TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

DISSERTATION

Scale and stream network structure in geostatistical hydrological analyses

(Geostatistische hydrologische Analysen unter Berücksichtigung von Skalenaspekten und Gewässernetzstruktur)

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

A.o. Univ. Prof. Dipl. Ing. Dr. techn. Günter Blöschl

E222

Institut für Wasserbau und Ingenieurhydrologie

eingereicht an der Technischen Universität Wien

Fakultät für Bauingenieurwesen

von Jon Olav Skøien, Cand.Scient.

0427750

Landstraßer Gürtel 29/2, 1030 Wien

Wien, im November 2006

Kurzfassung

Für viele hydrologischen Aufgabenstellungen, wie etwa Risikoanalysen, die Bestimmung von Eingangsgrößen in hydrologische Modelle, und die Erstellung von Karten hydrologischer Größen, ist es notwendig, flächendeckend die räumliche Verteilung hydrologischer Variablen zu kennen. Zu diesen Variablen zählen der Niederschlag, Bodenfeuchte, Grundwasserstände, und die hydraulische Durchlässigkeit sowie der Abfluss aus Einzugsgebieten. Meist sind nur Punktmessungen bzw. räumlich aggregierte Messungen vorhanden und es ist deshalb notwendig, diese Variablen an Stellen ohne Messungen zu berechnen. Geostatistische Methoden stellen ein interessantes Instrumentarium dar, solche Variablen zu analysieren und zu bestimmen. Diese Analysen werden durch die Fehler in die Korrelationsstruktur beeinflusst, die in der Regel unbekannt sind. Auch wurde die Gewässernetzstruktur bislang in den vorliegenden Methoden kaum berücksichtigt. Ziel dieser Dissertation ist es, die systematischen und zufälligen Fehler der geostatistischen Schätzung zufolge der räumlichen Anordnung (Distanz zwischen den Messungen, Größe des gesamten abgedeckten Bereichs und Fläche einer Messung) zu bestimmen, und geostatistische Methoden zu entwickeln, die es erlauben, hydrologische Variablen auf dem Gewässernetz an Stellen ohne Daten zu schätzen.

Die erste Zielsetzung wird in den Kapiteln 2-4 behandelt. Zufallsfelder werden generiert, aus denen Stichproben gezogen werden, um die räumliche Korrelationsstruktur zu schätzen und die Skaleneffekte auf die geschätzte Korrelationsstruktur zu bestimmen. Die Ergebnisse geben die Größe der systematischen und zufälligen Fehler als Funktion der räumlichen Anordnung an und geben Hinweise für die Interpretation der Daten und zur Korrektur der systematischen Fehler. Diese Ergebnisse werden in Kapitel 4 für die Analyse der charakteristischen Skalen von Niederschlag, Abfluss, Grundwasser und Bodenfeuchte in einer geostatistischen Variogrammanalyse verwendet.

Die zweite Zielsetzung wird in den Kapiteln 5-7 ausgearbeitet. In den Kapiteln 6 und 7 wird eine als topologisches Kriging oder Top-kriging bezeichnete neue Methode entwickelt, die es ermöglicht, mit dem Abfluss in Zusammenhang stehende Kenngrößen für ein Gewässernetz zu schätzen. In Kapitel 6 wird Top-kriging verwendet, um die 100 jährlichen Hochwasserspenden in Österreich zu bestimmen einschließlich deren Unsicherheit. Sowohl die Schätzwerte der Spenden als auch die Schätzwerte der Unsicherheit sind genauer als vergleichbare Ergebnisse, die mit dem konventionellen Ordinary Kriging Ansatz erzielt werden können. In Kapitel 7 wird Top-kriging erweitert unter Verwendung der Ergebnisse

aus Kapitel 5 und einem einfachen Wellenablaufmodell, um Zeitreihen des Abflusses mit hoher zeitlicher Auflösung für Einzugsgebiete ohne Abflussmessungen zu berechnen. Die mit Top-kriging erzielten Ergebnisse sind wesentlich genauer als die Ergebnisse des konventionellen Niederschlag-Abflussmodellansatzes, jeweils ausgedrückt durch die Modelleffizienzen.

Abstract

Spatially continuous information about hydrological variables is needed for a range of purposes in water resources management, such as risk assessment, estimating inputs for hydrological models and producing maps of hydrological variables. These variables include precipitation, soil moisture, groundwater levels, hydraulic conductivity and runoff. Usually, only point measurements or spatially aggregated measurements are available, and it is necessary to estimate the variables at locations where no observations have been made. Geostatistical methods are attractive tools both for analysing and estimating these variables. However, results are potentially influenced by errors in the correlation structure used in geostatistics. Also, the potential of incorporating the stream network organisation into geostatistical methods has not been fully exploited so far.

The objectives of this thesis are to address both issues; to identify biases and uncertainties of geostatistical estimates due to the spatial arrangement of the measurements (the distance between measurements, the size of the domain of interest, the area of a measurement); and to develop geostatistical methods for estimating hydrological variables along a stream network at locations where no data are available.

The first objective is dealt with in Chapters 2-4. Random fields are synthetically generated and the spatial correlation structure is estimated from samples from these random fields to examine the effects of the spatial arrangement of the measurements on estimates of the correlation structure. The results quantify the biases and uncertainties of the estimated correlation structure as a function of the spatial arrangement of the measurements, and assist in interpreting data and correcting for biases. The results are used in Chapter 4 for examining characteristic scales of precipitation, runoff, groundwater and soil moisture in a geostatistical variogram analysis.

The second objective is dealt with in Chapters 5-7. In Chapters 6 and 7, a novel method referred to as topological kriging or Top-kriging is developed for estimating stream flow related variables. In Chapter 6, Top-kriging is used to estimate the specific 100-year floods in Austria including their uncertainty. Both the estimates and the uncertainty estimates of Top-kriging are more accurate than those obtained by the traditional ordinary kriging method. In Chapter 7, Top-kriging is extended, using results from Chapter 5 and a simple routing model, to estimate high resolution runoff time series for ungauged catchments. The Top-kriging results are much better than those from the deterministic rainfall-runoff modelling approach in terms of model efficiencies.

Acknowledgements

Many people have given me guidance, help and tips during the time I have spent working with my PhD. First of all, this thesis would have never appeared in its present form, and maybe not at all, if it had not been for the motivation, support and help from my supervisor Prof. Günter Blöschl. He is an encouraging person to work with. He gives freedom to pursue own ideas, but also lets me know when it is time to let the ideas evolve into results. He has been willing to share an invaluable amount of time throughout all stages of my work. I will also thank Prof. Lars Gottschalk, who started to analyse correlation of runoff long time before he gave me my first hydrology lesson, who has shared of his experience in the field, who was my supervisor during my graduate studies, and from whom I learned a lot of what has developed into this thesis.

The work presented in this thesis was done during the five years that I have spent at the Institute for Hydraulic and Water Resources Engineering at the Vienna University of Technology. I thank all present and former colleagues of the institute, led by Prof. Dieter Gutknecht, for the good atmosphere of which I as a foreigner could benefit from both professionally and socially. A special thank goes to Ralf Merz, Duro Parajka and Christian Reszler, who have helped me in many ways during these years, such as providing me with data, solving problems and as discussion partners.

I was invited for a short period to Cemagref in Lyon, France, where I had fruitful discussions with Irina Krasowskaia, Etienne Leblois and Eric Sauquet, in addition to Lars Gottschalk. I also had constructive talks with Chuck Kroll, David Rupp and Ross Woods during a short stay at NIWA in Christchurch, New Zealand.

Gregor Laaha has also given me constructive feedback, particularly for Chapter 6. Andrew W. Western helped with data and comments for Chapter 4.

This thesis would not have been possible without a huge data set that has been made available by the Austrian Hydrographic Service (HZB) and the Austrian Central Institute of Meteorology and Geodynamics (ZAMG). Financial support for my work was provided by the Austrian Academy of Science, project HÖ 18.

The wholehearted support from my parents has always been a huge encouragement for me.

Finally, I want to thank Veronica Altieri for the love, support and encouragement she gives me.

Table of contents

Chapter 1. Introduction

Hydrology deals with a large range of processes in different tasks. Precipitation, soil moisture, groundwater and runoff are important factors, for example, in analysing land use impacts, for estimating hydropower production, for estimating flood risks, for drinking water purposes and in addressing ecological problems. Knowledge about the state of these processes has to be inferred from measurements, which are usually point measurements (e.g. precipitation gauges, TDR-instruments, groundwater levels) or aggregated values (e.g. radar and satellite images, pumping tests, stream gauges). However, for many purposes (e.g. maps and input variables for hydrological models), it is necessary to have spatially continuous information about the variables. The variable has hence to be estimated at locations where no data are available. The problem gets more complex if the variable is to be estimated in a spatio-temporal domain.

Geostatistical methods are attractive both for such estimation purposes, and for interpreting measurements. The assumption behind geostatistics is that the process of interest can be seen as a random field with a correlation structure. The kriging estimator is recognised as the best linear unbiased estimator (BLUE), best meaning that the mean squared error is a minimum, linear meaning that the estimate is a weighted mean of the data in the area, and unbiased meaning that the mean expected error is zero (Journel and Huijbregts, 1978, p. 304). Still, the use of geostatistical methods for analysing and modelling along stream networks has been limited. A proper use of these methods can in some cases be a solution to the Problem of Ungauged Basins (PUB). This is one of the central problems in hydrology (Sivapalan et al., 2003), how to obtain estimates for catchments without observations.

The traditional approach is to use deterministic rainfall-runoff models with parameters estimated from catchment properties or inferred from neighbouring, gauged catchments. This is the only way if the aim of the estimation is forecasting, or if it is otherwise necessary to know the causal relationships between precipitation and runoff. However, for many applications, it may not be necessary to invoke causal relationships directly. For example, when assessing the hydropower potential one is interested in the flow duration curves of ungauged sites (e.g. Castellarin et al., 2004) and there is usually no need to vary rainfall characteristics. Another application is environmental flows where one is interested in the

runoff dynamics and seasonality at ungauged sites without the need to examine scenarios (Gippel, 2005). For these applications, an alternative to the traditional rainfall-runoff modelling approach may be to directly estimate the runoff time series from observed time series of neighbouring catchments without recourse to rainfall data. Whereas precipitation has short spatial and temporal correlation lengths, and where measurements contain large uncertainties, runoff exhibits longer correlation lengths and represents aggregated values that can be observed with a much higher certainty.

An obvious choice of estimating runoff time series directly from observed runoff of neighbouring catchments are hence geostatistical methods. The scientific discipline of hydrology is not new to geostatistics. Many examples exist (e.g. Daviau et al., 2000; Adamowski and Bocci, 2001; Eaton et al., 2002; Merz and Blöschl, 2005). For runoff characteristics, the use has been limited to regionalisation of catchment parameters (e.g. Merz and Blöschl, 2005, Parajka et al., 2005). An exception is the work by Gottschalk and co-workers (Gottschalk, 1993a, b, Sauquet et al., 2000) who were the first to suggest that runoff was a spatially correlated random process that could be analysed and modelled through proper use of geostatistical methods. Their focus was on the use of a geostatistical distance as a measure of correlation between observations. This is an approximation of the true correlation structure, but at the same time computationally quicker than the use of the real correlation structure.

Estimation of stream flow characteristics is firstly a scale problem, secondly the topology of the stream flow networks has to be taken into account. The effect of measurement scale has been observed by several researchers (Gelhar, 1993, Blöschl, 1999), and a number of studies described the effects of measurement scales (Russo and Jury, 1987, Western and Blöschl, 1999) but a coherent analysis of the scale effects has been lacking in the literature. A thorough examination of measurement scale effects on estimates of the correlation structure was therefore essential to solve the problem of making estimates along a stream network. It was also necessary to study how the correlation structure changes when we move from precipitation to groundwater, soil moisture and runoff. The results of these analyses does also have applications beyond what they are used for in this thesis, e.g. in helping researchers optimising their sampling network, in choosing the right method for estimating correlation structure for their problem, and in interpretation of their results.

The aim of this thesis is hence to analyse the variability and scale effects in hydrology based on the geostatistical approach, both from a general point of view, starting with analyses of

synthetic random fields, then geostatistically analysing a series of hydrologic variables, and finally analysing and estimating in the more specific case of stream flow variables. In other words the aims of the thesis are to

- Examine the impact of scale on geostatistical analyses in hydrology
- Analyse, using geostatistical methods, what are typical characteristic space and time scales in hydrology, and their relationship to each other.
- Develop methods of representing variability that take into account the hydrologic characteristics (e.g. for up/downscaling)
- Develop methods for interpolating along stream networks (e.g. for regionalisation and PUB)

The thesis addresses these issues in the following way.

Chapter 2 and 3 deal with difficulties in estimating the correlation structure of a process. Because of incomplete information of the hydrologic variables of interest, estimates of the statistical characteristics of the variables will be associated with both uncertainties and biases. This has to some degree been studied earlier by Russo and Jury (1987) and Western and Blöschl (1999), but a coherent analysis as in Chapter 2 and 3 has not been published earlier. The chapters take particularly interest in the errors introduced by the spatial arrangement of the measurements. The underlying assumption is the simplest case of spatially correlated random fields, i.e. two dimensional fields of a variable that is normally distributed and exhibits an exponential variogram. Sampling scale effects will then be examined as they result from the finite spacing, extent and support of measurements – the scale triplet of Blöschl and Sivapalan (1995). The support, for example, can be interpreted as the catchment size for the case of runoff. Chapter 2 include non-parametric analyses where biases and random errors of the mean, variance and (spatial) integral scale of a random variable such as runoff and soil moisture have been estimated. Chapter 3 include parametric analyses where the biases and random errors of the correlation length, the sill and the nugget of an exponential variogram have been estimated.

Chapter 4 presents analyses of a set of hydrological variables in the context of characteristic scales. Characteristic scales have rarely been used in hydrology, even though their merits are undisputed (Dooge, 1986, Seyfried and Wilcox, 1995, Blöschl, 1999). Using the variogram as the tool for describing the correlation structure both in space and time, it is possible to derive characteristic scales both in space and time for the processes. Of the scaling effects discussed

in Chapter 2 and 3 particularly the support effect is pronounced in these analyses. The chapter also suggests that space and time are linked for many processes, and that it is possible to describe the joint effect of space and time through a characteristic velocity. The filtering effect of catchments is suggested in this chapter.

