $r(\Delta s)$ being significantly different from the porosit, clustering exists. When $r(\Delta s) > p$ a positive

dependence exists, which means physically that the clustering of voids is predominant. On the contrary, for $r(\Delta s) < p$, clustering of solids is effective.

The probability of having at least two successive voids is $r(\Delta s)$, of having three successive voids is $r^2(\Delta s)$, and in general having at least j successive voids is

$$P(n_\nu \geq j) = [r(\Delta s)]^{j-1} \tag{7.73}$$

It has been shown by Feller (1967) that

$$P(n_\nu = j) = P(n_\nu \geq j) - P(n_\nu \geq j + 1) \tag{7.74}$$

Hence, substitution of Eqs. 7.73 and 7.74 yields the probability of having j uninterrupted successive voids on an infinite sampling line as

$$P(n_\nu = j) = [1 - r(\Delta s)][r(\Delta s)]^{j-1} \tag{7.75}$$

It indicates that void lengths are geometrically distributed with parameter $r(\Delta s)$. The expectation of void length in an infinite characteristic function can be found from Eq. 7.75 as

$$E(n_\nu) = \frac{1}{1 - r(\Delta s)} \tag{7.76}$$

For independent processes, $p = r(\Delta s)$; this equation gives $E(n_\upsilon) = 1/(1 - p)$, which has been presented by Feller (1967). The purpose of a simulation model is to generate statistically indistinguishable synthetic characteristic functions from the observed line characteristic function. In other words, on the average, statistical parameters such as the mean and variance must be preserved in the synthetic characteristic functions. It is possible to generate geometrically distributed void lengths, υ, with parameter $r(\Delta s)$ as

$$\nu = 1 + \frac{\log\xi}{\log r(\Delta s)} \tag{7.77}$$

where υ is an integer-valued random variable and ξ is a uniformly distributed random variable between 0 and +1. The solid lengths, g, can be generated similar to Eq. 19 as

$$g = 1 + \frac{\log\xi}{\log r_g(\Delta s)} \tag{7.78}$$

where $r_g(\Delta s)$ is the first autorun coefficient calculated from the line characteristics after its multiplication by –1.

Since the void and solid lengths occur alternatively on a sampling line, the initial length can either be selected randomly or according to the final length type on the

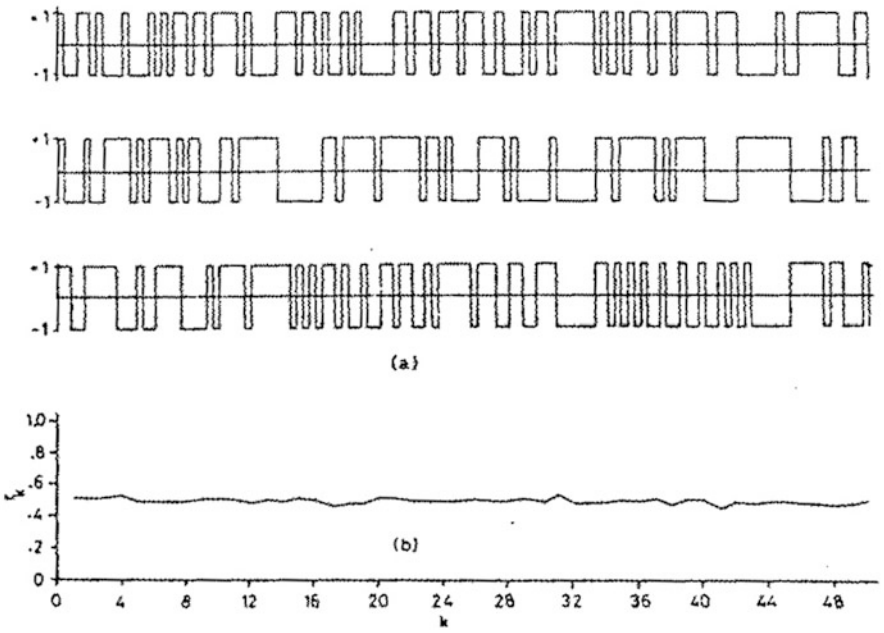


Fig. 7.25 Synthetic line characteristic function $[r(k\Delta s) = 0.5,\ p = 0.5]$

observed sampling line, and then void and solid lengths are arranged in sequence. However, whichever way is adopted does not make any difference in the final product.

The aforementioned modeling technique for the porous media based on the autorun model has been implemented on a digital computer for parameters $p = 0.5$; $r(\Delta s) = r_g(\Delta s) = 0.5$, 0.7, and 0.9. Very long sequences of synthetic line characteristic functions (10,000 points) have been generated, but three samples with length of 125 points each are presented in Figs. 7.25a, 7.26a, and 7.27a, together with their sample autorun functions in Figs. 7.25b, 7.26b, and 7.27b, respectively. Figure 7.25a represents the independent structure of void-solid occurrences resulting from the autorun model, whereas Figs. 7.26a and 7.27a are samples of dependent structure.

It is clear that on the average the lengths of voids or grains are shorter in Fig. 7.25a than in others; in addition Fig. 7.25a is richer in the number of crossings than others. As the autorun coefficient increases, these lengths, as well as the number of crossings, increase. Line characteristic functions in Figs. 7.25a, 7.26a, and 7.27a are possible realizations of fine, medium, and coarse-grained sandstones, respectively.

On the other hand, generalization of the stochastic model proposed by Roach (1968) for finite sampling lines can be achieved by the autorun coefficient so as to cover dependent porous medium configuration. For this purpose, the following

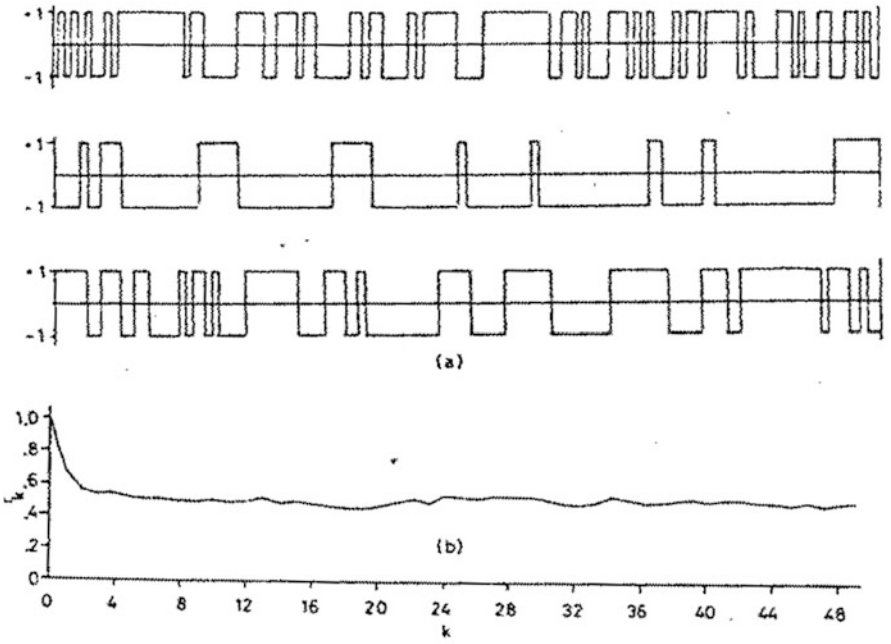


Fig. 7.26 Synthetic line characteristic function $[r(k\Delta s) = 0.7,\ p = 0.5]$

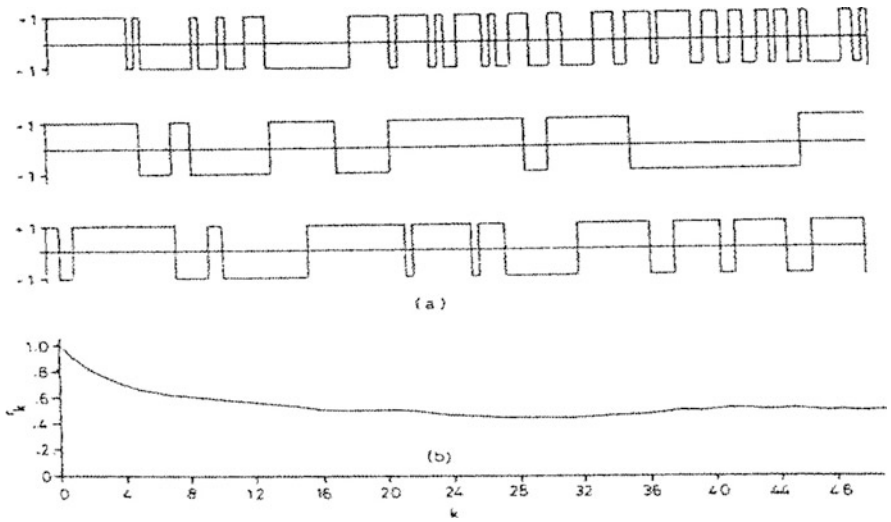


Fig. 7.27 Synthetic line characteristic function, $[r(k\Delta s) = 0.9,\ p = 0.5]$

conditional probabilities of the combinations of two successive events at equally spaced points along a sampling line can be written as

$$\begin{aligned} P(\nu/\nu) &= r(\Delta s) \\ P(g/\nu) &= 1 - r(\Delta s) \\ P(g/g) &= 1 - \frac{p}{1-p}[1 - r(\Delta s)] \\ P(\nu/g) &= \frac{p}{1-p}[1 - r(\Delta s)] \end{aligned} \tag{7.79}$$