Chapter 5 continues the analyses of precipitation and runoff as random fields, but the variables will be treated as outcomes of spatio-temporal processes. Variance in space and time will be treated jointly, as opposed to independent spatial and temporal analyses as in Chapter 4, and the catchments will be seen as spatio-temporal filters. The assumption is that the measurement of instantaneous runoff is the convolution of the runoff generation process within the catchment, and within a certain time interval, similar to Woods and Sivapalan (1999). As a consequence, it is possible to back-calculate spatio-temporal point variograms of the processes. The analyses are done on data with higher temporal resolution than in Chapter 4.

Chapter 6 presents the topological kriging approach (called Top-kriging) for interpolation along stream networks for hydrological variables that can be approximated as a spatial aggregation process. The approach is to some degree a modification of the model suggested by Sauquet et al. (2000), but it is a more general model than what Sauquet et al. (2000) suggested. Austrian flood data are used as an example for the application of the model.

Chapter 7 presents an extension of the Top-kriging approach from Chapter 6, using the spatio-temporal filtering assumptions from Chapter 5. Methods are developed for the estimation of runoff hydrographs in ungauged catchments, taking both the spatial and the temporal aggregation into effect and taking advantage of the spatio-temporal correlations of runoff. Various spatio-temporal estimation methods are tested. This chapter goes beyond the work of Gottschalk et al. (Gottschalk, 1993a, b, Sauquet et al., 2000) in several ways. Firstly, the correlations between catchments are not described by a geostatistical distance, but by the use of a point variogram. Secondly, the methods are more focused on estimating non-conservative variables, and on the estimation of high resolution runoff time series.

Chapter 2 and 3 are on the fundamentals of scale effects and hence use hypothetical random fields. The remaining chapters use a comprehensive hydrological data set in Austria to illustrate the methods developed and how they can be used to make better hydrological predictions in ungauged catchments than has been possible before.

Chapter 2. Sampling scale effects in random fields and implications for environmental monitoring

2.1 Introduction

Scientists in numerous environmental disciplines are interested in the spatial distribution of a variable in the landscape at a given point in time. The variable of interest can be sampled at a number of locations and subsequently analysed to estimate those statistical characteristics of the variable across the landscape that are deemed relevant in a particular context. The three statistical characteristics most relevant to spatial distributions are the mean, the spatial variance and the integral scale of the variable of interest. The mean is a measure of the average behaviour within a domain. The spatial variance is a measure of how different the variables are in space. The integral scale is a measure of the average distance over which a variable is correlated in space. Small integral scales indicate that the variable varies erratically over short distances while large integral scales indicate that the variable varies smoothly over short distances and variability starts to get significant at larger distances. This chapter focuses on these three statistical characteristics.

The mean, the spatial variance and the integral scale are of interest in a number of disciplines within the environmental sciences to represent water quality parameters in estuaries (Caeiro et al., 2003); characteristics of biological communities (Wallace et al., 2000, Nunan et al., 2002, Franklin and Mills, 2003, Jurado-Expósito et al., 2004), radon concentration in soils (Oliver and Khayrat, 2001), hydraulic conductivity in aquifers (Gelhar, 1993, p. 292), and soil moisture characteristics of landscapes (Mohanty et al., 2000, Western et al., 2004), among many other examples. In these studies the three statistical characteristics are used for a range of purposes, as parameters for understanding of processes involving hypothesis testing, for quantifying a resource, for compliance, for assisting management decisions and for risk assessment.

All of these studies have in common that it is rarely possible to sample the variable of interest exhaustively and with infinite detail. There will always be some constraints imposed by the measurement procedure and the costs associated with the sampling. For example, soil moisture

can be sampled by time domain reflectometry *in situ* probes that, typically, provide an estimate of soil moisture in the top 30 cm of the soil over a measurement area of 1 dm^2 (Western and Blöschl, 1999). Alternatively, remotely sensed data can be used to provide estimates of surface soil moisture over an area of, say, 1 ha or more, depending on the sensor that is used (Entin et al., 2000). Similar measurement areas apply to most environmental measurements. In micro meteorological studies, for example, the effective fetch is the area over which the sensors on a micrometeorological tower provide information. It changes with the meteorological conditions and can be on the order of hundreds of metres in length (Schmid, 2002). Another example are groundwater pumping tests that integrate information over the depression cone. Here, the measurement area is on the order of 1 ha, depending on the subsurface characteristics and the withdrawal rate (Anderson, 1997). All of these examples relate to the sampling scale that is variably termed grain, footprint or support depending on the discipline (Blöschl and Sivapalan, 1995, Bierkens et al., 2000).

There are two other scales that are important in the spatial arrangement of environmental measurements. These are the spacing and the extent. The spacing is the characteristic distance of two sample locations, i.e., it is small if the samples are narrowly spaced and large if only a few samples are taken over a given domain. Soil hydraulic conductivity measurements in dedicated studies may have spacings of meters (Zehe and Blöschl, 2004) while rain gauges in a region are typically spaced at tens of kilometres. Water quality and air quality samples, often, have larger spacings. This is because measurement costs of small spacings are usually prohibitive, so in many practical cases, the number of samples over a domain is a function of the budget available rather than based on scientific analyses of what would be needed to sample the variable of interest accurately. The extent is the overall size of the domain sampled. It may range from centimetres to hundreds of kilometres, depending on the application. The extent is often selected as the area of interest but sometimes it can be much smaller.

The three scales, spacing, extent and support, have been termed the scale triplet by Blöschl and Sivapalan (1995). In the two-dimensional case, as is of importance in environmental monitoring in landscapes, Blöschl (1999) suggested to define them as characteristic length scales: spacing (L_S), extent (L_E), and support (L_A) (Fig. 2.1).The spacing can be related to the size of the domain, A_{dom}, and the number of samples N, the extent to the size of the domain, and the support can be related to the support area A.

$$L_S = \sqrt{\frac{A_{dom}}{N}}$$

$$L_E = \sqrt{A_{dom}} \qquad (2.1)$$

$$L_A = \sqrt{A}$$

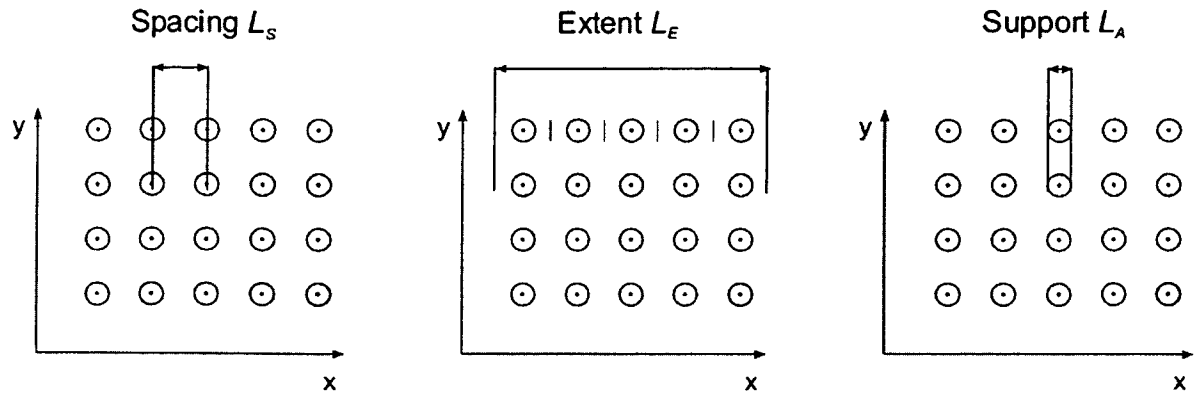


Figure 2.1 The sampling scale triplet for the two dimensional spatial case: spacing L_S , extent L_E and support L_A. Modified from Blöschl and Sivapalan (1995).

When estimating the mean, the spatial variance and the integral scale of the variable of interest, the spatial sampling scales will have some bearing on the results. The estimates from a finite number of samples, collected over a finite support area will be different from the true statistical characteristics of the variable in the landscape one is interested in because of a number of reasons. If the spacing of the samples is large relative to the underlying variability, one will not sample the small scale variability. If the extent of the samples is small relative to the underlying variability, one will, conversely, not sample the large scale variability. Finally, if the support of the samples is large relative to the underlying variability, the data will appear much smoother than the true spatial distribution of the variable of interest. These effects will result in bias (or systematic error) and uncertainty (or random error) in the estimates of the mean, the spatial variance and the integral scale.

It is not always clear that estimates may have been biased due to sampling scale effects. Franklin and Mills (2003), for example, compared a number of studies of similar microbial communities at different scales. The authors of the original studies had suggested that different factors may have contributed to structuring those communities. Based on a discussion of Rahel

(1990), Franklin and Mills (2003) however suspected that these differences were more likely to reflect viewpoints of different scales rather than differences in the way communities are organised.

While different spatial sampling strategies have attracted considerable interest in the literature (e.g. Morrissey et al., 1995; Thompson, 2002; Webster and Oliver, 2001; McDonald, 2003), research on quantifying sampling scale effects on the statistical characteristics estimated from the data has been more fragmented. The main contribution probably has been from geostatistics which emerged in the 1950s focussing, among other things, on how the spatial variance of a variable (such as the ore grade) changes with the support area or sampling volume (see e.g., Journel and Huijbregts, 1978). The methods to estimate the change in variance amount to filtering the underlying distribution by a linear filter which reduces the variance and changes the spatial correlation structure. There have been numerous applications throughout the environmental sciences from soil assessment to the estimation of extreme rainfall for catchments (e.g. Blöschl and Sivapalan, 1995, Webster and Oliver, 2001) among many other examples.

The role of the spacing of the monitoring sites has attracted less attention in the environmental sciences. Russo and Jury (1987) examined the effects of spacing on estimates of the three statistical characteristics based on Monte Carlo simulations with synthetically generated random fields. They found that large spacings will lead to overestimates (i.e. positive biases) of the integral scale and large random errors. The bias was small as long as the spacing was smaller than half the correlation length. This can be related to the Nyquist frequency in time series analysis which is the highest frequency of a signal that can be resolved by a given sampling rate (Nyquist, 1924). In a somewhat related study, Beckie (1996) analysed sampling scales associated with the resolution of a single groundwater field test and used filter functions to estimate how the unresolved small scale variability of, say, hydraulic conductivity, changes as a function of the spacing and the support of the measurements. He found that increasing spacing increases the unresolved small scale variability, while increasing support decreases it. This is because also the overall variability decreases as an effect of increasing support.

The effect of the extent of the data on the integral scale of hydraulic conductivity has been discussed by Gelhar (1993). Figure 6.5 in Gelhar (1993) suggests that the spatial scale over which hydraulic conductivity is correlated (i.e. the integral scale) tends to be on the order of 10% of the extent of the domain, irrespective of the type of underlying aquifer. This is disturbing as it would imply that the integral scales estimated from the data are mainly

controlled by the sampling setup rather than by the distribution of the variable one is meaning to capture. The findings of Gelhar (1993) have been corroborated by a number of other studies such as Blöschl (1999) for the case of snow cover patterns. Di Federico and Neuman (1997) and Cintoli et al. (2005) suggested that the concept of truncated power variograms can be used to explain these sampling effects.

One of the important concepts common to the analysis of sampling scale effects is that sampling involves some sort of filtering, and observed properties may hence be heavily biased by the sampling scales (Cushman, 1984, 1987). While a number of aspects of sampling scale effects have been examined in the literature in various contexts, a coherent analysis for landscapes (i.e. two dimensional fields) is still lacking. This is the purpose of this chapter. Specifically, the aim of this chapter is to examine what are the biases and the random errors (or uncertainties) introduced by the sampling scale triplet (spacing, extent and support) into estimates of the mean, the spatial variance and the spatial integral scale of a variable in a landscape. The chapter builds on an initial analysis of Western and Blöschl (1999) who used real soil moisture data to examine sampling scale effects. In contrast to Western and Blöschl (1999), however, the data in this chapter consist of a large number of synthetically generated random fields from which samples at various sampling scales are taken. This allows us to analyse the biases and random errors of the mean, the variance and the integral scale in a more controlled manner than is possible with real data.

2.2. Method

2.2.1 Generating random fields

The data in this chapter consist a large number of two-dimensional random fields that was generated. It was then sampled from these fields with a spatial arrangement of the samples conforming to a specified scale triplet of spacing, extent and support. The random field represents the spatial distribution of the variable of interest and each realisation is one possible outcome of the sampling. From the samples it was first calculated the mean, spatial variance and integral scale, and then examined how similar these estimates were to the mean, spatial variance and integral scale of the underlying population of the random field. The mean, spatial variance and integral scale of the underlying population are considered to be the true values and this is what one is intending to estimate to the highest possible accuracy. Figure 2.2 illustrates schematically the procedure. The top charts show the underlying true distribution

and the bottom charts show, schematically, the sampled values for large spacings, small extents and large supports.