On the average, the probability $P(\upsilon)$ that a point lies in a void space is equal to p, whereas the probability $P(g)$ that a point is within a solid space is $q = 1 - p$. The void-solid sequence probability is given in Eq. 7.79 as $P(\upsilon/g)$ which defines a crossing point in the line characteristic function. The probability that a solid is followed by a void, $P(g, \upsilon)$, is equal to $P(g)P(\upsilon/g)$ and explicitly

$$P(g, \nu) = p[1 - r(\Delta s)]$$

However, the probability $P(g, \upsilon, g)$ of having void length equal to $1\Delta s$ can be obtained as $P(g, \upsilon)P(g/\upsilon)$ or as

$$P(g, \nu, g) = p[1 - r(\Delta s)]^2 \tag{7.80}$$

Similarly, in general, the probability of having the void length equal to $n\Delta s$ becomes

$$P(g, \nu, \nu, \nu, \nu \ldots, \nu, g) = p[1 - r(\Delta s)]^{n-1}[1 - r(\Delta s)]^2 \tag{7.81}$$

The number of voids, N_υ, of length $n\Delta s$ in a set of N equally spaced points along the sampling line can be obtained as

$$N_\nu = Np[1 - r(\Delta s)]^{n-1}[1 - r(\Delta s)]^2 \tag{7.82}$$

For the independent process, $r(\Delta s) = p$, and then Eq. 7.82 yields the same result as originally proposed by Roach (1968). Furthermore, a similar expression for the solid can be found as

$$N_g = N(1-p)\left[1 - r_g(\Delta s)\right]^{n-1}\left[1 - r_g(\Delta s)\right]^2 \tag{7.83}$$

Figure 7.28 shows the change in the number of voids per unit length, N_υ/N with the autorun coefficient for given void length of $1\Delta s$, $2\Delta s$, and $3\Delta s$ at different porosities, $p = 0.1$, 0.2, 0.3, and 0.4. It is obvious that the number of voids with length $1\Delta s$ decreases continuously with the increase of the autorun coefficient. However,

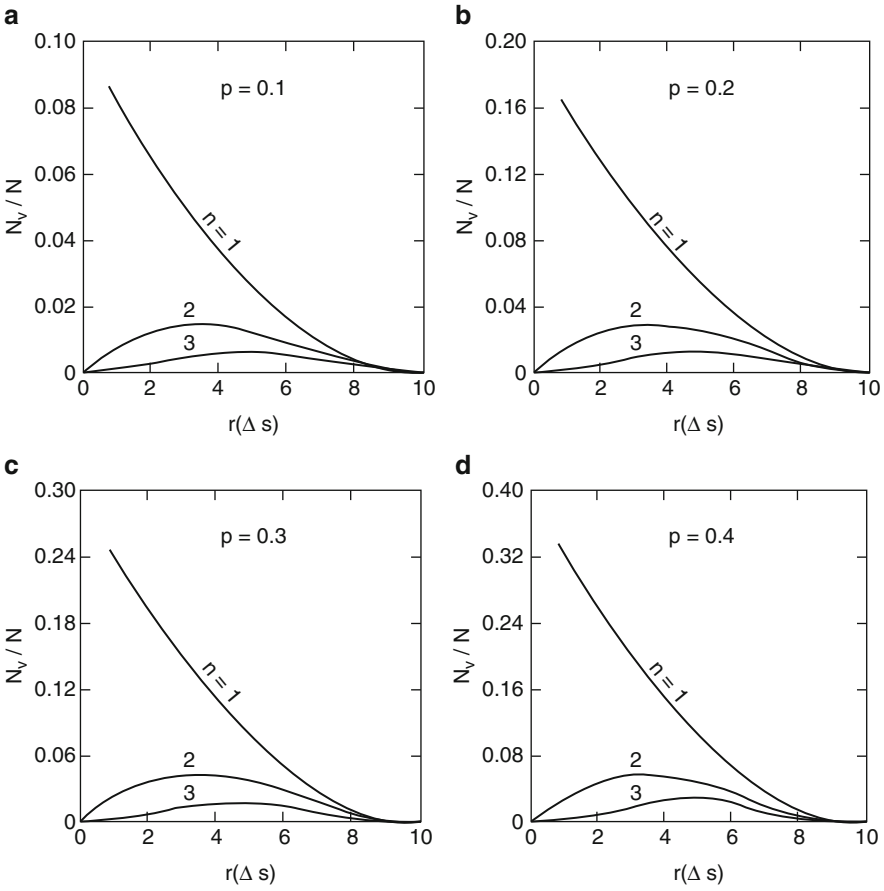


Fig. 7.28 (**a–d**) Number of voids on a line characteristic function

such a relationship is not valid for other void lengths because the number of voids with lengths other than $1\Delta s$ has maximum values for autorun coefficients other than zero. For instance, this maximum occurs at about $r(\Delta s) = 0.3$ for $p = 0.3$. Furthermore, whatever the porous medium parameters are in the long run, the number of voids of any length will be more than voids with smaller lengths. It can be concluded from Fig. 7.28 that the difference between the numbers of voids per unit length for voids of length $1\Delta s$ and others is relatively larger at low autorun coefficients. For a line characteristic function with $N = 10{,}000$ points, $p = 0.3$, and $r(\Delta s) = 0.2$, the number of voids for void lengths of $1\Delta s$, $2\Delta s$, and $3\Delta s$ are 1920, 384, and 77, whereas with $r(\Delta s) = 0.8$, they are 120, 96, and 77, respectively.

7.6 CSV Technique for Identification of Intact Length Correlation Structure

Different rock quality description indices that are presented into the rock mechanics literature, which are based on intact lengths, assume that the occurrences of discontinuities and accordingly intact lengths are independent from each other. Unfortunately, the validity of such an assumption is questionable, and, therefore, the final equations for rock quality assessments need to be validated prior to their use.

Otherwise the results can be regarded only as initial approximations, which might be rough estimations of the rock quality. It is rather obvious that due to different geological phenomena, the rock discontinuities appear in local cluster forms which show that the intact lengths cannot be considered as independent from each other. In order to check whether the intact lengths along a scanline are independent or dependent from each other, herein, the CSV method has been presented in detail. The basic concept of standard CSV is presented and so simplified that in the case of independent intact length occurrences, its variation with the total intact length appears as a straight line which passes from the origin. Any systematic deviation from straight line implies that the intact length is dependent, and therefore classical rock quality description index must not be used, and instead measures which take into account the correlation of intact lengths must be preferred in rock quality studies. Besides theoretical CSV models are developed for the independent, Markov and ARIMA (1, 1) processes. These models provide a general guide in the identification of intact length correlation structure as well as the numerical value of such a correlation.

The application of the methodology developed, herein, is presented for scanline measurements at Tytherington Quarry in England. The Markovian type of correlation structure is identified for all of the scanlines from this quarry.

Mechanical and/or hydraulic behavior of jointed rock masses requires an accurate representation of joint geometry including intact and trace lengths, orientation, intensity, etc. The spacing between two successive joints along an arbitrary sampling line, namely, scanline, is referred to as the intact length. Most of the mechanical behavior and the quality classification of jointed rock masses are directly related to the intact length occurrences along different directions. According to Terzaghi (1965), the intact lengths along the mean pole direction can be regarded as the true spacing, whereas along any other directions, they can be expressed as a function of the true spacing and the angle between the sampling and mean pole directions. In case of several joint sets, within the same rock mass, it is rather difficult to measure true spacing for an individual set. In practice, however, scanline measurements include intact lengths between joints belonging to different sets.

Many investigators have examined empirical joint spacing distributions based on measured intact lengths, and generally two different types of theoretical distribution functions were proposed, namely, lognormal (Steffen 1975; Bridges 1975; Barton 1977; Şen and Kazi 1984) and negative exponential distribution

(Call et al. 1976; Priest and Hudson 1976; Einstein and Baecher 1983; Wallis and King 1980). Besides, gamma distribution of intact lengths has been proposed by Şen (1984). However, it seems that the negative exponential distribution appears to be the most frequently used one due to its simplicity. In order to decide whether the intact lengths fit the negative exponential or lognormal distribution, most often the conclusions are based on visual comparison but Beacher et al. (1977) have used goodness-of-fit tests in the decision making about the best distribution. For such a test the primary requisite is the histogram of intact lengths. However, any test of goodness of fit ignores the internal structure in the intact length occurrence along a scanline, i.e., the intact length occurrences are assumed to have a complete random behavior without any dependence on each other. We cannot really assume that nature is nice enough to present us with independent identically distributed intact lengths. In nature, the creation of fractures depends on many interactive phenomena such as pressure, temperature, volcanic activity, earthquakes, etc. As a consequence, prior to any assessment of intact lengths with the classical techniques, one should confirm their random occurrences. All of the aforementioned studies imply that the intact lengths occur randomly which might not be the case in actual situations, and subsequently overestimations result in the evaluations. Hence, a key question in the intact length evaluation is whether they have an independent structure or not. Provided that the independence is verified, then the use of classical techniques leads to reliable answers. Otherwise, new techniques should be devised so as to take into consideration the dependence structure of intact lengths. This point has been observed first by Higgs (1984) who suggested an empirical technique called "the profile area method" by which the dependence structure within the intact length measurement series is taken into consideration indirectly but in an effective way. Later, Şen (1990b) provided analytical formulation for the profile area method which showed explicitly that the rock quality measures are not only functions of simple statistical parameters such as the mean and variance values but, additionally, the serial correlation coefficient. The classical techniques appear as a special case of the analytical formulations when the serial correlation coefficient is practically equal to zero. On the other hand, Eissa and Şen (1990) made an extensive computer simulation by using Monte Carlo techniques so as to generate serially dependent intact lengths according to a very simple Markov model. The calculations of rock quality designation (RQD) for different serial correlation coefficients indicated that increase in correlation coefficient results in increase in rock quality.