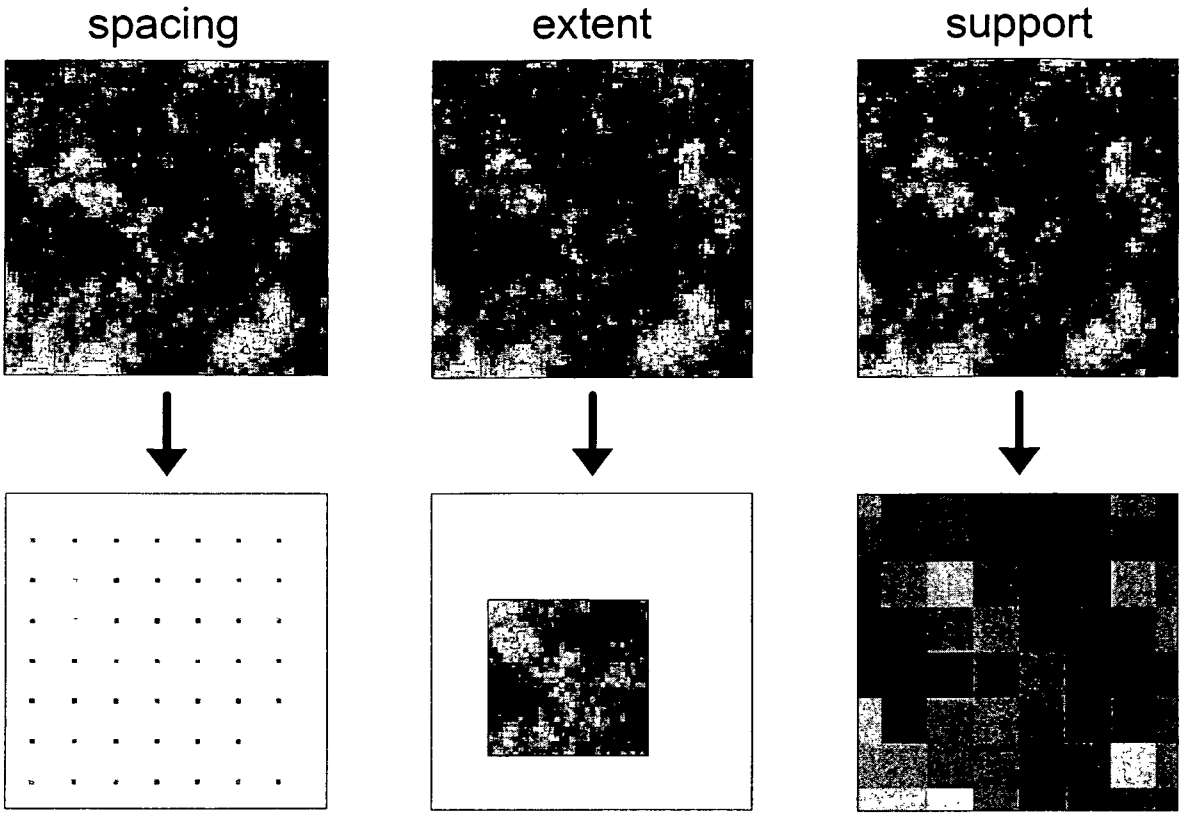


Figure 2.2 Schematic of sampling scale effects. The top row shows the underlying true distribution and the bottom row is the actual information sampled. Left: With large spacings, the small scale variability will not be sampled. Centre: With small extents, the large scale variability will not be sampled. Right: With large supports the samples will appear much smoother than the true spatial distribution.

The main assumptions in generating the random fields were (a) that the fields were stationary, i.e., the statistical characteristics of the population of the random field (such as the mean and the variance) were assumed not to change with spatial location. The second assumption was (b) that the univariate distribution of the random field conforms to a normal distribution with population mean μ and population variance σ^2. The final assumption was (c) that the spatial correlation structure of the population of the random field can be represented by an exponential variogram:

$$\gamma(h) = \sigma^2(1 - \exp(-(h/\lambda))) \qquad (2.2)$$

where λ is the correlation length and h is the distance between two points in the random field (Journel and Huijbregts, 1978). The variogram describes the average variance between two points separated by distance h. The correlation length is a measure of the average distance over which a variable is correlated in space. For the variogram in Eq. 2.2, the integral scale happens to be identical with the correlation length while for different variograms the integral scale can be smaller or larger than the correlation length, depending on the shape of the variogram. The exponential variogram is consistent with a first-order autoregressive or Markov process (Webster and Oliver, 2001) and is hence the simplest assumption one can make about the spatial variability of a random field. With these prescribed statistical characteristics two dimensional random fields were generated using the Turning Band Method (Mantoglou and Wilson, 1981, 1982) on a square grid of 1024*1024 points with grid size of Δx. For each combination of the sampling scale triplet 1000 realisations were generated.

2.2.2 Sampling

Dimensionless sampling scales L_S^*, L_E^* and L_A^* will be used throughout this chapter, i.e. the spacing, extent and support scaled by the true correlation length λ of the population of the random field:

$$L_S^* = L_S / \lambda$$

$$L_E^* = L_E / \lambda \qquad (2.3)$$

$$L_A^* = L_A / \lambda$$

In a first analysis the spacing and extent were varied jointly. From each of the 1000 realisations for each extent, 16, 100 and 1024 samples were taken. For a fixed number of samples the spacing increases linearly with the extent, similar to Eq. 2.1.

$$L_S^* = L_E^* / \sqrt{N} \qquad (2.4)$$

where N is the number of samples. For these analyses the support was set as $L_A^* = 0$, i.e. the samples from the grid were point samples.

In a second analysis the support was varied. The extent was fixed as $L_E^* = 10$, with spacing being a function of the number of samples according to Eq. 2.4. The support was changed by arithmetically averaging the point values over $(L_A/\Delta x)^2$ points. Again, 16, 100 and 1024

samples were sampled. The effect of increasing support was represented by the aggregation of an increasing number of points from the generated field for each sample. The support areas were assumed to be quadratic. Two monitoring schemes where used in both analyses, sampling on a regular square grid and random sampling.

In all instances with $L_A > L_S$, measurements will be partly overlapping and the measurements closest to the border of the domain will also include values from outside what is defined as the extent of the domain.

It was assumed that all sample values were error free, i.e. it was only analysed the biases and random errors (uncertainties) introduced by a less than exhaustive sampling of the underlying distribution. In a real world study, instrument errors will introduce additional uncertainty.

2.2.3 Estimation of moments and integral scale

From the N samples $z(\mathbf{x}_i)$ at locations $\mathbf{x}_i$ the sample mean ($\bar{z}$) and sample variance (s^2) were estimated as:

$$\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z(\mathbf{x}_i) \tag{2.5}$$

and

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N} (z(\mathbf{x}_i) - \bar{z})^2 \tag{2.6}$$

The sample variogram was estimated by the traditional estimator of Matheron (1965):

$$\hat{\gamma}(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \tag{2.7}$$

where $h = |\mathbf{h}|$ is the spatial lag between two points. The summation over the number of pairs $n(h)$ is within bins which was chosen at logarithmical intervals. The integral scale of the population is defined as (Taylor, 1921, Russo and Jury, 1987, Western and Blöschl, 1999):

$$J = \int_0^{\infty} \left[1 - \frac{\gamma(h)}{\sigma^2}\right] dh \tag{2.8}$$

The sample integral scale was estimated from the sample variogram as:

$$\hat{J} = \sum_{i=1}^{N_b} \left[1 - \frac{\hat{\gamma}(h_{i-1}) + \hat{\gamma}(h_i)}{2s^2} \right] \Delta h_i \tag{2.9}$$

where N_b is the bin number where $\bar{\gamma}(h_{N_b}) = s^2$. In other words, this is the bin number for which the sample variogram first intersects a horizontal line $\bar{\gamma}(h) = s^2$. If the variogram reached the sample variance between two bins, the exact intersection point was found by linear interpolation. It was assumed $\hat{\gamma}(h_0) = 0$ and $\Delta h_i = h_i - h_{i-1}$.

From each of the 1000 realisations for a given combination of the scale triplet, the statistical characteristics were estimated by Eqs. 2.5, 2.6 and 2.9. From the set of estimates the ensemble mean, i.e. the mean value of the 1000 estimates was estimated for each of the three statistical characteristics, in addition to the variance around the ensemble mean. The variance around the ensemble mean represents the random error or uncertainty of the estimates of Eqs. 2.5, 2.6 and 2.9 one encounters if only one sample set (i.e. one replica) is available.

The estimates in Eqs. 2.5, 2.6 and 2.9 were plotted in nondimensional form scaled by the (true) population characteristics:

$$\bar{z}^* = \frac{\bar{z} - \mu}{\sigma} \tag{2.10}$$

$$s^{2*} = \frac{s^2}{\sigma^2} \tag{2.11}$$

$$\hat{J}^* = \frac{\hat{J}}{\lambda} \tag{2.12}$$

If the ensemble mean of the nondimensional mean (Eq. 2.10) is zero, the estimates of the mean are unbiased. If the ensemble mean of the nondimensional variance (Eq. 2.11) is unity, the estimates of the variance are unbiased. Similarly, if the ensemble mean of the nondimensional integral scale (Eq. 2.12) is unity, the estimates of the integral scale are unbiased. If the variance (or standard deviation) of the sample estimates around the ensemble mean are small for any of the statistical characteristics, their uncertainty is small.

In addition, an alternative case was considered termed "single realisation case", in which the estimates were normalised differently. For each realisation the estimated sample mean and sample variance were compared independently with the mean and the variance of the entire random field of the same realisation. This is the case where the interest is in the statistical

characteristics of one realisation, not the population characteristics. The equations corresponding to Eqs. 2.10-2.12 but for the single realisation case (denoted by subscript R) are then:

$$\bar{z}_R^* = \frac{\bar{z} - \bar{z}_{dom}}{\sigma} \tag{2.13}$$

$$s_R^{2*} = \frac{s^2}{s_{dom}^2} \tag{2.14}$$

$$\hat{J}_R^* = \frac{\hat{J}}{\hat{J}_{dom}} \tag{2.15}$$

where the subscript *dom* refers to estimates based on exhaustive sampling over the domain of size A_{dom} (Eq. 2.1).

An application of this case is the estimation of the spatial average, variance and integral scale of precipitation for a certain event, or the concentration of a nutrient in a particular area at one point in time. For each realisation the realisation mean was subtracted from the estimated mean. The average of these deviations over the ensemble (according to Eq. 2.5) will show if the estimates are biased compared to the realisation mean, while the variance of these deviations (according to Eq. 2.6) will give the uncertainty of such an estimate. A similar procedure was followed for variance, except that the estimated variance was divided by the realisation variance to normalise it. The integral scale estimated from the samples was normalised by the realisation integral scale, i.e. the integral scale estimated from all points of the generated field of the same realisation.

It should be noted that, for a given sampling scale triplet, the choice of estimation methods for the statistical characteristics may affect the results to some degree. The method chosen in this chapter is one of the more commonly used methods. A comparison of methods is beyond the scope of this chapter and some guidance on the performance of different methods is given in Cressie (1991).

2.2.4 Analytical expressions for bias and random error

The sampling analyses from random fields are also compared with analytical expressions for the bias of the mean, spatial variance and integral scale as a function of the scale triplet taken from Western and Blöschl (1999). In their derivation they made a number of simplifying

assumptions which can be tested here. Among other things they suggested that it is possible to superpose the effects of the spacing, extent and support on the estimation biases. The equations used are summarised in Section 2.2.4.1. Expressions for the random error or uncertainty of the estimates are given in Section 2.2.4.2 for the mean as a function of the spacing and extent as well as for the variance for statistically independent samples.

2.2.4.1 Bias of estimates

Bias due to large support

The assumption here is that the variable aggregates linearly, i.e.

$$z_A = \frac{1}{A}\int_A Z(x)ds \tag{2.16}$$

where z_A is the average value of a random field Z within an area A where $L_A = \sqrt{A}$ is termed the support in terms of units length. The variance of the aggregated field, $\sigma^2_{L_A}$, then is the difference between the total variance, σ^2, and the variance within the support of the measurement, $\sigma^2_{WL_A}$, (Journel and Huijbregts, 1978):

$$\sigma^2_{L_A} = \sigma^2 - \sigma^2_{WL_A} \tag{2.17}$$

with

$$\sigma^2_{WL_A} = \int_0^{R\max} \gamma_T(r)\cdot f_1(r|L_A)dr \tag{2.18}$$

where γ_T denotes the true variogram. r is the distance between two randomly chosen points in the aggregated region, $R_{\max}$ is the maximum distance in the region and $f_1(r|L_A)$ is the probability density function (pdf) of distances r within the region (Rodríguez-Iturbe and Mejía, 1974). For a square region, f_1 has been derived by Ghosh (1951) as:

$$f_1(r) = \frac{4r}{L^4_{L_A}}\phi_1(r) \tag{2.19}$$

where

$$\phi_1(r) = \frac{1}{2}\pi L_A^2 - 2L_A r + \frac{1}{2}r^2 \quad \text{for } 0 \le r \le L_A \tag{2.20a}$$

$$\phi_1(r) = L_A^2\left\{\sin^{-1}\frac{L_A}{r} - \cos^{-1}\frac{L_A}{r}\right\} + 2L_A\sqrt{r^2 - L_A^2} - \frac{r^2 + 2L_A^2}{2} \quad \text{for } L_A \le h \le \sqrt{2}L_A \quad (2.20b)$$

Eq. 2.18 can then be evaluated numerically. The variogram of the averaged process can be found in a similar way (Journel and Huijbregts, 1978):

$$\gamma_{L_A}(h) = \int_0^{R\max} \gamma_T(r) \cdot f_2(r|(h, L_A))dr - \sigma^2_{L_A W} \quad (2.21)$$

where f_2 is the pdf of distances of two points, each of them is located randomly in one of two regions. The two regions are separated by a centre-to-centre distance h and have side lengths (support) L_A. For square regions and $h \ge L_A$, f_2 has been derived by Sivapalan (1986) as:

$$f_2(r) = \frac{2r}{L_A^4}\phi_2(r) \text{ where}$$

$$\phi_2(r) = L_A\sqrt{r^2 - e^2} - \frac{(r-e)^2}{2} - eL_A\cos^{-1}\left(\frac{e}{r}\right) \qquad \text{for } e \le r \sqrt{L_A^2 + e^2}$$

$$\phi_2(r) = \frac{L_A^2}{2} + e(r - \sqrt{r^2 - L_A^2}) - eL_A\sin^{-1}\left(\frac{L_A}{r}\right) \qquad \text{for } \sqrt{L_A^2 + e^2} \le r \le h$$

$$\phi_2(r) = L_A g\cos^{-1}\left(\frac{h}{r}\right) - eL_A\left[\sin^{-1}\left(\frac{L_A}{r}\right) - \cos^{-1}\left(\frac{h}{r}\right)\right]$$
$$- e\sqrt{r^2 - L_A^2} - 2L_A\sqrt{r^2 - h^2} + \frac{L_A^2}{2} + h^2 + r^2 - gr \quad \text{for } h \le r \le \sqrt{L_A^2 + h^2} \quad (2.22)$$

$$\phi_2(r) = L_A g\sin^{-1}\left(\frac{L_A}{r}\right) - L_A^2 + g\sqrt{r^2 - L_A^2} - gr \qquad \text{for } \sqrt{L_A^2 + h^2} \le r \le g$$

$$\phi_2(r) = L_A g\left[\sin^{-1}\left(\frac{L_A}{r}\right) - \cos^{-1}\left(\frac{g}{r}\right)\right] + g\sqrt{r^2 - L_A^2}$$
$$+ L_A\sqrt{r^2 - g^2} - \frac{L_A^2 + r^2 + g^2}{2} \qquad \text{for } g \le r \le \sqrt{L_A^2 + g^2}$$

with $e = h - L_A$ and $g = h + L_A$

and it can be evaluated numerically if the regions overlap (i.e. $h < L_A$). The expected integral scale can then be found by numerical integration similar to Eq. 2.6:

$$J_{L_A} = \int_0^\infty \left(1 - \frac{\gamma_{L_A}(h)}{\sigma^2_{L_A}}\right)dh \quad (2.23)$$