However, neither the empirical method (Higgs 1984) nor the simulation technique (Eissa and Şen 1990) is suitable for practical purposes. It is, therefore, suggested herein the use of CSV concept as suggested by Şen (1989a, b) in evaluating the intact length occurrences and accordingly the rock quality classification. The CSV technique provides graphs which give additional interpretations in intact length occurrences and especially in their regional behaviors.

7.6.1 Intact Length CSV

Intact length measurements in the field are dependent on the relative positions of the scanline within the study area. The variability of intact lengths along a scanline leads to the concept of regional variability of intact length with the change in scanline orientation. This variability determines the regional behavior as well as the predictability of the intact lengths on the basis of which the rock quality appreciations can be assessed. Large variability implies that the degree of dependence of intact length on each other might be rather small even for scanlines close to each other. Such a variability may be a product of either one of the active geological phenomena such as tectonic, volcanic, depositional, erosion, recharge, or other activities.

In order to quantify the degree of variability of regional changes, variance techniques have been used already (Higgs 1984). On the other hand, Eissa and Şen (1990) and Şen (1991) have employed autocorrelation methods in the intact length assessment. However, these methods cannot account for the regional dependence due to either non-normal distribution (Şen 1978) and/or irregularity of sampling positions (Subyani and Şen 1990).

The SV method has been proposed by Matheron (1963) to get rid of the aforementioned drawbacks. Its elegancy is that the regionalized variable PDF is not important in obtaining the SV, and furthermore, it is effective for irregular data positions. It is to be recalled herein that the classical variogram, autocorrelation, and autorun techniques all require equally spaced data values. However, the discontinuities along a scanline are irregularly spaced, and, therefore, use of classical techniques is highly questionable, except that these techniques provide approximate results only. The SV technique, although suitable for irregularly spaced data, has practical difficulties as summarized by Şen (1989a, b). Among such difficulties is the grouping of distance data into classes of equal or variable lengths for SV construction, but resulting SV results in an inconsistent pattern and does not have a nondecreasing form.

However, adaptation of CSV gives with the same data always a nondecreasing pattern without grouping of distances but rather their arrangement in ascending order. By this arrangement each one of the distances is considered individually in the regional variability of the intact lengths. In general, CSV, $\gamma_s(d_k)$ is defined as a successive summation of square differences as

$$\gamma_c(d_k) = \sum_{i=1}^{k} d\left(d^i\right) = \frac{1}{2}\sum_{i=1}^{k} \left(Z_i - Z_{i-1}\right)^2 \tag{7.84}$$

in which $d(d^i)$ indicates half-square difference at ith order in an ordered intact length arrangements where superscript i indicates the rank; Z_i is the intact length corresponding to rank i.

7.6.2 Theoretical CSV Model

Equation 6.1 in its present form is useful to obtain experimental CSV from a given set of intact lengths which will be performed in the application. However, the interpretation of these CSVs requires a set of relevant theoretical models for their comparison. It is possible to derive theoretical models provided that the underlying generating mechanism of the intact lengths is known or assumed. To start with let us take expectation of both sides in Eq. 7.84 leading to (after the expansion of parenthesis term on the right-hand side)

$$E[\gamma_c(h_k)] = \frac{1}{2}\left[E\left(Z_i^2\right) - 2E(Z_iZ_{i-1}) + E\left(Z_{i-1}^2\right)\right] \tag{7.85}$$

Let us assume that the intact lengths are second-order stationary, which implies that $E\left(Z_i^2\right) = E\left(Z_{i-1}^2\right) = \sigma^2$ and $E(Z_iZ_{i-1}) = \text{cov}(Z_iZ_{i-1}) + \sigma^2$ where $\text{cov}(Z_i, Z_{i-1})$ indicates the covariance between Z_i and Z_{i-1} and finally σ^2 is the variance of intact lengths. It is to be noted that second-order stationary implies that the statistical parameters are independent from i, i.e., from rank. Consideration of these conditions and the substitution of the relevant values into Eq. 2 lead to

$$E[\gamma_c(h_k)] = \sigma^2\left[k - \sum_{i=1}^{k}\frac{\text{cov}(Z_i, Z_{i-1})}{\sigma^2}\right] \tag{7.86}$$

or since from statistical time series analysis (Box and Jenkins 1970) the ratio term in the parenthesis is defined as the lag-k autocorrelation function, ρ_i, then Eq. 7.86 can be rewritten succinctly as

$$E[\gamma_c(k_k)] = \sigma^2\left(k - \sum_{i=1}^{k}\rho_i\right) \tag{7.87}$$

in which ρ_i represents the correlation within the intact length sequence, and this expression is a general formulation of the theoretical CSV. The following specific models of CSV can be derived from the available stochastic processes in the literature as employed by Eissa and Şen (1990):

1. Independent Model CSV: This is the most widely used assumption in the rock quality assessments that have appeared in the literature (Priest and Hudson 1976; Hudson and Priest 1979; Şen 1984, 1990b). The uncorrelatedness of intact lengths implies $\rho = 0$, and consequently, the simplest model of all emerges as

$$E[\gamma_c(k_k)] = k\sigma^2 \tag{7.88}$$

which means that in the CSV model of such intact length occurrences, the variance plays the dominant role only. The graphical representation of

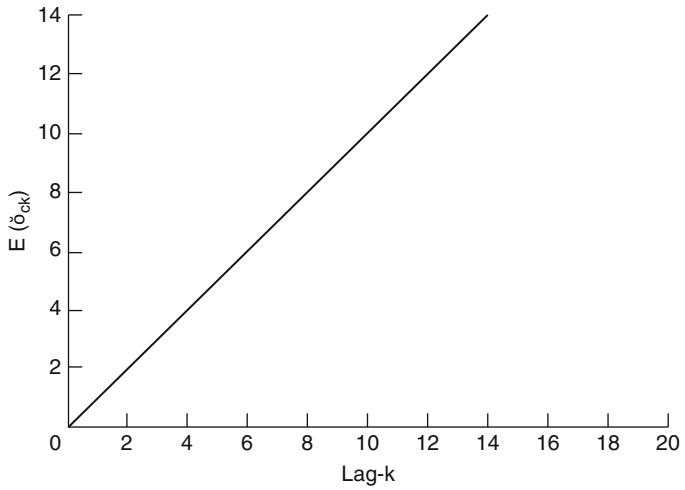


Fig. 7.29 Theoretical standard CSV model of independent intact lengths

independent CSV model appears as a straight line that passes through the origin as shown in Fig. 7.29. The slope of this straight line gives the variance of intact lengths.

As the variance value becomes smaller and smaller, the independent process CSV becomes closer to the horizontal axis which represents the distances. For instance, in the case of equal intact lengths along a scanline, the variance becomes zero and hence the horizontal axis represents the CSV. Otherwise, for large variances the CSV model becomes closer to the vertical axis indicating significant and random differences between successive intact lengths.

2. Markov Model CSV: When the intact lengths are serially dependent and the correlation coefficient decreases according to a power law with the increase of lag, then the generating mechanism of them is Markov process, which has the autocorrelation structure as $\rho_0 = 1$, but $\rho_i = \rho_1^i (i = 1, 2, \ldots)$, where ρ_i is the lag-one correlation coefficient which may assume any value between -1 and $+1$. For this model the CSV model becomes

$$E[\gamma_c(k_k)] = k\left(1 - \rho_1 \frac{1 - \rho_1^k}{1 - \rho_1}\right)\sigma^2 \tag{7.89}$$

which reduces to Eq. 7.87 for $\rho_1 = 0$. Positive ρ_1 1 value means that the average, long intact lengths follow long intact lengths and short intact lengths follow short intact lengths. On the other hand, negative ρ_1 implies that long intact lengths follow short intact lengths or vice versa. Equation 7.84 has power form for small k values but becomes straight line for large k values as

$$E[\gamma_c(k_k)] = k\left(1 - \frac{\rho_1}{1 - \rho_1}\right)\sigma^2 \tag{7.90}$$

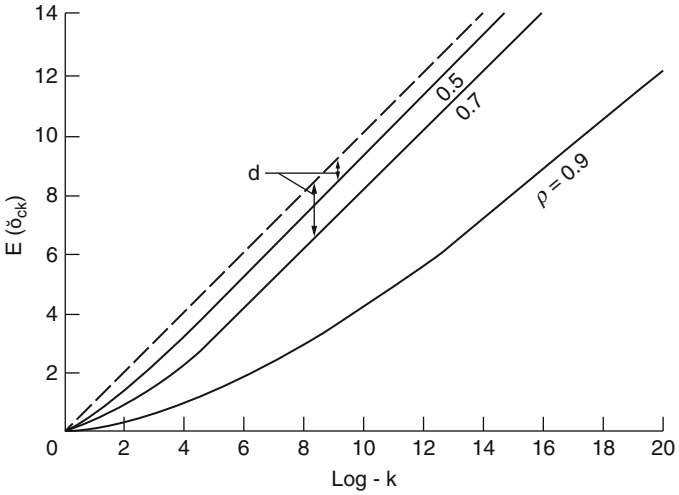


Fig. 7.30 Theoretical standard CSV model for Markovian intact lengths

Division of both sides by σ^2 gives rise to the definition of standard CSV, $E[\gamma_c(k_k)]/\sigma^2$, as

$$\Gamma(k) = k\left(1 - \frac{\rho_1}{1-\rho_1}\right) \tag{7.91}$$

The graphical representation of $\Gamma(k)$ is given in Fig. 7.30.