Bias due to small extent

The variance $\sigma^2_{L_E}$ of a random field within a domain of extent L_E can be found similar to Eq. 2.18 as:

$$\sigma^2_{L_E} = \int_0^{R\max} \gamma_T(r) \cdot f_1(r|L_E) dr \tag{2.24}$$

Western and Blöschl (1999) suggested that the expected variogram of a random field with a small extent can be assumed to be similar to the true variogram for lags that are small compared to the extent L_E, and that it will be equal to $\sigma^2_{L_E}$ for lags that are large compared to the extent with a sharp crossover at a certain lag distance B where the true variogram reaches $\sigma^2_{L_E}$, i.e.

$$\gamma_{L_E}(h) = \gamma(h) \text{ for } 0 \le h \le B \tag{2.25a}$$

$$\gamma_{L_E}(h) = \sigma^2_{L_E} \text{ for } h > B \tag{2.25b}$$

For an exponential variogram (Eq. 2.2) this gives:

$$\gamma_T(B) = \sigma^2_{L_E} \text{ and } B = -\lambda \cdot \ln\left(1 - \frac{\sigma^2_{L_E}}{\sigma^2_T}\right) \tag{2.26}$$

where λ is the true correlation length of the underlying process. This gives the expectation of the integral scale:

$$J_{L_E} = \frac{\lambda}{\sigma^2_{L_E}} \cdot \left[\sigma^2_{L_E} + \left(\sigma^2_T - \sigma^2_{L_E}\right) \cdot \ln\left(1 - \frac{\sigma^2_{L_E}}{\sigma^2_T}\right)\right] \tag{2.27}$$

Bias due to large spacing

For any large spacing L_S, the expectation of the sample variance $\sigma^2_{L_S}$ will be equal to the true variance:

$$\sigma^2_{L_S} = \sigma^2 \tag{2.28}$$

For lags larger than the average spacing, L_S, Western and Blöschl (1999) suggested that the sample variogram can be approximated by the true variogram and for shorter lags it can be assumed to be linear between the origin $\gamma(0) = 0$ and $\gamma(L_S)$, i.e.

$$\gamma_{L_S}(h) = \frac{h}{L_S}\gamma_T(L_S) \text{ for } h \le L_S \tag{2.29a}$$

$$\gamma_{L_S}(h) = \gamma_T(h) \text{ for } h > L_S \tag{2.29b}$$

The expectation of the integral scale for an exponential variogram then is:

$$J_{L_S} = \lambda\left[\exp\left(\frac{-L_S}{\lambda}\right) + \frac{L_S}{2\lambda}\cdot\left(1+\exp\left(\frac{-L_S}{\lambda}\right)\right)\right] \tag{2.30}$$

which approaches the true correlation length λ for small spacings.

2.2.4.2 Error boundaries of estimates of mean

For grid sampling, the uncertainty (in terms of the error variance Var) of the estimates of the mean $\overline{z_{L_E}}$ as a function of extent L_E is (Rodríguez-Iturbe and Mejía, 1974):

$$\begin{aligned} Var(\overline{z_{L_E}}) &= \frac{1}{N^2}E\left[\sum_{i=1}^{N} Cov(z_i, z_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(z_i, z_j)\right] \\ Var(\overline{z_{L_E}}) &= \frac{\sigma^2_{L_E}}{N^2}\left[N + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\rho(h_{ij})\right] \end{aligned} \tag{2.31}$$

where $\sigma^2_{L_E}$ is the variance of the random field within the region with extent L_E. For brevity, it is here used the covariance $Cov = \sigma^2_{L_E} - \gamma_{L_E}$ and the correlogram $\rho = Cov/\sigma^2_{L_E}$ of the random variables in points i and j with a distance h_{ij}, instead of the variogram. Eq. 2.31 can be solved numerically, given the number of samples, N, and a correlogram.

For random sampling, the distance becomes a random variable, and Rodríguez-Iturbe and Mejía (1974) replaced the distance with the expectation of distance:

$$Var(\overline{z_{L_E}}) = \frac{1}{N^2}E\left[\sum_{i=1}^{N} Cov(z_i, z_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov(z_i, z_j)\right] \tag{2.32}$$

$$Var(\overline{z_{L_E}}) = \frac{\sigma^2_{L_E}}{N^2}\left\{N + N(N-1)E\left[\rho(h_{ij}|L_E)\right]\right\} \tag{2.33}$$

$E\left[\rho(h_{ij}|L_E)\right]$ represents the expected value of the correlation between two points randomly located in a square of side length L_E and can be expressed in a similar way to Eq. 2.24:

$$E\left[\rho(h_{ij}|L_E)\right] = \int_0^{R\max}\rho(r)f_1(r|L_E)dr \tag{2.34}$$

where f_1 is taken from Eq. 2.19.

For large spacings, $\rho(h_{ij}|L_E) = 0, \quad i \neq j$. The expression then reduces to:

$$Var(\overline{z_{L_E}}) = \frac{\sigma^2_{L_E}}{N} \tag{2.35}$$

which is the normal expression for the expected random error of the mean for uncorrelated samples. The uncertainty of the estimates of variance for uncorrelated samples can be found as:

$$Var(s^2) = \frac{2\sigma^4_{L_E}}{N} \tag{2.36}$$

2.3 Results

2.3.1 Estimation of the population characteristics

2.3.1.1 Spacing

Figure 2.3 shows the results for the case where spacing is large relative to the underlying correlation length λ. The monitoring scheme is gridded sampling and three cases of 16, 100, and 1024 samples are considered. The charts in the top row of Fig. 2.3 indicate that large spacings do no introduce biases into the mean. For large spacings relative to the underlying correlation length λ, $L_S^* > 3$, the uncertainty of the mean (shown as error bars in the figures) does not change with spacing. This is because the samples are practically uncorrelated. The expression in Eq. 2.35 for uncorrelated samples provides a close match to the uncertainties of the sampling analysis as shown in Fig. 2.3. For small spacings, the uncertainties increase with decreasing spacings which is, however, related to small extents and will be discussed in section 2.3.1.2. The uncertainty of the mean decreases with increasing number of samples in all cases as would be expected.

The estimates of the spatial variance (Fig. 2.3, second row) are unbiased at large spacings, similar to the estimates of the mean and, again, the expression in Eq. 2.36 provides a close match. The sampling properties of uncorrelated random variables are very well defined as would be expected.

The estimates of the integral scale (Fig. 2.3, bottom row) are biased as a function of spacing and are increasingly overestimated with increasing spacings. For 16 samples, there is only a very short range of spacings ($L_S^* \approx 2$) where the integral scales are unbiased. For a larger

number of samples this range increases. The uncertainties associated with the integral scales (error bars in Fig. 2.3) are relatively small which indicates that the integral scale can be estimated with good confidence but the biases are large. The expression in Eq. 2.30 closely follows the results of the sampling analysis which indicates that the scale effects of the spacing can be well predicted.

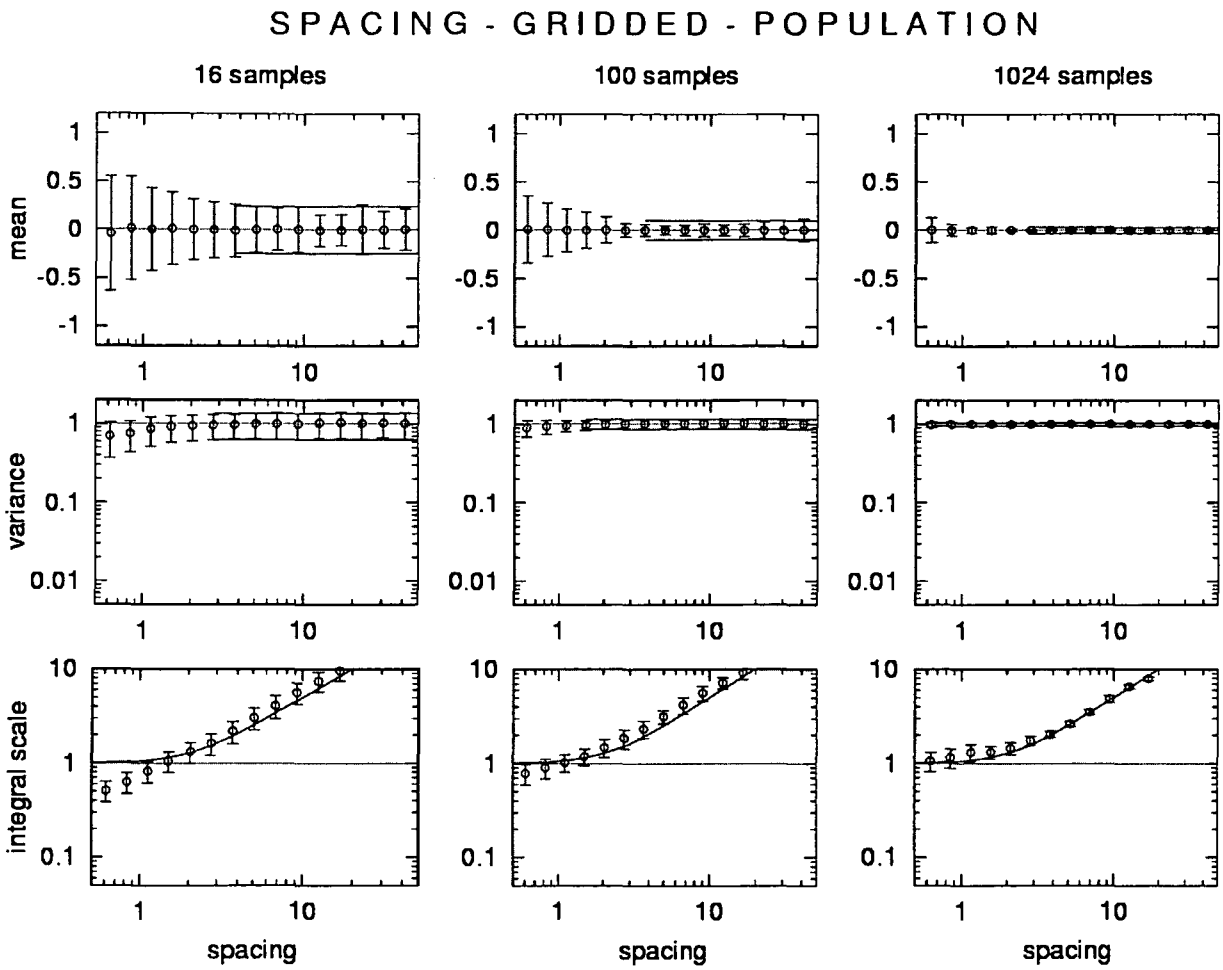


Figure 2.3 Effect of large spacings L_S^* on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for gridded sampling. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. The lines in the charts show Eqs. 2.35, 2.36 and 2.30.

Figure 2.4 shows the same information as Fig. 2.3, but here the monitoring scheme was random rather than gridded. The results for the mean and spatial variance are similar to the gridded scheme but those for the integral scales are different. The uncertainties of the

estimates are considerably larger than for gridded sampling. On the other hand the biases are much smaller. For a large number of samples ($N = 1024$) the estimates of the integral scale are almost unbiased over a wide range of spacings. As indicated in Fig. 2.4 (bottom row), Eq. 2.30 overestimates the biases. This is because Eq. 2.30 applies to gridded rather than random sampling.

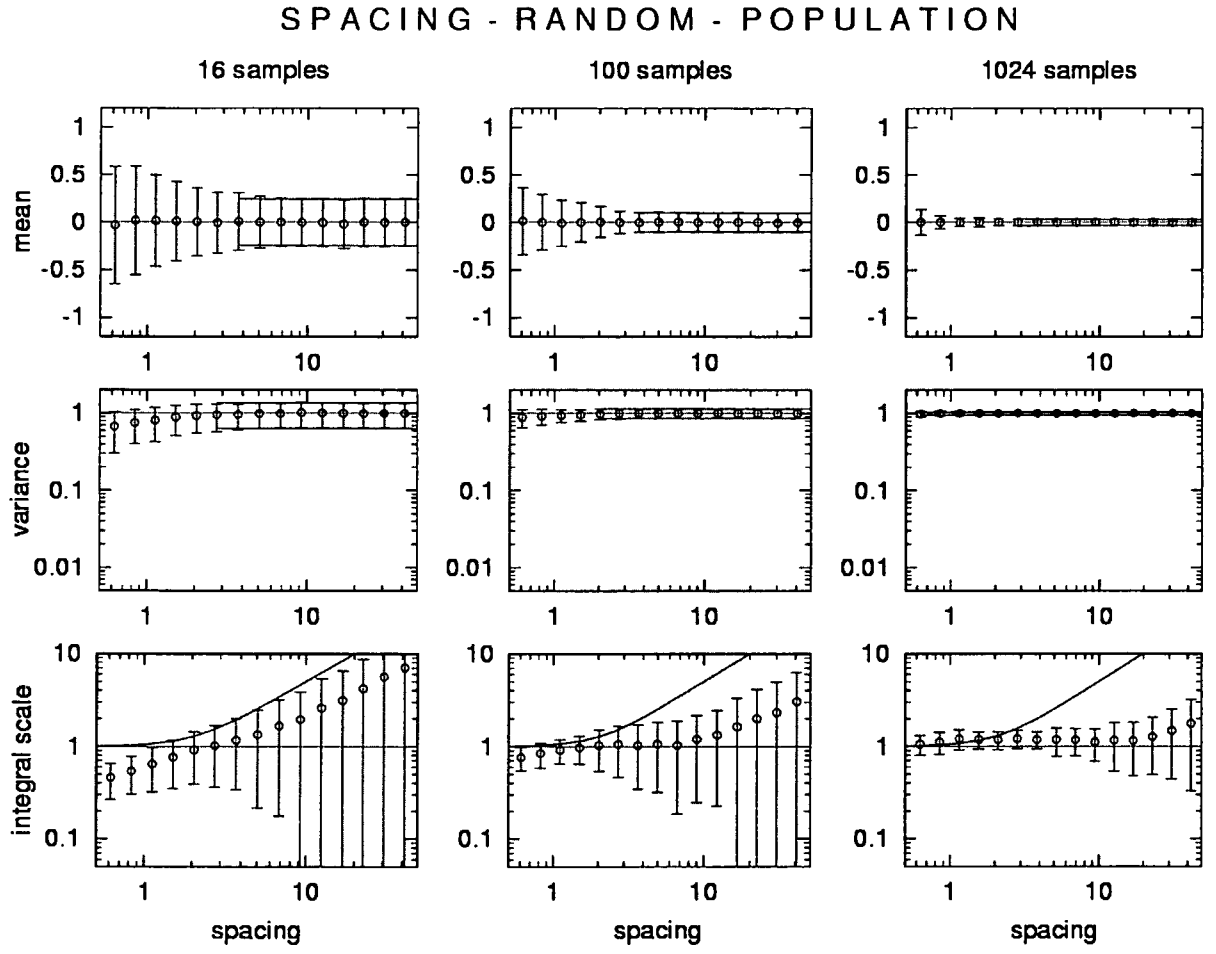


Figure 2.4 Same as Fig. 2.3 but for the case of random sampling.

2.3.1.2 Extent

Figure 2.5 shows the results for the case where the extent is small relative to the underlying correlation length λ. For comparison the graphs have been extended to large extents which imply large spacings for a given number of samples. The top row in Fig. 2.5 shows that the mean is unbiased. The uncertainty, however, increases drastically with decreasing extent. In the limit of very small extents the uncertainty is unity. This is the standard deviation of the underlying field which means no information on the underlying distribution is obtained from

such a monitoring scheme. On the other hand, the case for large extents shown in Fig. 2.5 top row implies large spacings (e.g., with 100 samples, an extent of 100 implies a spacing of 10). It is interesting that the uncertainties introduced by small extents (when only a small fraction of the landscape is sampled) are much larger than those introduced by large spacings (where in the limit the data are uncorrelated). For the case of small extents all samples are heavily correlated and very little is learned about the variability of the population. This behaviour is also borne out by the expressions in Eq. 2.33.

The second row in Fig. 2.5 shows that the variance is significantly biased for small extents. Once $L_E^* < 10$ the variance is underestimated and this bias does not depend on the number of samples. Eq. 2.24 gives a good prediction of the biases although at very small extents Eq. 2.24 slightly underestimates the biases.

The bottom row in Fig. 2.5 indicates that the integral scale is heavily biased for small extents. The underestimation can be almost two order of magnitudes. It is interesting that the underestimation occurs for $L_E^* < 10$. This means that the monitoring domain needs to be at least 10 times the underlying correlation length for not introducing extent biases. For larger extents there is some overestimation but this is a spacing effect as discussed above. For all cases, the uncertainty of the estimated integral scale is small in spite of the very large biases. The lines shown in the bottom row in Fig. 2.5 are a product of Eqs. 2.27 and 2.30 to account for extent and spacing effects simultaneously. They closely represent the sampling results.

For the case of small extents, the biases and uncertainties of the estimators are independent of the spatial arrangements of the samples within the domain. Random sampling hence produces very similar results as those in Fig. 2.5 so they are not shown here.

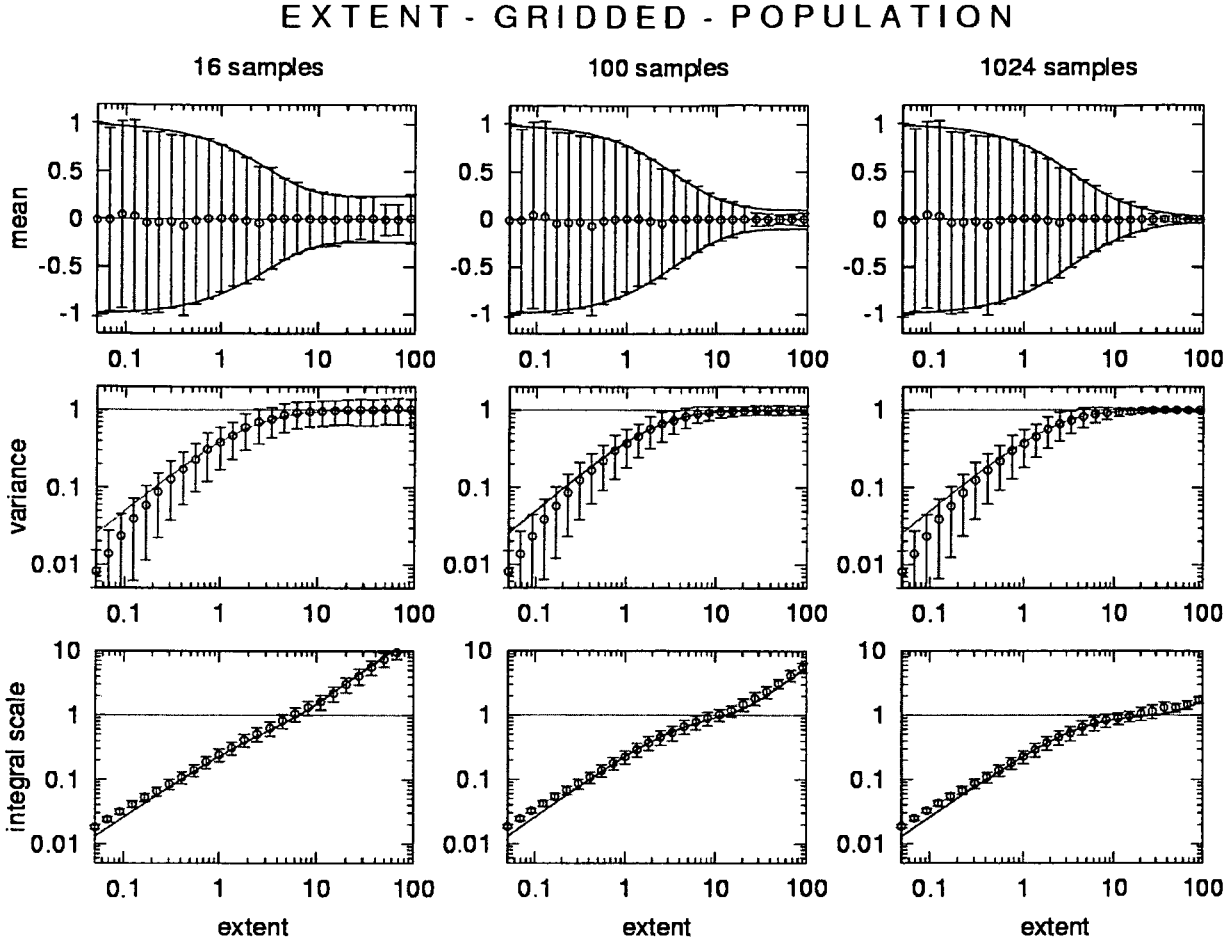


Figure 2.5 Effect of small extents L_E^* on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for gridded sampling. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. Lines in the charts show Eqs. 2.33, 2.35, 2.24, and the product of Eqs. 2.27 and 2.30.

2.3.1.3 Support

Figure 2.6 shows the results for the case where the support is large relative to the underlying correlation length λ. The first row in Fig. 2.6 indicates that the estimates of the mean are unbiased. The uncertainties of the mean do not change with the size of the support. This may seem counter intuitive as one would perhaps expect the uncertainty of the mean to decrease with increasing support. The reason for the uncertainty not to change with support is that this case examines estimation of the mean of the population rather than that of a single realisation (shown later). Most of the uncertainty comes from the variability between different realisations (what Journel and Huijbregts, 1978, p. 192, terms fluctuation variance). The uncertainty due to not exhaustively sampling a single realisation is however much smaller. A

comparison of Fig. 2.6, top row, suggests that this is indeed the case as the uncertainty does not decrease with increasing number of samples (and hence decreasing spacing). The magnitude of the uncertainty is controlled by the extent of the data chosen for these cases ($L_E^* = 10$, see Fig. 2.5).

The variances (Fig. 2.6, second row) are, however, significantly biased. They are underestimated once the support exceeds about one third of the underlying correlation length. The vertical lines in Fig. 2.6 show the support values where $L_A^* = L_S^*$, i.e. the domain is fully covered by the samples. For larger supports, the samples overlap which introduces particularly large biases. The uncertainties of the variances are not very large and there is a slight tendency for decreasing uncertainty with increasing number of samples. Eqs. 2.17 and 2.18 are close to the sampling results although they slightly underestimate the variances for large supports. This is because Eqs. 2.17 and 2.18 have been derived for non-overlapping samples.

The third row in Fig. 2.6 shows moderate biases of the integral scale as a result of large supports. Again, once the samples overlap, the integral scales are overestimated. For 16 samples, there is a spacing effect which results in an overestimation for all supports. Eq. 2.23 does not account for these spacing effects. For large supports, Eq. 2.23 overestimates the bias results of the sampling analysis. This is a border effect in the sampling analysis as the largest distance between the centres of the samples will be the size of the domain.

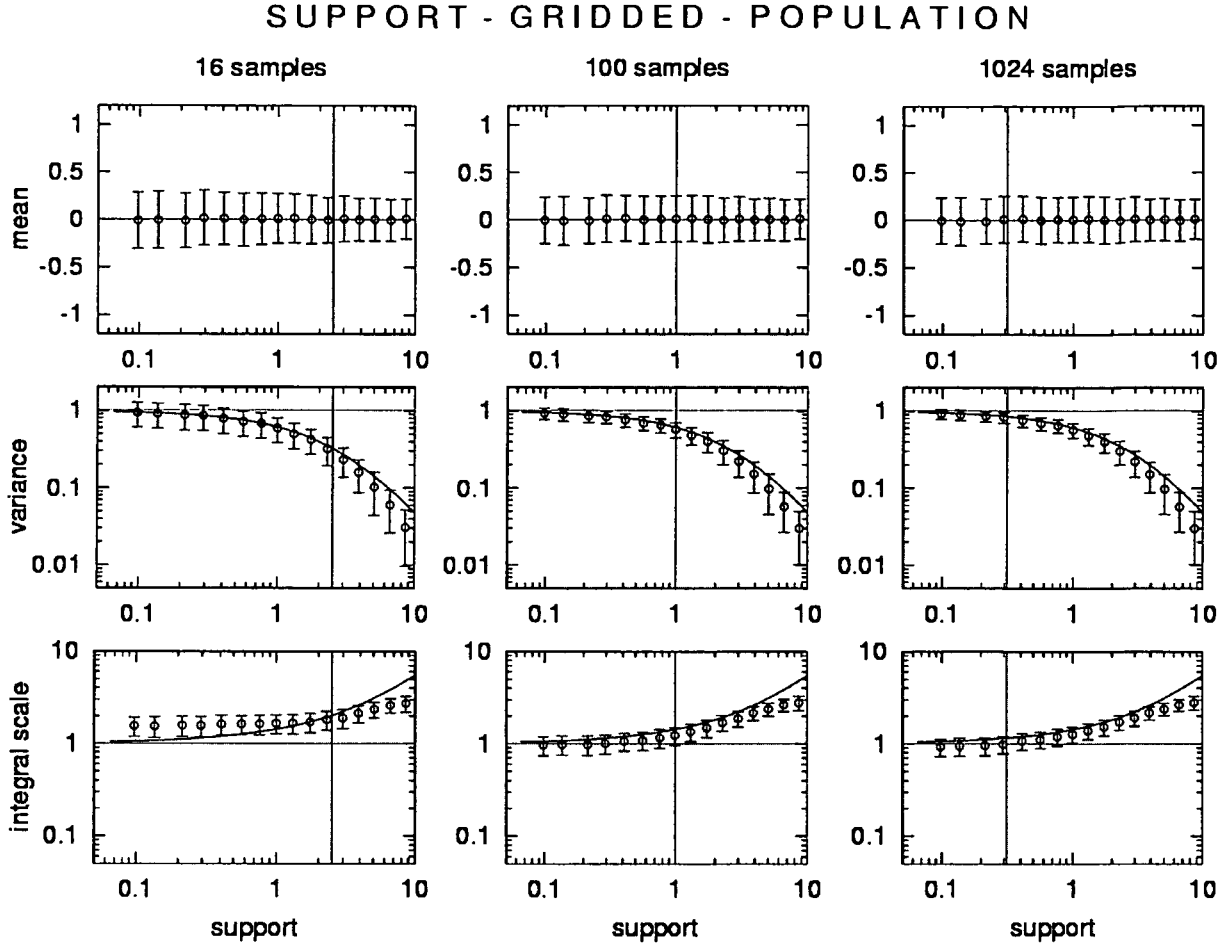


Figure 2.6 Effect of large supports L_A^* on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for gridded sampling. The extent is $L_E^* = 10$ in all cases. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. Lines in the charts show Eqs. 2.17 and 2.23. The vertical lines show $L_A^* = L_S^*$, i.e. the support where the domain is fully covered by the samples.

2.3.1.4 Summary of scale triplet effects

For comparison the results for the sampling scale effects are summarised in Fig. 2.7. It is interesting that the uncertainties in the mean are largest for small extents as compared to other scale effects. The biases in the variance are large both for small extents and large supports. The integral scale will always be biased and the bias is closely related to the magnitude of any of the scales of the scale triplet.