1. ARIMA Model CSV: These models have been proposed by Box and Jenkins (1970) and applied to the intact length simulation by Eissa and Şen (1990). The autocorrelation structure of this model is given as

$$\rho_0 = 1$$
$$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}$$
$$\rho_i = \phi\rho_1^{i-1}$$

for $i > 2$. ARIMA model represents intact lengths which are more persistent than the Markov model case. The substitution of autocorrelation structure into Eq. 7.87 yields after necessary algebraic manipulation to

$$E[\gamma_c(k_k)] = \left(k - \rho_1\frac{1-\phi^k}{1-\phi}\right)\sigma^2 \tag{7.92}$$

where ϕ is the model parameter representing extra persistence on the Markov model. The standard CSV model is

$$\Gamma(k) = k - \rho_1 \frac{1 - \phi^k}{1 - \phi} \tag{7.93}$$

which provides power type of curves for small k values but as k becomes bigger and bigger, it asymptotically approaches to a straight line portion which can be expressed as

$$\Gamma(k) = k - \frac{k\rho_1}{1 - \phi} \tag{7.94}$$

The graphical representation of ARIMA (1, 1) model standard CSVs is presented in Fig. 7.31.

The comparison of this figure with the Markovian standard CSVs in Fig. 7.30 indicates that in the case of ARIMA (1, 1) model, the attendance of CSV to a straight line appears at bigger distances. It is obvious that ARIMA model CSV reduces to Markov and independent model cases for $\phi = \rho_1$ and $\rho_1 = 0$, respectively. Equation 7.94 helps to identify the correlation structure of intact lengths. In order to find the correlation coefficient, it is sufficient to take the derivative of Eq. 7.94 with respect to k which leads to

$$\Delta\gamma_c = \frac{d\Gamma(k)}{dk} = 1 - \frac{\rho_1}{1 - \phi} \tag{7.95}$$

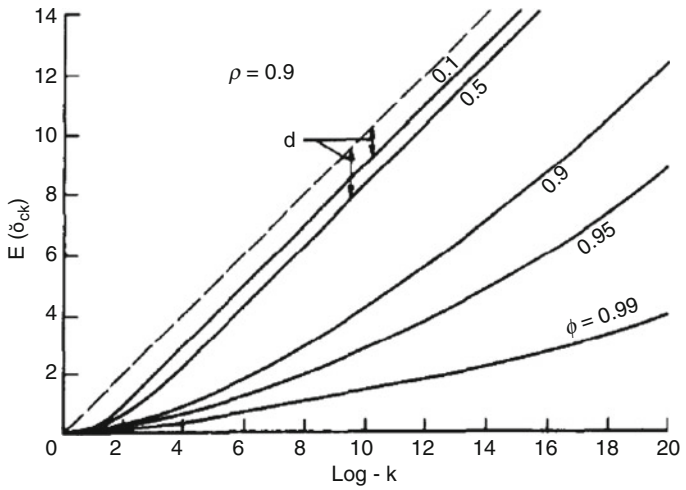


Fig. 7.31 Theoretical standard CSV model for ARIMA intact lengths

in which $\Delta\gamma_c$ indicates the slope of standard CSV at large distance. For independent process case $\rho_1 = 0$ and hence $\Delta\gamma_c = 1$.

This is the case where the RQD formulations in terms of the average discontinuity number as given by Priest and Hudson (1976) or Şen (1984) for different distributions are valid. Otherwise, deviations from the straight line that passes through the origin on a standard CSV plot indicate invalidity of all these formulations since the intact lengths are correlated.

If the correlation structure is of Markovian type, i.e., $\phi = \rho_1$, then Eq. 7.95 can be rearranged for the first-order (lag-one) autocorrelation coefficient estimation as

$$\rho = \frac{1 - \Delta\gamma_c}{2 - \Delta\gamma_c} \tag{7.96}$$

Example 7.1 The field data for the implementation of the methodology developed herein are extracted from a previous study by Ryckes (1984) on modes of failure and the stability of rock slopes. The scanline measurements are carried out at the Tytherington Quarry, which is located halfway between Thornbury and Tytherington about 16 km north of Bristol in the county of Avon. The Paleozoic rocks of the lower coal series in the area are affected by many movements in the past which led to faults, folds, and unconformities each of which led to different pattern of discontinuity spacing, i.e., intact length distributions. Due to these different mechanisms, it is not possible to expect that the intact length sequence along any scanline in this area has independent correlation structure. In order to depict the regional discontinuity pattern in this area, three scanlines were set up at different directions. The bedding plane orientation is almost horizontal for the first scanline which will be referred herein as SL1. The second scanline, SL2, has a direction of 20° toward the southwest, whereas the third scanline, SL3, has an intermediate inclination to the former. All of the scanlines were set up horizontally as two joint sets were recognized to be vertical. The geometrical and statistical summaries of these three scanlines are presented in Table 7.6. Due to its

Table 7.6 Scanline data characteristics

Parameter	SL1	SL11	SL12	SL2	SL3
Length (m)	13.30	4.82	6.19	7.80	11.19
Intact length number	75	35	40	79	73
Average	15.48	13.78	15.48	10.00	15.33
Median	8.00	10.00	7.50	7.00	11.00
SD	219.5	14.87	22.18	10.14	14.05
RQD (%)	84.5	88	79.5	74	86.7
Distribution	Lognormal (Fig. 7.4)	Exponential (Fig. 7.5)	Lognormal (Fig. 7.6)	Exponential (Fig. 7.7)	Lognormal (Fig. 7.8)
x^2-square	0.260	0.350	7.880	9.046	2.34

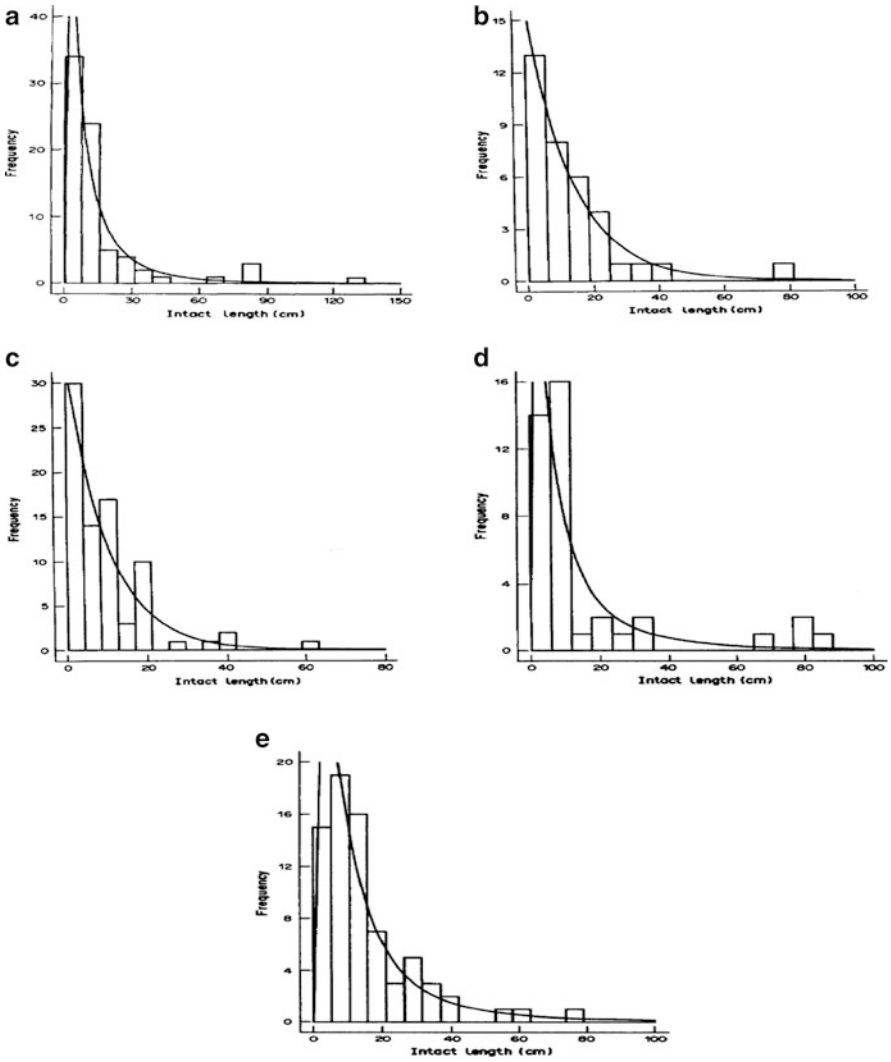


Fig. 7.32 Sample and theoretical PDFs for intact lengths along (**a**) SL1, (**b**) SL11, (**c**) SL12, (**d**) SL2, (**e**) SL3

heterogeneous structure, the first scanline will be examined in two subsets, which will be referred to as SL11 and SL12.

The frequency distribution functions for scanlines are shown in Fig. 7.32.

The theoretical PDFs have been fitted to the experimental counterparts by using χ^2 tests. For five different scanlines, two types of intact length distributions, namely, negative exponential and logarithmic normal, emerge as representative. This difference in the intact length distribution indicates very clearly that their

occurrences within the rock mass are heterogeneous, i.e., direction dependent, and consequently, even from this observation, one can qualitatively conclude that intact lengths have dependent structure. However, classical statistical tests such as the one used for distribution function identification do not give any clue about the correlation structure of intact lengths.

Eissa and Şen (1990) suggested the use of correlation function technique in intact length structure exploration. However, correlation techniques are valid only for the normally distributed intact lengths (Şen 1978). As shown above none of the intact lengths at Tytherington Quarry are distributed normally, and consequently the use of correlation techniques is not meaningful. However, the non-normal intact lengths, especially lognormally distributed ones, can be transformed into normal distribution easily but the transformation of negative exponential distribution exposes great difficulties. Nevertheless, even if the transformation is possible, the transformed intact lengths will not reflect genuine properties of original lengths.

As mentioned in the previous section, the CSVs are robust and valid for any PDF. In fact, the central limit theorem of classical statistics states that whatever the underlying probability distribution of a random variable, its successive summations or averages will have normal distribution. The experimental CSVs for each one of the scanlines considered in this study are presented in Fig. 7.33. At a first glance, one can observe the following significant points:

1. None of the CSVs appear as a straight line passing through the origin. This is tantamount to saying that the intact length occurrences along any scanline cannot be considered as independent processes but they emerge all from dependent processes. This further implies that in the creature of discontinuities within the rock mass, uniform conditions did not prevail but rather a complex combination of multitude geological events such as tectonics, cooling, volcanic activities, etc. took place jointly.
2. The initial portions of each experimental CSVs shifts toward the distance axis. Such a shift implies the existence of positive correlation between successive intact lengths. It further implies that in general big intact lengths follow big intact lengths and small intact lengths follow small intact lengths. Otherwise, if negative correlation should prevail, in the occurrence of intact lengths, then the initial CSV portion would shift toward the vertical axis.
3. Each experimental CSV shows fluctuation about a straight line at large distance. As stated before the existence of such a straight line as a portion in the CSV implies that over the distance range of this straight line portion, the intact lengths are independent from each other. This is the only range where the classical rock quality designation formulations keep their validity. Local deviations from the straight line indicate the hidden or minor dependencies in the intact length evaluation.
4. Over an initial distance range, R, the experimental CSV appears as a curve. This range is defined quantitatively as the distance between the original point and the abscissa of the initial point of the late straight line portion as mentioned in step (3). It is worth mentioning, herein, that as long as the threshold value necessary

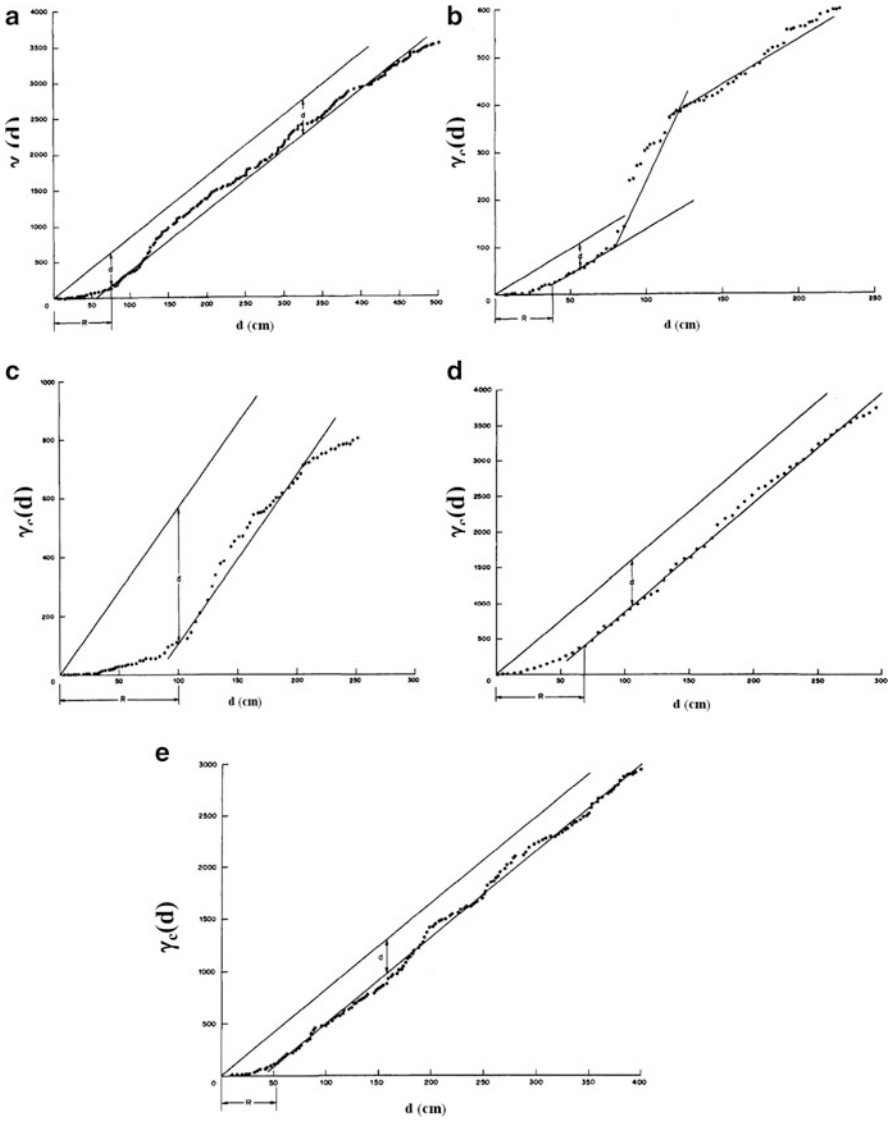


Fig. 7.33 Sample standard CSV of (**a**) SL1, (**b**) SL11, (**c**) SL12, (**d**) SL2, (**e**) SL3

in rock quality designation (RQD) calculation is greater than this range value, then the calculation is theoretically sound and valid. Otherwise, for threshold values less than this range, value will give RQD values which are unreliable. Hence, CSV provides a criterion for checking the validity of classical RQD values based on 0.1 m (4 in.). Such a valuable criterion cannot be obtained with the correlation techniques. These techniques are valid only for the distance domain over which the CSV appears as a straight line.

Table 7.7 CSV parameters

Parameter	SL1	SL11	SL12	SL2	SL3
Range, R(m)	0.74	0.37	1.00	0.68	0.52
Vertical distance, d(m^2)	475.0	54.0	465.0	650.0	325.0
Slope, Δy_c(m)	8.26	2.00	5.70	15.38	8.06
Correlation coefficient, p	0.495	0.497	0.493	0.460	0.489

5. Existence of straight line at large distances identifies the underlying generating mechanism of the intact lengths as a Markov process. It is already mentioned in the previous section that if the curvature continues even at reduced rates at large distances, then the ARIMA (1,1) process becomes effective. In this section, all the scanlines considerations have Markovian dependence.
6. The slope of the large distance straight line portion is related to the standard deviation of the underlying intact lengths. It is possible to consider this slope as the population standard deviation of the intact lengths.
7. The vertical distance between the late distance straight line and the one drawn parallel to it passing through the origin reflects the magnitude of correlation coefficient of the intact lengths. Obviously, the smaller this distance, the smaller will be the intact length correlation. Under the light of the abovementioned points, the relevant numerical values concerning the initial range, R; vertical distance, d; and slope $\Delta\gamma_c = 1$ of final straight line, and the correlation coefficient estimation from Eq. 6.13 is summarized in Table 7.7.

It is obvious from this table that the correlation coefficient along each scanline has more or less the same magnitude, and, therefore, the intact length dependence may be regarded as regionally isotropic. It implies that only one type of model which is already identified as Markovian type can be used in intact length description. The use of this model in intact length generation is outside of this paper's scope. Last but not the least, the following particular points for each scanline can be inferred from the CSV graph comparison from Fig. 7.33:

1. The CSVs of SL1, SL2, and SL3 have the same pattern in that they are composed of initial curvature portion and large distance straight line. However, local deviations around this straight line are rather more persistent in the case of SL1 than others. This indicates that the intact lengths along SL1 have the same secondary internal structure. In order to discover this structure, SL1 is divided into two mutually exclusive portions, namely, SL11 and Sl12.
2. The CSVs for SL11 and SL12 in Fig. 7.33b, c are distinctively different than others. They exhibit not a single straight line but two successive straight lines. This implies that along the SL1 direction, there is another secondary geological event which plays a role in the occurrence of discontinuities along this scanline. A reliable explanation on this point can be arrived only after a detailed geological study of the quarry considered.

One of the most recent and significant questions about the intact lengths is whether they have independent or dependent occurrences along a given scanline. In the past, without consideration of such a question, all of the rock quality classifications derived from the intact length properties were based on the assumption that the intact lengths are identically and independently distributed. The difficulty was due to the lack of a reliable technique in quantifying the correlation structure of intact lengths. However, in this paper, standard CSVs are proposed as a practical tool in measuring the intact length correlations. The CSV calculations are straightforward without any difficulty or ambiguity. The CSV graphs show the change of half-square differences between intact lengths with ordered distance. The intact lengths are independent only when the standard CSV variation with distance appears as a straight line that passes through the origin. Otherwise, they are dependent, and according to the dependence structure, standard CSV graphs take different shapes. However, they have some common property such as that at small distances they appear as curves, whereas at big distances again straight lines dominate but their extensions do not pass through the origin. The slope of the straight line portions on standard CSV plot gives opportunity in calculating the intact length correlation coefficient. For instance, if this slope is close to unity, then and only then the intact lengths can be assumed as independent, and consequently theoretical RQD relationships with the average number of discontinuities are reliably used in any rock evaluation project. The application of methodology developed in this paper has been performed for field data obtained at Tytherington Quarry in England.