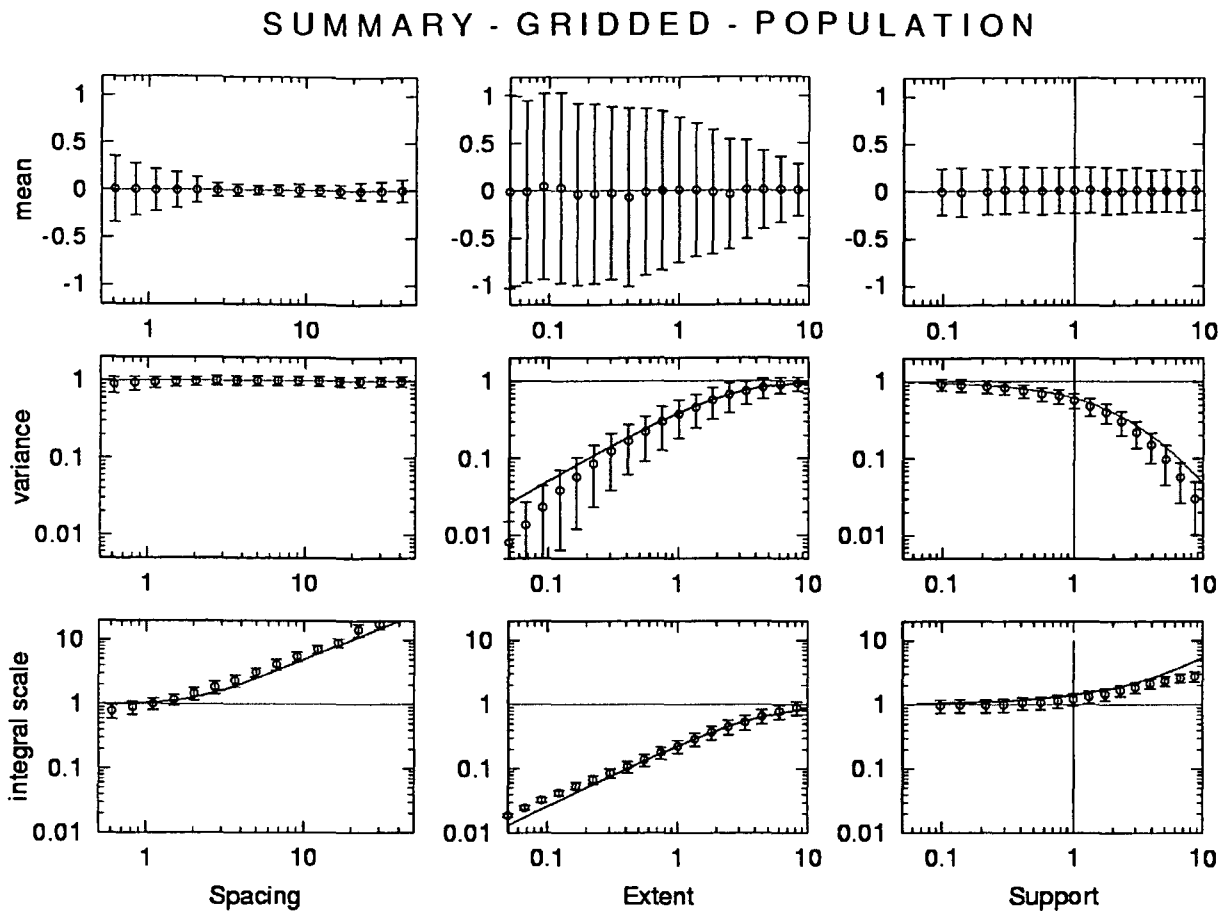


Figure 2.7 Effect of the sampling scale triplet on the sample mean $\bar{z}^*$, spatial variance s^{2*} and integral scale $\hat{J}^*$ for 100 gridded samples. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean.

2.3.2 Estimates of single realisations

2.3.2.1 Spacing

Figures 2.8 to 2.11 present the scale effects for the analyses where interest lies in the statistical characteristics of one single realisation (Eqs. 2.13 to 2.15), not the population characteristics. Figure 2.8 gives the results for the case where spacing is large relative to the underlying correlation length λ. The mean is unbiased (Fig. 2.8 top row), similarly to the population case in Fig. 2.3. The estimation uncertainty of the mean increases with the spacing which is the exact opposite of the population result in Fig. 2.3, top row. For large spacings (relative to the underlying correlation length) the samples are uncorrelated, so the estimates for the single realisation are no better than the estimates for the population (in both cases 0.25 for 16 samples, for example, as given by Eq. 2.35). On the other hand, as the spacing decreases

(relative to the underlying correlation length) the correlation of the samples increases. While this increases the estimation uncertainty of the population (Fig. 2.3) it decreases the estimation uncertainty of the realisation (Fig. 2.8). In the limit, if the samples are perfectly correlated, there will be no uncertainty introduced from missing out on the locations between the samples.

The second row in Fig. 2.8 indicates that the variance is, again, unbiased. For large spacings, the samples are uncorrelated and the results are identical to those in Fig. 2.3.

The third row in Fig. 2.8 shows that the integral scale for the single realisation case is biased in a similar way as in the population case. For large spacings, the integral scale is substantially overestimated. The bias is predicted well by Eq. 2.15.

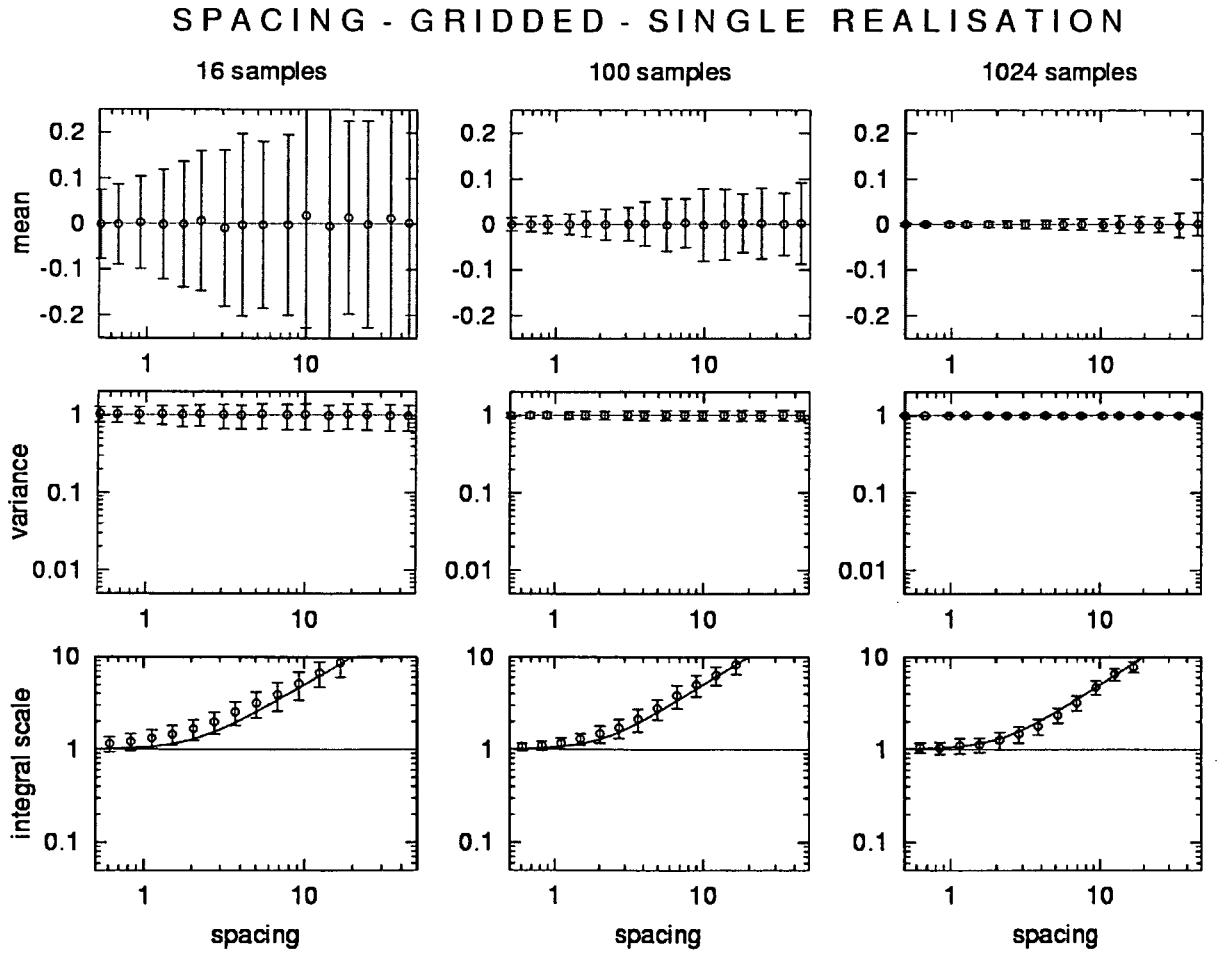


Figure 2.8 Effect of large spacings L_S^* on the sample mean $\bar{z}_R^*$, spatial variance s_R^{2*} and integral scale $\hat{J}_R^*$ for gridded sampling. This is the case where the interest resides in the statistical characteristics of a single realisation, not the population characteristics. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean. Lines in the charts for the integral scale show Eq. 2.30.

2.3.2.2 Extent

Figure 2.9 shows the results for the case where the extent is small relative to the underlying correlation length λ, again in the single realisation mode. Small extents (relative to the underlying correlation length) imply that the samples are highly correlated or, equivalently, that the variable varies very smoothly over the domain sampled. Because of this, the difference between the sampled values and the entire underlying random field (considered as the truth here) is small in all instances. If this is the case, all statistical characteristics will be essentially unbiased (as compared to each realisation) and the uncertainty is always very small which is indeed the case in Fig. 2.9 for small extents. In other words, in the limit of very small extents (relative to the underlying correlation length), everything is known about the underlying distribution (of the one realisation) from a limited number of samples. This is in stark contrast to Fig. 2.5 where in the limit of very small extents nothing is known about the underlying distribution (of the population) no matter how many samples are taken.

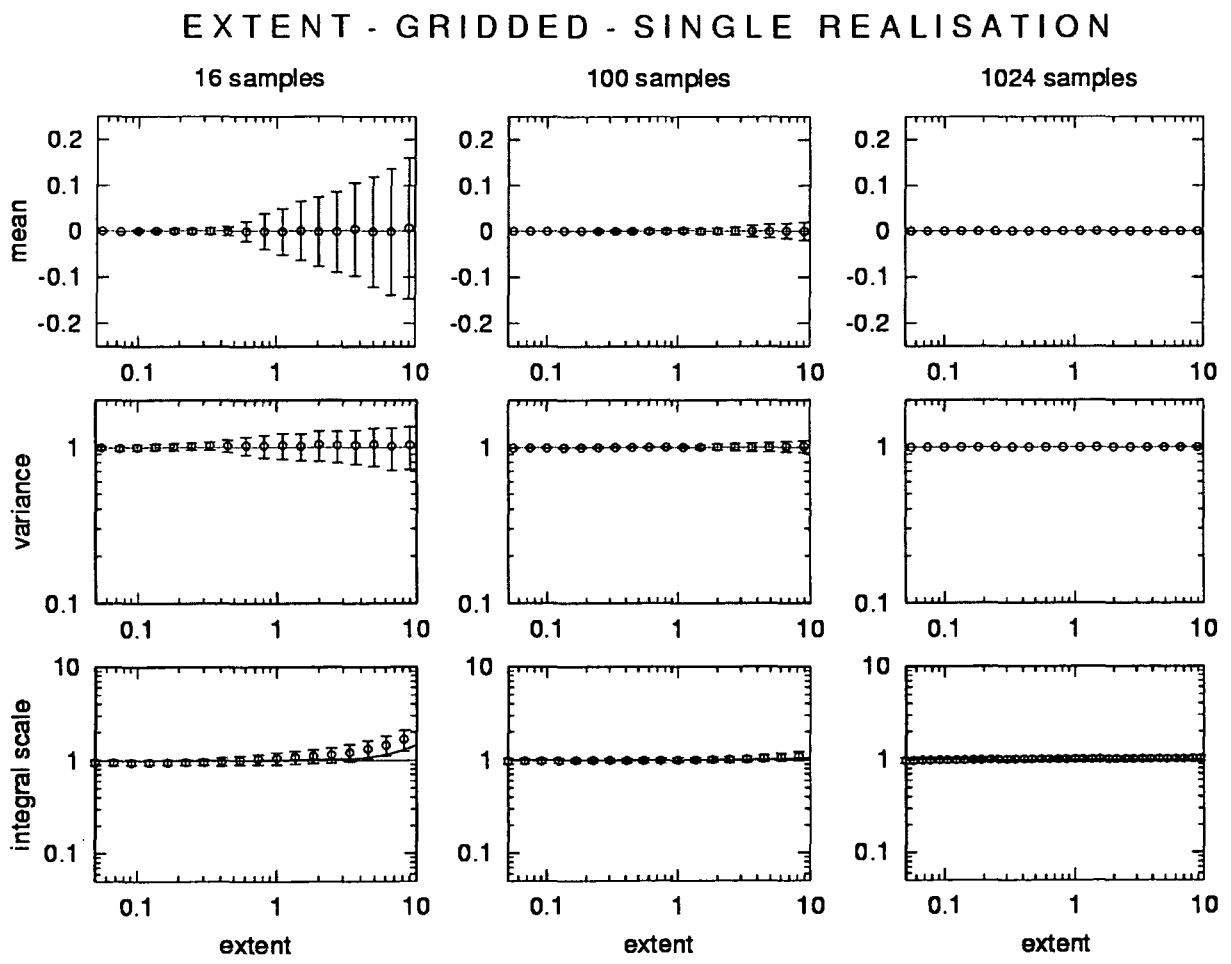


Figure 2.9 Same as Fig. 2.8 but for the case of small extents L_E^*.

2.3.2.3 Support

Figure 2.10 shows the results for the case where the support is large relative to the underlying correlation length λ. The mean is unbiased (Fig. 2.10 top row). The uncertainty of the mean decreases with the number of samples. This is a result of the spacing as the extent was held fixed to $L_E^* = 10$ in all cases, which implies smaller spacings for a larger number of samples. There is an interesting dependence of the uncertainty of the mean on the support. For small supports, the uncertainty decreases with increasing support until $L_A^* = L_S^*$. This is because an increasingly larger fraction of the domain of interest will be covered by the samples. For $L_A^* = L_S^*$ (indicated by the vertical line in Fig. 2.10) the domain is fully covered by the samples, so there is no uncertainty involved in the estimation of the mean. However, as the support increases further, the samples will overlap and the uncertainty starts to increase. This is because some parts of the domain will be more frequently sampled than others and because the part the support area will lie outside the domain A_{dom}.

The second and third rows in Fig. 2.10 show similar results as their counterparts of the population case in Fig. 2.6. The biases are essentially identical. However, there are minor differences in the uncertainties. The uncertainties of the variance in Fig. 2.10 are slightly smaller than those in Fig. 2.6. This is because of different spacing effects. On the other hand the uncertainties of the integral scale in Fig. 2.10 are slightly larger.

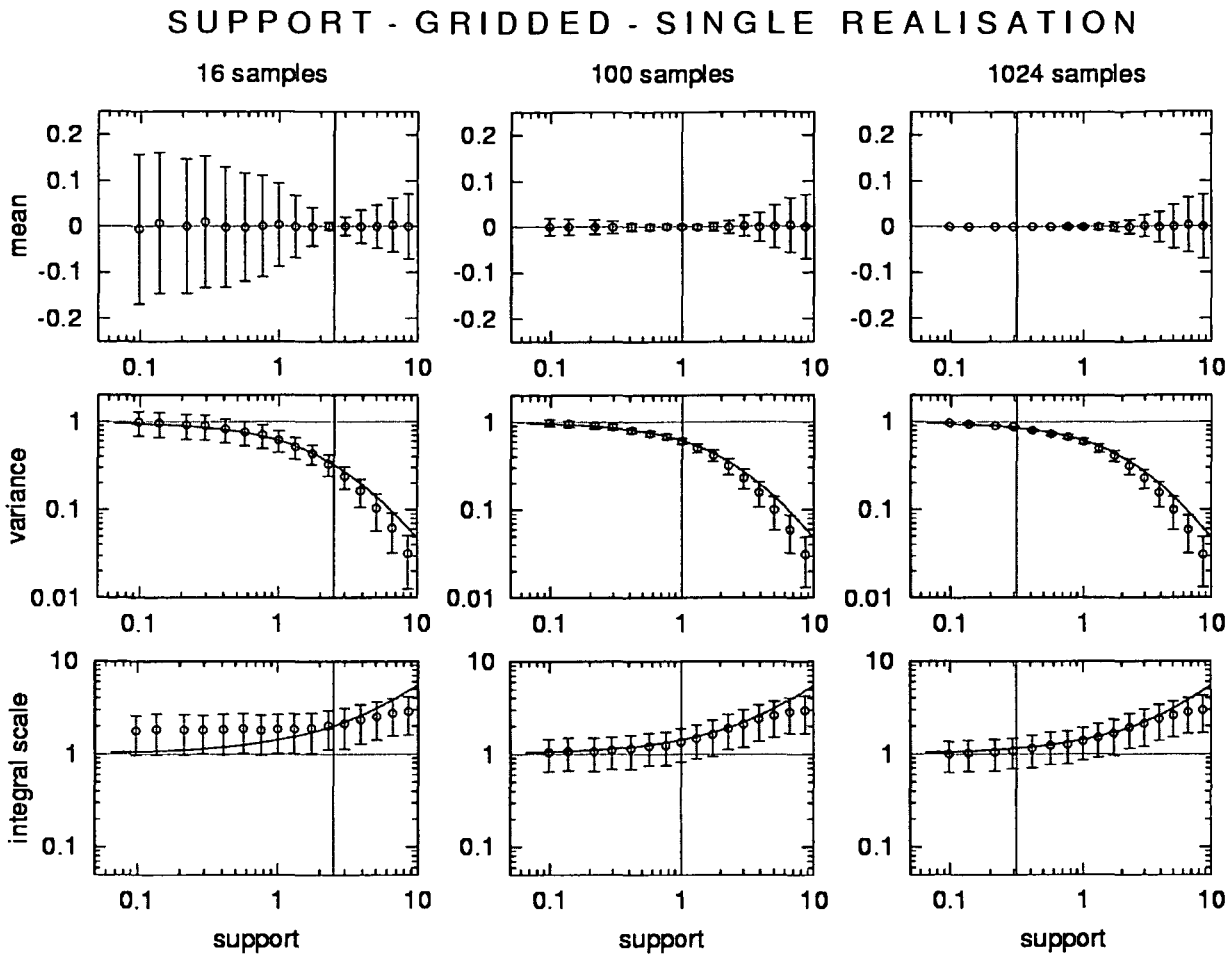


Figure 2.10 Same as Fig. 2.8 but for the case of large supports L_A^*. The extent is L_E^*= 10 in all cases. Lines in the charts show Eqs. 2.17 and 2.23. The vertical lines show $L_A^* = L_S^*$, i.e. the support where the domain is fully covered by the samples.

2.3.2.4 Summary of scale triplet effects

For comparison, the results of the sampling scale effects are summarised in Fig. 2.11 for the analyses where the interest lies in the statistical characteristics of one single realisation. The main difference to the population case in Fig. 2.7 is that small extents neither introduce biases nor uncertainties. This is simply because in the single realisation mode, it is not interesting what happens outside the domain sampled. However, in the population mode it is indeed the population, including the distribution outside the domain sampled that is of interest. The sampling effect of large spacings on the mean is related. For large spacings, the samples are uncorrelated, so yield increasingly smaller uncertainties in the populations case (Fig. 2.7) but increasingly larger uncertainties if one is interested in one single realisation only (Fig. 2.11). The spacing and support effects on both the variance and the integral scale of the single

realisation case are similar to those of the population case. The biases in the variance are large for large supports. The integral scale will be biased for large spacings and large supports and the bias is closely related to the magnitude of these two scales.

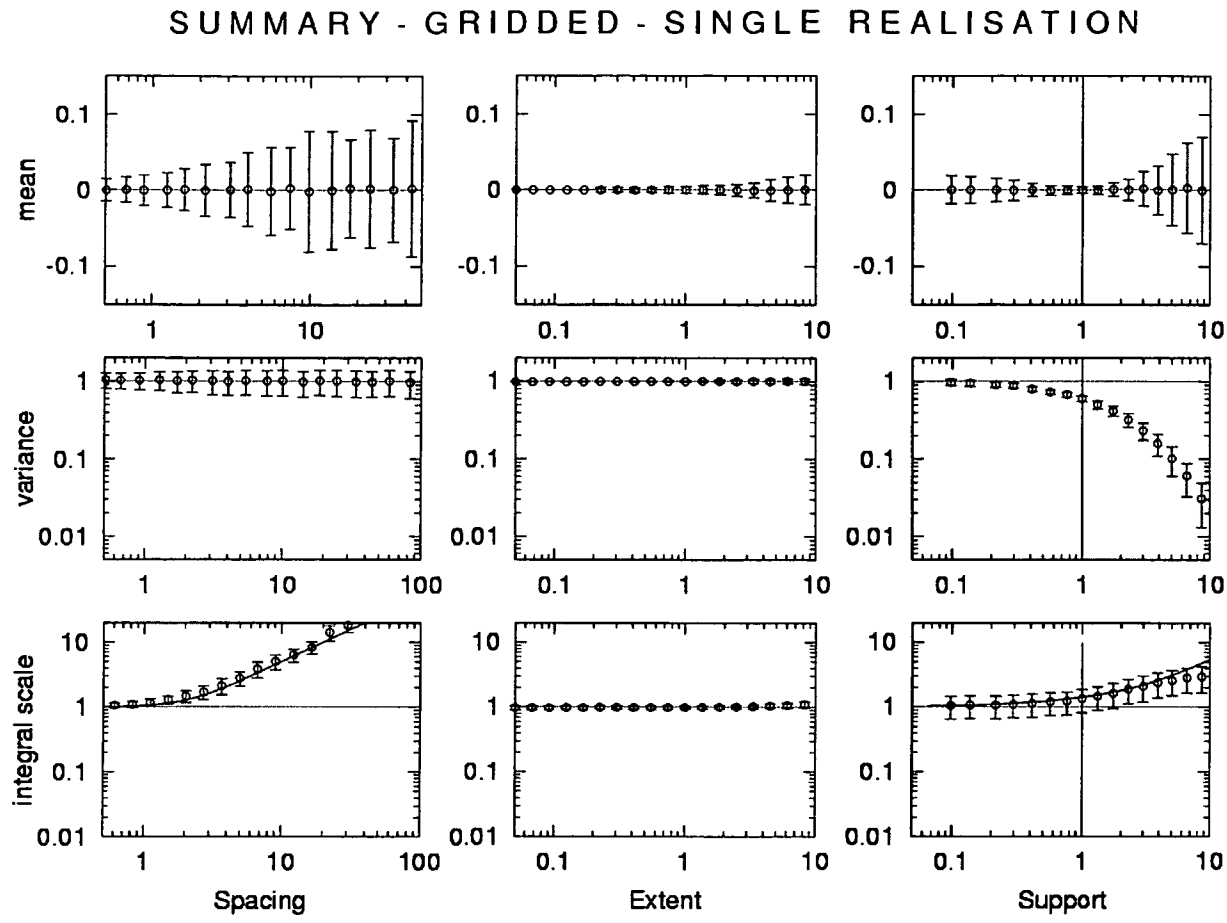


Figure 2.11 Effect of the sampling scale triplet on the sample mean $\bar{z}_R^*$, spatial variance s_R^{2*} and integral scale $\hat{J}_R^*$ for 100 gridded samples. This is the case where the interest resides in the statistical characteristics of a single realisation, not the population characteristics. The circles show the ensemble mean and the error bars represent the standard deviation around the ensemble mean.

2.4 Implications for environmental monitoring

2.4.1 Importance of sampling scale effects

The sampling analyses in this chapter have shown that there may exist enormous biases in the three statistical characteristics examined here, the mean, the spatial variance and the integral scale, which result from sampling scale effects. Even in the absence of instrument error the biases can be up to two orders of magnitude for the cases considered here. Also, the uncertainty or random error can be very significant.

The findings in this chapter are consistent with previous work. The biases are similar to those of Western and Blöschl (1999) of their analysis of soil moisture data in a catchment although this chapter analyses the scale effects in a more controlled manner. Also, this chapter gives more precise estimates of the uncertainties involved as the synthetic sampling analysis allowed multiple realisations to be drawn. Similar to Russo and Jury (1987) the results of this chapter suggest that large spacings, relative to the underlying correlation length, tend to overestimate the integral scale. Similar to Gelhar (1993) the results found here suggest that small extents of the domain result in overestimates of the integral scale. The biases in the variance due to large supports are consistent with those given in the geostatistical literature. It is interesting that the biases of the integral scale are closely related to the magnitude of any of the components of the scale triplet where in most cases large sampling scales result in an overestimation of the integral scales. Intuitively, this is because of dimensional reasons as both the integral scale and the sampling scale are related to units of length. In most cases, the expressions given in Appendix A predict the biases estimated by the sampling analysis well. In examining these sampling biases, however, there are a lot of subtleties involved.

For typical monitoring scales, the results of this chapter suggest that usually there is bias due to more than one of the components of the scale triplet. Each of the components produces a scale effect of its own which, for some statistical characteristics, are similar (such as the integral scale) but for others are different, and the resulting bias is the combination of all three effects. This mixing of the effects of spacing, extent and support is not always obvious. Qi and Wu (1996), for example, examined the effect of support on three spatial autocorrelation indices, i.e., the Moran Coefficient, the Geary Ratio and the Cliff-Ord statistic based on topographic and biomass data of the Malaysian Peninsular. They observed an overall decline in spatial autocorrelation with increasing support but this was more likely an effect of the spacing of adjacent units increasing at the same time. In a number of monitoring systems, such as remote sensing data, changing (or selecting) grid size usually involves a change of both

spacing and support, so the grid size effect on the statistical characteristics is a combination of spacing and support effects. The comparison of the synthetic resampling analysis with the analytical results suggested that it is indeed possible to superpose the effects of spacing, extent and support to good accuracy (e.g. Fig. 2.5) to get an overall assessment of the total magnitude of the biases involved.

The results in this chapter showed that the sampling scale effects can be very important but it is not only the sampling scale triplet that controls the biases and uncertainties it is also whether one is interested in population statistics or the statistics of one single realisation. The former case is where one is interested in the mean value of, say, concentration over a very large area, i.e. the population mean. In this case it is the small extent that introduces the largest uncertainties into estimates of the mean. For the case of small extents all samples are heavily correlated and very little is learned about the variability of the population. Gelhar (1993, p. 40) and Priestley (1981, p. 319) describe how the number of effective independent samples will be reduced when the samples are correlated. The uncertainties decrease with increasing spacing as the samples become less correlated and hence contain more information about the population. This means that the estimates are quite accurate inside the domain but if one extrapolates further away (in the case of small extents) the results will be highly uncertain as the data represent only a small fraction of the entire population. On the other hand, the second case, termed single realisation case in this chapter, is where one is interested in the characteristics (e.g. the mean) of, say, precipitation for a certain event, or the concentration of a nutrient in a particular area at one point in time where the extent is the same as the size of the domain of interest. In this case very little uncertainty is introduced by extent effects and the spacing effect on the mean is the exact opposite of that in the population case. The uncertainties increase with increasing spacing as the underlying pattern becomes less smooth and hence more erratic. Because of the magnitude of the sampling scale effects there are important ramifications for sampling design and data analysis.