7.7 Multi-directional RQD Simulation

A simple fracture network model is proposed based on the negative exponential PDF of the discontinuities in all directions. In the fracture generating procedure, the midpoint coordinates of each fracture is generated by a uniform PDF; the fracture directions are simulated by normal PDFs and the fracture lengths are derived from a lognormal PDF. Three different sets of fractures are presented on a planar surface with different random seeds. The main purpose of this paper is to assess the rock quality designation (RQD) on the basis of one, two, and three directional fracture sets. All previous studies up to date theoretically or practically were confined to one direction only. A graphical procedure is presented for depicting the population estimate of the average number of discontinuities along a scanline. The substitution of this estimate into the relevant RQD equation provides the rock quality estimation in a better representative manner than the classical estimates.

Fracture network models are useful tools for understanding the geotechnical and geohydrological behaviors of rock masses. These networks can be constructed by adopting a model for the fracture geometry (position, direction, trace length, aperture, roughness, etc.), estimating the statistical distribution of the appropriate geometric parameters through field measurements, and then generating realizations

of statistically indistinguishable networks. Once the geometry of a particular realization is specified, then either flow through the network or the geotechnical properties can be studied. For example, one might benefit from such a procedure in the study of average fracture lengths in various directions, extreme values of apertures or traces, average permeability of fracture network under various boundary conditions, etc. Last but not least, fracture network models might be a part of a larger study of hydrologic response to a perturbation such as the construction of an underground opening containing nuclear waste.

Simple simulation studies were initiated by Goodman and Smith (1980) for unidirectional discontinuity spacing with the purpose of obtaining an experimental relationship between the rock quality designation (RQD) and the average number of discontinuities. On the other hand, Baczynski (1980) has proposed a zonal concept for spatial distribution of fractures in rock masses. However, many fracture network studies have been undertaken especially for the water flow through interconnected fractures. Among such studies are the equivalent porous medium permeability in network of discontinuous fractures by Long (1983), the groundwater flow through fractures by Rouleau (1984), and the mass transport in fractured media by Smith and Schwartz (1984). Recently, extensive computer simulation studies were performed by Şen (1984, 1990a) and Eissa and Şen (1990) in assessing the rock quality classifications. To the best of author's knowledge so far, all of the fracture simulation studies in the geomechanics domain are performed unidirectionally for RQD.

7.7.1 Fracture Network Model

In general, the dimensions, orientation, trace lengths, apertures, and other geomechanical properties of each fracture are randomly distributed in nature. Therefore, their replications through Monte Carlo simulations lead to realizations of a possible portion of the fractured medium with the same statistical properties as the observed ones. Once a representative simulation model is developed, it is then possible to obtain the geomechanical properties of the fracture network in detail. Herein, the main geomechanical property is the directional rock quality designation (RQD). In order to achieve such a goal in this paper, the discontinuity network properties are assumed and defined by the following characteristics:

1. The discontinuities within a fracture network are assumed to occur randomly with their midpoint coordinates, obeying the uniform PDF within a unit frame of 1×1. It is possible to enlarge such a basic frame by suitable horizontal and vertical scales in order to obtain desired frame dimensions. The distribution of discontinuity midpoint location in the frame obeys a 2D Poisson process.
2. Throughout this study the discontinuities are assumed to be planar and of finite trace lengths so that intersection by any arbitrary plane gives rise to a group of linear traces. These traces can be represented by an average length and a measure

of variability in terms of standard deviation. In this study, the trace lengths are assumed to originate from a lognormal PDF. Besides, this trace length distribution is independent of the intersection plane orientation.

3. The discontinuity orientations for each set are assumed to have a normal PDF with an average orientation from the north direction, in addition to a reasonable variation expressed by the standard deviation. The occurrences of different directions within each set are independent from each other.

The plausible values of each one of the PDFs mentioned above can be estimated from field studies by examining either borehole cores or exposed rock faces. However, reliable measurements of trace lengths can be obtained from extensive rock faces only. Parameter estimations for the PDFs in describing discontinuities should be determined from both linear and areal measurements. Priest and Hudson (1981) and Hudson and Priest (1983) have presented brief explanations of the scanline measurements, i.e., linear measurements, which provide information about the frequency of discontinuity occurrences along a scanline as well as variability of intact lengths and aperture sizes to a certain extent. However, areal measurements provide significant supplementary information about the discontinuity directions, trace lengths, and midpoints.

7.7.2 RQD Analysis

In its simplest form, the rock quality designation (RQD) is defined as the percentage of intact length summation, S, that is greater than a threshold value such as 0.1 m (4 in.) along a scanline length, L, as

$$\mathrm{RQD} = 100\frac{S}{L} \tag{7.97}$$

The rock mass classification is already given in Table 7.1. The practical difficulties in using Eq. 7.97 are that:

1. It gives a single sample value of RQD along one direction. The practical difficulty may arise in trying to get a different scanline direction so as to assess the possible heterogeneities within the rock mass. For instance, along a highway cut or tunnel excavation, only longitudinal scanlines can be measured for RQD calculations. Lateral scanlines are possible only in large diameter tunnels and large highway cuts. Therefore, it is necessary to set up a fracture network model from the available scanline measurements; and then by using this model, many scanlines in a multitude of desired directions can be taken synthetically.
2. Most often in nature, either due to narrow rock surface exposure or weathering or geomorphologic features, it may not be possible to make measurements along rather long scanlines. This gives rise to another source of error or bias in the RQD calculations according to Eq. 6.1, (Sen and Kazi 1984). However, a

well-defined representative fracture network model provides unlimited opportunities in making scanline measurements as long as desired.

3. The RQD calculations from Eq. 7.97 do not give any insight into the intact length distribution. As mentioned by Şen (1984), different distributions provide different formulations of the RQD calculation. However, in a fracture network model, as developed in this paper, the intact length distributions are known, and therefore one can make more accurate RQD estimations.
4. In a scanline measurement, different sets of fractures are measured irrespective of their individual probability distribution functions. Therefore, it is not possible to assess the effect of each set separately from others. However, in a fracture network model, such assessments can be made rather easily, and the contribution of each fracture set on the RQD value can be quantified individually.
5. Along the scanline adopted in the field, there may occur only intact lengths which are all greater (or smaller) than the threshold value. In such a situation, Eq. 7.97 leads to the rock mass as either excellent or very poor. In order to account completely for all of the possible intact lengths, the population distribution of fractures at every direction should be known. Such a task can be achieved easily once the fracture network model is established for the rock mass.

Because of the aforementioned points, it is worth to investigate the RQD value behavior within a rock mass based on a specific intact length distribution which is adopted herein as the negative exponential distribution.

The negative exponential distribution function, $f(x)$, of intact lengths can be expressed through Eq. 7.27 with λ as the average number of discontinuity, i.e., fractures, say along one direction. Priest and Hudson (1981) derived the expected RQD value based on this distribution as in Eq. 7.34 with the threshold value, t. The difference between Eqs. 7.97 and 7.27 is such that the former gives only sample estimate of RQD, whereas the latter provides population estimation. Of course, in any study the latter type of estimation must be preferred, but its calculation is rather difficult. Theoretically, very long scanline measurements might yield approximately the population RQD value with the use of Eq. 7.97. However, practical difficulties such as small rock surface exposures make such an approximation almost impossible. Besides, the short scanline estimates lead to biased estimations, i.e., underestimation of the RQD value even if many scanlines were available (Şen and Kazi 1984). Therefore, in the RQD analysis, as proposed herein a new technique of its estimation by using Eq. 7.27 is suggested. The field measurements will not be processed as usually done by Eq. 7.97 but rather by Eq. 7.27.

To this end, let us consider the cumulative probability distribution of the intact lengths by taking integration of Eq. 7.27 from 0 to any x value. After necessary calculations, this leads to

$$F(x) = 1 - e^{-\lambda x} \tag{7.98}$$

where $F(x)$ is the cumulative probability which is a nondecreasing function of the intact length, x. It assumes any value in the range 0–1. However, in order to bring

this equation into a more practical form, first of all the probability, $P(x)$, of an intact length to be greater than a given length, x, is considered as $P(x) = 1\ F(x)$ and then natural logarithms of both sides in Eq. 7.98 is leading to

$$\ln P(x) = \ln[1 - F(x)] = -\lambda x \tag{7.99}$$

By taking the differentiation of both sides in this expression, it is possible to obtain $d\ln P(x) = -\lambda x$; hence, finally,

$$\frac{dx}{d\ln P(x)} = -\frac{1}{\lambda} \tag{7.100}$$

This equation shows a linear relationship between $\ln P(x)$ and x, and the slope of straight line is equal to $1/\lambda$. Consequently, the following points can be advised for a practicing geologist in order to find the λ value:

1. Find the probability, $P(x)$, of each intact length measurement.
2. Plot on a semilogarithmic paper intact lengths versus the probabilities on the logarithmic axis.
3. Fit a straight line through the scatter of field points.
4. Calculate the slope value of the straight line. This slope value is, in fact, equal to $1/\lambda$.
5. Substitute the 7 l value into Eq. 6.3 and find the population RQD value.

The only practical difficulty in this procedure is encountered in step 1 for calculating the cumulative probability. However, there appears to be two distinctive approaches. One is the usual way of constructing a histogram for the intact lengths and then transforming it into a cumulative PDF. However, this application raises practical difficulties, especially when the number of intact lengths is small, which is almost the case in any field survey. Another way to obtain the cumulative probability is by means of an empirical approach based on the rank, m, of any intact length within a data size, n. Hence, the cumulative probability can be expressed as

$$P(x) = \frac{m}{n+1} \tag{7.101}$$

This empirical approach is used very often in the practical studies, and it does not give any difficulty in the applications. The RQD analysis based on this empirical law, together with the generated intact lengths, is given in the following section.

7.7.3 RQD Simulation Results

The simulation studies in this paper are based upon a computer program which has been developed to generate in two dimensions a random network of discontinuity

Table 7.8 Input data for fracture network generation

		Set 1		Set 2		Set 3	
Property	Distribution	Mean	Variance	Mean	Variance	Mean	Variance
Orientation	Normal	30	9	130*	9	0	4
Trace length (m)	Lognormal	0.2	0.05	0.18	0.04	0.16	0.04
Intact length (m)	Negative exponential	0.013	2×10^{-4}	0.013	2×10^{-4}	0.013	2×10^{-4}
Location	Bristol	0.5	0.5	0.5	0.5	0.5	0.5

pattern. Three different orientations are adopted for the fracture orientation, each of which is assumed to be normally distributed about a given value. Table 7.8 shows the relevant parameters taken for the simulation of fractures in this paper. However, in any future study, they may be changed according to the field conditions.

This program can typically handle up to any number of sets, each containing any desired number of discontinuities. The program finds the beginning and ending points' coordinates of any trace. The results of this program appear as a numerical representation of a 2D discontinuity network. Such a network is one of the example samples of random discontinuity geometry generated according to the input data shown in Table 7.8.

The sample area size in these generations is considered as 1×1. However, its transformation into any desired size is possible by the multiplication with any desired scale. Some sample realization of generated discontinuity sets are shown in Fig. 7.34 in detail as individual and combined fracture sets.

It is important to notice in the individual sets, i.e., Fig. 7.34a, that some fracture traces intersect with the general frame of the area. These fractures are referred to as censored trace lengths. The generation scheme gives their total trace lengths, but in practice they are never known. On the other hand, some short length traces appear entirely within the sample area. Consideration of the whole set on one sample area, as in Fig. 7.34c, makes the picture rather complex as it is the case in nature.

In order to assess the effect of different fracture sets individually or collectively on the RQD calculations, the sample area is overlain with scanlines in the east-west direction at different levels between 0 and 1. Of course, it is possible to take a scanline at any level, but the discussion here will concentrate on five different levels which are at 0.1, 0.3, 0.5, 0.7, and 0.9. The computer program lays these scanlines, at these levels, on the sample area of a generated fracture network; subsequently, it finds individually and collectively the intact lengths along them, ordering them in an ascending order, and attaches to each intact length an order to be used in Eq. 7.101. The results are presented on semilogarithmic paper as shown in Fig. 7.35.

A visual inspection of these figures indicates that there appear wide ranges of data scatter, and with the consideration of more than one set of fracture, the scatter becomes less. The best straight lines are fitted through the scatter of points; their slopes are found by considering a full cycle on the logarithmic axis. The

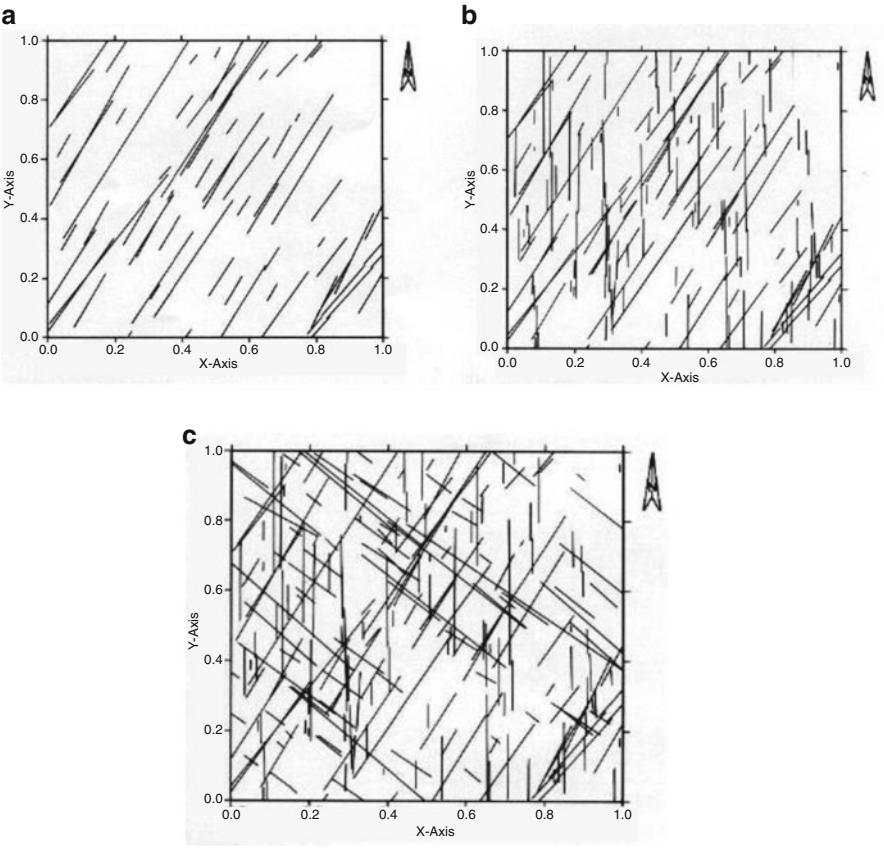


Fig. 7.34 (**a**) A fracture set with 30° orientation from the north, (**b**) two fractures sets with 0° and 30° orientation from the north, (**c**) three fractures sets with 0°, 30°, and 130° orientation from the north

corresponding length on the horizontal axis is equal to the slope, i.e., $1/\lambda$ 1 parameter. This slope calculation procedure is presented in Fig. 7.35c. The results are summarized in Table 7.9.

As is obvious from Table 7.9, the best rock quality is obtained when only one fracture set is considered. As expected, an increase in the number of fracture sets can cause a decrease in the RQD value. There appears a practically significant difference between the RQD values based on the single set of fracture and those on two or three fracture sets. The rock quality is of an excellent type when one set of fractures is considered; whereas the quality deteriorates and becomes very good and fair for two and three fracture sets, respectively. The last row in Table 7.9 gives average RQD values for all the scanlines considered in this study. It is obvious from the comparison of these values that when two sets are considered, the RQD deterioration relative to one set is almost 15 %, whereas for three sets consideration is almost 30 %.

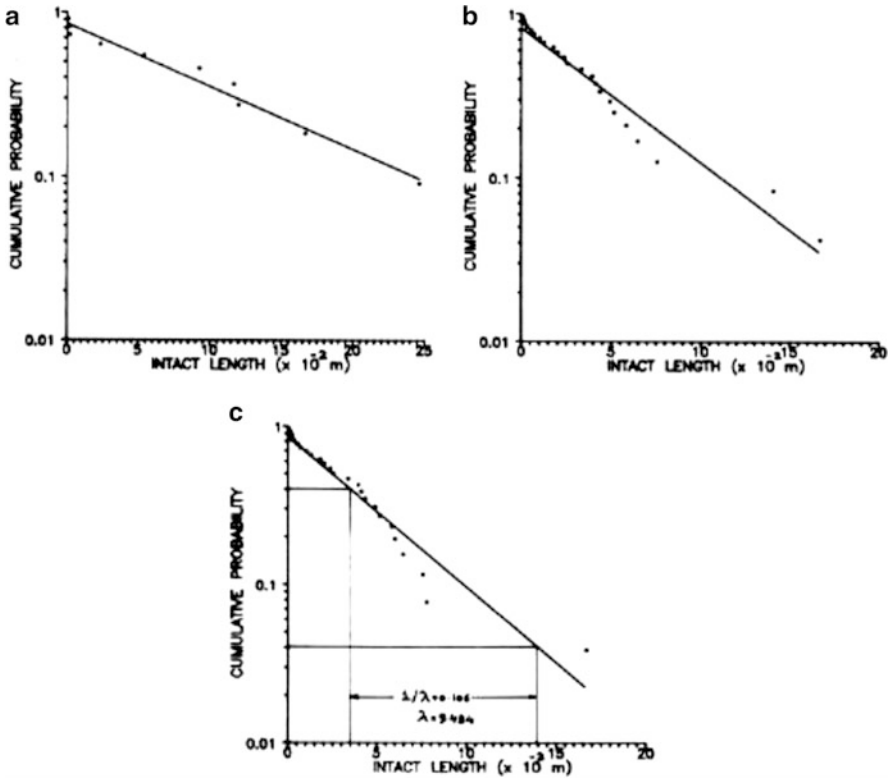


Fig. 7.35 Cumulative probability plots of intact lengths at the 0.1 level. (**a**) Single set at 30°, (**b**) two sets at 0° and 30°, (**c**) three sets at 0°, 30°, and 130°

Table 7.9 Slope and RQD values for $t = 0.1$ m

	Level	One set	Two sets	Three sets
0.1	Slope	0.267	0.120	0.106
	RQD %	94.5	79.8	75.6
0.3	Slope	0.214	0.135	0.128
	RQD %	92.0	83.0	81.0
0.5	Slope	0.227	0.117	0.079
	RQD %	92.7	79.0	63.9
0.7	Slope	0.212	0.122	0.050
	RQD %	91.8	80.0	40.6
0.9	Slope	0.188	0.092	0.076
	RQD %	90.0	70.4	62.0
Average RQD %		92.2	78.4	64.6

References

Agterberg F (1975) New problems at the interface between geostatistics and geology. Geostatistics 75:403–421

Baczynski NRP (1980) Rock mass characterization and its application assessment of unsupported underground openings. Ph.D. Thesis. University of Melbourne, Australia, 233 p

Baecher GB, Lanney NA, Einstein HH (1977) Statistical description of rock properties and sampling. 18th U.S. symposium on Rock Mechanics, PP. 5C1 8

Barton CM (1977) Geotechnical analysis of rock structure and fabric in C.S.A Mine, Cobar, New South Wales. Applied Geomechanics, Technical paper 24, CSIRO, Australia

Barton N, Lien F, Lunde J (1974) Engineering classification of rock masses for the design of tunnel support. Rock Mech 6(4):189–236

Bieniawski ZT (1974) Geomechanic classification of rock masses and its application in tunneling. International Society of Rock Mechanics, 3rd Denver Colorado September 1974 Proceedings, vol 2A pp 2732

Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden Day, San Francisco, 475 pp

Brady BHG, Brown ET (1985) Rock mechanics for underground mining. Allen and Unwin, London

Bridges MC (1975) Presentation of fracture data for rock mechanics. Second Australia New Zealand Conference on geomechanics, Brisborne, Australia, pp 144–148

Brown ET (1981) Rock characterization testing and monitoring. Pergamon Press, Oxford, 221 pp

Call RB, Savely J, Nicholas DE (1976) Estimation of joint set characteristics from surface mapping data. 17th U.S. symposium on Rock Mechanics, pp 2B21–2B29

Cliff AD, Ord JK (1973) Spatial autocorrelation. Pion, London

Coates DF (1964) Classification of rocks for rock mechanics. Int J Rock Mech Min Sci Geomech Abstr 1(3):421–429

Cording EJ, Mahar JW (1978) Index properties and observations for design of chambers in rock. Eng Geol 12:113–142

Cramer (1938) Specific surface?

Cruden DM (1977) Describing the size of discontinuities. Int J Rock Mech Min Sci Geomech Abstr 14:133–137

Deere DU (1963) Technical description of rock cores for engineering purposes, Felsmechanik und in Qenieurgeologie. Rock Mech Eng Geol 1(1):16–22

Deere DU (1964) Technical description of rock cores for engineering purposes. Rock Mech Eng Geol 7(7):16–22

Deere DU (1968) Geologic considerations. In: Stagg KG, Zienkiewicz OC (eds) Rock mechanics in engineering practice. Wiley, New York, pp 4–20

Ege JR (1987) Core index. A numerical core-logging procedure for estimating rock quality. US Geol Surv 954:115

Einestein HH, Baecher GB (1983) Probabilistic and statistical methods in engineering geology, specific methods and examples. Part I: exploration. Rock Mech Rock Eng 16:39–72

Eissa EA, Sen Z (1990) Intact length persistence in relation to rock quality designation. Int J Rock Mech Min Sci Geomech Abstr 28:411–419 (in print)

Feller W (1967) An introduction to probability theory and its application. Wiley, New York, 509 pp

Goodman RE (1976) Methods of geological engineering in discontinuous rock. West group publishing, St. Paul, p 484

Goodman RR, Smith HR (1980) RQD and fracture spacing. J Geotech Eng Div ASCE 106 (GT2):191–193

Hammersely JM, Handscomb DC (1964) Monte Carlo methods. Methuen and Co. Ltd, London. 178 pp

Higgs NB (1984) The profile-area and fracture frequency methods: two quantitative procedures for evaluating fracture spacing in drill core. Bull Assoc Eng Geol 21(3):377–386
Hudson JA, Priest SD (1979) Discontinuities and rock mass geometry. Int J Rock Mech Min Sci Geomech Abstr 16:339–362
Hudson JA, Priest SD (1983) Discontinuity frequency in rock masses. Int J Rock Mec Min Sci Geomech Abstr 20(2):73–89
International Society for Rock Mechanics (1978) Standardization of laboratory and field tests. Suggested methods for the quantitative description of discontinuities in rock masses. Int J Rock Mech Min Sci Geomech Abstr 15:319–368
Jenkins, Watt (1968) Autocorrelation function???
Journel AG (1974) Simulation conditionelle de gisements miniers-theorie et pratique. These de Doctor Ingenieur, Universite de Noncy
Kazi A, Sen Z (1985) 1Volumetric RQD; an index of rock quality. In: Proceedings of the international symposium on fundamentals of Rock Joints Bjorkliden pp 99102
Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwaterstrand. J Chem Metall Min Soc S Afr 52:119–139
Krumbein WC (1970) Geological models in transition to geostatistics, in geostatistics. Plenum Press, New York, pp 143–161
Kulhawy FH (1978) Geomechanical model for rock foundation settlement. J Geotech Eng Div ASCE 104(GT2, Proc Paper 13547):211–228
Long JCS (1983) Investigation of equivalent porous medium permeability in networks of discontinuous fractures. Ph. D. Thesis, University of California, Berkeley
Louis C, Pernot M (1972) Three-dimensional investigation of flow conditions of Grand Maison Dam Site. In: Proceedings of the symposium on international society for Rock Mechanics, Percolation through fissured rocks
Matern B (1960) Spatial variation. Medd Statons Skogsforsknings institut 49:144 p
Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266
Matheron G (1965) Les variables regionalises et leur estimation. Masson et Cie, Paris, 306 p
Minoura, Coley (1971) Thin section in spatial simulation
Mood AM (1940) The distribution theory of runs. Ann Math Stat 11:427–432
Otaibi A (1990) Geotechnical investigation and engineering geological maps of Al-Nagabah area. Unpublished Thesis, Faculty of Earth Sciences, King Abdulaziz University, Jeddah
Piteau DR (1970) Analysis of the genesis and characteristics of jointing in the Nchanga open pit for the purpose of ultimately assessing the slope stability. Report, Nchanga Consolidated Copper Mines Ltd
Priest SD, Hudson J (1976) Discontinuity spacing in rock. Int J Rock Mech Min Sci Geomech Abstr 13:135–148
Priest SD, Hudson JA (1981) Estimation of discontinuity spacing and trace length using scanline surveys. Int J Rock Mech Min Sci Geomech Abstr 18(3):183–197
Rice SO (1945) Mathematical analysis of random noise. Bell Sys Tech J 24:46–156
Roach (1968) For independent processes simulation
Roulean A, Gale JE (1985) Statistical characterization of the fracture system in the Stripa Granite, Sweden. Int J Rock Mech Min Sci Geomech Abstr 22(6):353–367
Rouleau A (1984) Statistical characterization and numerical simulation of a fracture system — application to groundwater flow in the Stripa granite. Ph.D. thesis, University of Waterloo, Ontario
Ryckes KD (1984) A routine method to evaluate the modes of failure and the stability of rock slopes. Unpublished thesis, Imperial College of Science, Technology and Medicine, 246 pp
Şen Z (1978) Autorun analysis of hydrologic time series. J Hydrol 36:75–85
Şen Z (1979a) Application of autorun test to hydrologic data. J Hydrol 42:1–7
Şen Z (1979b) Effect of periodic parameters on the autocorrelation structure of hydrologic series. Water Resour Res 15(6):1639–1642

Şen Z (1980) Statistical analysis of hydrologic critical droughts. J Hydraul Div ASCE Proc Pap 14 134, 106(HY1):99–115
Şen Z (1984) RQD models and fracture spacing. J Geotech Eng Am Soc Civ Eng 110(2):203–216
Şen Z (1985) Autorun model for synthetic flow generation. J Hydrol 81:157–170
Şen Z (1989a) Cumulative semivariogram models of regionalized variables. Int J Math Geol 21 (3):891–903
Şen Z (1989b) Comment on "The generation of multi-dimensional autoregressive series by herringbone method". Math Geol 21(2):267–268.22
Şen Z (1990a) RQP RQR and fracture spacing. Technical note. Int J Rock Mech Min Sci Geomech Abstr 22:135–137 (in print)
Şen Z (1990b) Spatial simulation of geologic variables. J Math Geol 22(2):175–188
Şen Z (1991) The profile-area method for statistical evaluation of rock mass quality. Bull Assoc Eng Geol XXVIII:351–357
Şen Z (1974) Small sample properties of stationary stochastic models and the hurst phenomenon in hydrology. Unpublished Ph.D. thesis, Imperial College of Science and Technology, University of London, 286 pp
Şen Z, Kazi A (1984) Discontinuity spacing and RQD estimates from finite length scanlines. Int J Rock Mech Min Sci Geomech Abstr 21(4):203–212
Sharp WE, Aroian LA (1985) The generation of multidimensional autoregressive series by herringbone method. Math Geol 17:67–79
Smith L, Schwartz FW (1984) An analysis on the influence of fracture geometry on mass transport in fractured media. Water Resour Res 20:1241–1252
Steffen O (1975) Recent developments in the interpretation of data from joint surveys in rock masses. 6th regional conference of Africa on oil Mechanical and Foundation, II, pp1726
Subyani A, Şen Z (1990) A comparative study between cumulative semivariogram and semivariogram for regular data. J Hydrol (in print)
Switzer P (1965) A random set process with a Markovian property. Ann Math Stat 36:1859–1863
Terzaghi KC (1946) Introduction to tunnel geology: rock defects and loads on tunnel supports. In proctor R.V. and
Terzaghi RD (1965) Sources of error in joint surveys. Geotechnique 15:237–304
Wallis PF, King MS (1980) Discontinuity spacing in a crystalline rock. Int J Rock Mech Min Sci Geomech Abstr 17:63–66
Watson (1975) Correlation slope at the origin??

Index

Z. Sen, *Spatial Modeling Principles in Earth Sciences*,
DOI 10.1007/978-3-319-41758-5