2.4.2 Implications for sampling design

The most direct implication probably is the importance of adjusting the sampling design to the scale of the underlying processes one is intending to capture. Stommel (1963) compares the process of sampling to a fish net and notes that a single net does not catch fish of all sizes. Quite obviously, if one is interested in small scale processes the sampling scales need to be adjusted to smaller scales, and the opposite is true for large scale processes. Informally, the

monitoring network is often adjusted to the scales of the variables one is interested in (e.g. Hatcher et al., 1987, Seyfried and Wilcox, 1995) but it is also possible to formalise this. The variable to gauge the scale of the underlying (true) variability used here is the correlation length. The correlation length is closely related to the integral scale. For an exponential variogram as has been used here, the integral scale and the correlation length are identical. For different variograms they differ but are usually of the same order of magnitude. In this chapter, it was possible to normalise the scale effects by the correlation length, i.e. all scale effects were a function of the scale of each component of the scale triplet (spacing, extent and support) relative to the underlying correlation length. If one uses the correlation length as the main measure of the spatial variability of the underlying processes one has to design the spatial arrangement of the monitoring network relative to the correlation length. A sampling design that is commensurate with the scale of the underlying variability would then involve the following questions:

- What are the processes that are to be monitored?

- What are the spatial scales of these processes (measured by the integral scale)?

- What are the spacing, extent and support of the monitoring network that give small biases and uncertainties for the statistical characteristics one is interested in, given the constraints on costs and hence number of samples?

As a rule of thumb, the extent should be significantly larger than the correlation length (for the population case), the spacing should be smaller than the correlation length and the choice of the support depends on the statistic one is interested in. Large supports are an advantage for the single realisation case but produce significant biases in the population case.

While these steps appear logical, to some degree, this is a Catch-22 situation. One needs to know the integral scale to estimate the correlation length for the sampling design but would need the sampling results to estimate the integral scale. To break the loop one can, however, use the integral scales from other studies one expects to exhibit similar spatial variabilities, particularly, studies where a large number of samples are collected and hence are spatially representative.

In the planning of the monitoring network it is not only the average spacing (or number of sampling points over a given area) that is important but also the spatial sampling arrangement (see e.g. Thompson, 2002). Depending on whether grid sampling or random sampling schemes are selected, the distribution of distances between pairs of points is different. Random sampling includes distances significantly smaller than the (average) spacing, so the small scale

variability can be better resolved. The biases associated with larger spacings in estimating the integral scale are hence smaller (Fig. 2.4). However, this is at the cost of larger uncertainties. Nested sampling (Webster and Oliver, 2001), as an alternative, allows for still smaller distances between points, so one would expect still smaller biases, but it hinges on the assumption that the variability is spatially stationary, i.e., the small area that is sampled in more detail is representative of a larger area that is sampled more sparsely. Bellehumeur and Legendre (1998) discuss the issues of stationarity in nested sampling and Stenger et al. (2002) provide an example of the impacts of non-stationarity with nested sampling. From a statistical perspective, stationarity can be tested for but the issue is not a purely statistical one. Ultimately it is the environmental processes one wants to capture and they need to be representative for nested sampling to make sense.

A similar issue of representativeness applies to the case of small extents. Small extents cause large uncertainties in all statistical characteristics examined here. If one seeks to increase the generality of the results of a study, one therefore needs to increase the extent as pointed out by Hewitt et al. (1998). On the other hand, it is not wise to increase the extent beyond what can be assumed to be the same process. Haugen (1978) describes how the certainty of estimated moments of turbulence motion increases with the temporal extent of the measurements but will be affected by the diurnal cycle if the extent is too long. Wiens (1989) examined the temporal and spatial extents of measurements at the same time, and argued that it is easy to be overly confident in the sampling results as lack of representativeness can cause serious biases. In the case the interest resides in a single realisation (nutrients over one area at one point in time, for example) sampling design is usually such that the sampling spans the entire domain, i.e. the extent is equal to the size of the domain of interest which, in the light of the results of this chapter, obviously is a prudent choice.

In selecting the support of a monitoring scheme, larger supports may be an advantage if one is interested in estimating the mean for the single realisation case. The support depends on the choice of instrument. Remote sensing sensors usually have rather large supports while in situ samples are rather small. There are, however, a large number of other factors that need to be considered when selecting an instrument, both technical (such as accuracy) and logistical, so the sampling support may not be the main consideration in choosing an instrument. A similar effect as large supports can be obtained by taking multiple point measurements over a small area and then averaging them which filters out fine-scale heterogeneity one may not be interested in (Root and Schneider, 1995). This may also filter out some of the measurement error, which however is beyond the scope of this chapter.

The number of samples is, of course, the most important issue in spatial sampling design. The costs associated with the monitoring usually are strongly positively correlated with the number of samples and the uncertainties of the estimates are strongly negatively correlated so there are tradeoffs involved (see e.g. Faures et al., 1995, Lopes, 1996). The results of this chapter suggest that for the realisation mean, spatial variance and integral scale, an increasing number of samples will in most cases reduce the estimation uncertainty. The exception is the case of small extents, relative to the underlying variability, where the biases and uncertainties do not depend on how many samples are taken within the extent of the sampling domain. Uncertainties in extrapolating beyond the sampling domain cannot be compensated by a larger number of samples within the domain.

2.4.3 Implications for data analysis and interpretation

As the sampling scales are likely to affect the measurements of a variable, even if the sampling design has been selected commensurate with the underlying variability, it can be difficult to compare measurements from different sources with each other. Theocharopoulos et al. (2001), for example, in a comparison of soil pollution sampling guidelines within the European Union pointed out that differences in the guidelines of individual countries caused problems in the objective definition of pollution, in the reproduction of the sampling by different teams, and in defining threshold values throughout the European Union. In a similar vein, Western et al. (1998) reviewed field studies of soil moisture around the world and concluded that the spatial statistical characteristics of soil moisture found in these studies were more closely related to the sampling scales than to the environmental conditions of the particular catchments.

It is, however, possible to account for these differences, and correct for sampling scale biases to some extent as pointed out in this chapter. To illustrate the case, let us consider a hypothetical example where bias is introduced into estimates of the integral scale by large sample spacings. Let us assume that 100 samples have been collected from a field of 1000 m by 1000 m extent, i.e. the average spacing is 100 m, and that the integral scale was estimated from the data as 80 m. The equation for the expected integral scale as a function of spacing (Eq. 2.30) cannot be analytically inverted, but an estimate of the true correlation length can be found by trial and error. If a true correlation length of 55 m is assumed, the dimensionless spacing is 1.8. Eq 2.30 then gives a dimensionless integral scale of 1.5 (i.e. a bias of 50%). This is consistent with its overestimation due to large spacings (80 rather than 55 m) so the initial guess was correct. The biases can be back-calculated in a similar way for the other

sampling scales. Let us consider an example of biases due to large supports where hydraulic conductivity has been sampled by groundwater pumping tests (support of 100 m). The samples gave a variance of the logarithms of hydraulic conductivity of $\sigma^2_{\ln K} = 1.2$ and an integral scale of $J = 88$ m but one is interested in the variance and the integral scale of the point process (i.e. if conductivity had been measured by laboratory core tests with a support of 10 cm). Again by iteration, it can be assumed that the (true) point integral scale is 40 m. The dimensionless support then is 2.5. Eq. 2.23 gives a dimensionless integral scale of 2.2 (i.e. a bias of 120%) which is consistent with the observed integral scale of 88 m. From Eq. 2.17 the dimensionless variance is 0.32, so the point variance is 3.75 rather than 1.2 what was observed for the large support.

While this back-calculation is useful to assess potential biases, in many cases, the uncertainty with this will be large. This is particularly true of the integral scale if a small number of samples is available. Figure 2.5, for example, shows that, if only 16 samples are available, the estimated integral scale becomes a function of extent only and is essentially independent of the underlying correlation length. It is on the order of 10 % of the extent of the domain. This is exactly the observation of Gelhar (1993, p. 292) and Blöschl (1999) in comparisons of case studies of aquifer hydraulic conductivity and snow cover variability, respectively, and implies that the integral scale is essentially unobservable if only a small number of samples is available. As the number of samples increases, the integral scale gets better defined. Also, random sampling decreases the bias over grid sampling. It is clear that a large number of samples is a prerequisite for reliably identifying the integral scales if they are to be used in either sampling design or data interpretation.

It should be noted that in this study a number of simplifying assumptions have been made. Most importantly, the random fields were assumed to be stationary with one single scale, the integral scale, representing the spatial variability. While, often, this is a reasonable assumption for a certain scale range, there are also examples where the underlying variability possesses more than one scale or a suite of nested scales (see .e.g. McBratney, 1992 for a soils example). In these cases a power law variogram may be a better choice than an exponential variogram. This is then termed fractal variability. There is a large body of literature on fractal random fields in the environmental sciences (see e,g, Chilés and Delfiner, 1999) and more research needs to be done on the sampling scale effects of this type of processes. Qualitatively, one would assume that the biases for fractal processes are similar to those found here (see .e.g. Blöschl, 1999), but the uncertainties most likely will be different.

Another assumption made here was that the instrument errors were assumed negligible. It is clear that, as the instrument errors increase, so will the uncertainties in the estimates of the statistical characteristics.

It has been taken a simple statistical approach in this chapter that intends to capture the first order sampling scale effects on the mean, the spatial variance and the integral scale of an environmental variable. There are applications where the interest resides in different characteristics of a variable such as trends or patterns. While for these characteristics the scale effects will be different, the analysis in this chapter can give some guidance on assessing their order of magnitude.

Chapter 3. Scale effects in estimating the variogram and implications for soil hydrology

3.1 Introduction

This chapter focuses on the use of geostatistics for characterizing spatial properties in the landscape, with a particular focus on vadose zone related variables. This variable can be soil moisture, concentrations of nutrients, pollutants or bacteria, or dynamic properties related to any of these, such as hydraulic conductivity or denitrification rates. As in the previous chapter, the variable of interest can be considered a realization of a random field.

The analyses have similarities with the analyses in the previous chapter, but whereas the previous chapter was concerned with non-parametric estimates, this chapter deals with parametric estimates of the variogram. The sill of the variogram is a measure of the magnitude of the spatial variability, the correlation length is a measure of the degree of correlation in space (i.e. the scale of the underlying process), and the nugget is a measure of small-scale variability.

Similar to Chapter 2, the focus lies on the scale triplet by Blöschl and Sivapalan (1995). In the two-dimensional case, as is of importance for environmental variables in landscapes, Blöschl (1999) defined them as characteristic length scales: spacing (L_S), extent (L_E), and support (L_A) (Fig. 3.1). When measuring, e.g., soil moisture in an area of size A_{dom}, the spacing (L_S) is the average distance between the centres of each measurement and is a function of the size of the domain, A_{dom}, and the number of samples N. The extent (L_E) reflects the physical boundaries of the domain of measurements, and the support (L_A) is related to the support area A of the measurement, e.g. the surface area of a soil sample or the footprint of a pixel from remote sensing. In the case of a square, these relations are (as described in the previous chapter):

$$L_S = \sqrt{\frac{A_{dom}}{N}} \tag{3.1}$$

$$L_E = \sqrt{A_{dom}} \tag{3.2}$$

$$L_A = \sqrt{A} \tag{3.3}$$

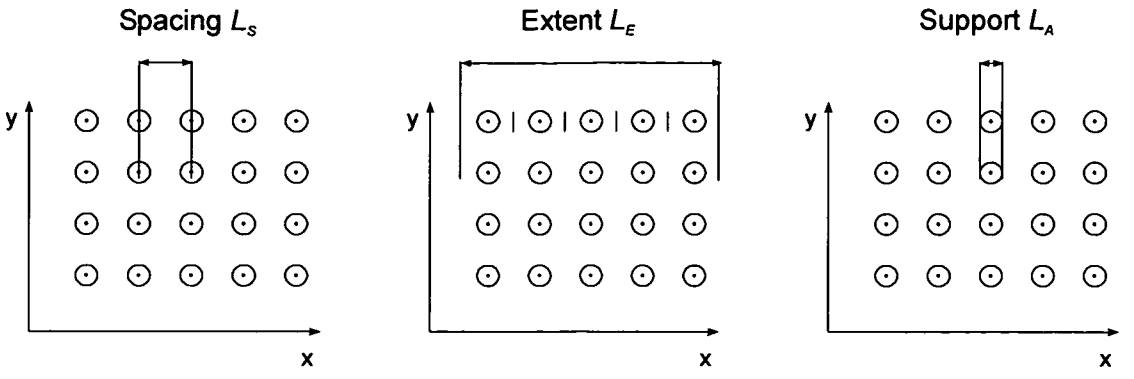


Figure 3.1 The sampling scale triplet, spacing L_S, extent L_E, and support L_A, for the two-dimensional case.

When the sampling scales are not commensurate with the scale of the underlying process, they will likely affect the estimated characteristics of the data. The data may then exhibit both biases and random errors that are related to the choice of sampling scale. Chapter 2 presents analyses where the effect of sampling scale on non-parametric estimates are studied, and also gives a review of related studies.

There have been several studies that examined uncertainties or random errors in the estimated variogram parameters. Cressie (1985) suggested that the uncertainty would depend on the variogram value and the number of samples in a bin. Shafer and Varljen (1990) proposed a jack-knifing method to estimate the confidence limits. Ortiz and Deutsch (2002) and Pardo-Igúzquiza and Dowd (2001) developed analytical expressions for describing the uncertainty of the estimates, taking the correlation between sample pairs into account, based on work by Menke (1989) and Woodbury and Sudicky (1991). Webster and Oliver (1992) examined the influence of the number of samples on the estimates of the local variogram for one particular experimental set-up. They concluded that a variogram computed from 150 samples might often be satisfactory, while one derived from 225 samples will usually be reliable. Their findings were for one experimental setup, and estimation of the local variogram. In different applications the number of samples needed for reliably estimating the variogram may be different.

Sample variograms can be estimated in different ways (Matheron, 1965, Cressie and Hawkins, 1980, Omre, 1984) most of which are a variant of plotting the squared differences of the data values of pairs of points against lag and then averaging them within each lag bin. Fitting a variogram model to the sample variogram can be done by, e.g., Ordinary Least

Squares (OLS), Generalised Least Squares (GLS), Weighted Least Squares with approximations (WLS) (all from Cressie, 1991, p. 94–97), Maximum Likelihood (ML), and Restricted Maximum Likelihood (REML) (Kitanidis, 1983, Kitanidis and Lane, 1985, Pardo-Igúzquiza, 1998). These methods assume that the type of variogram model is known. Gorsich and Genton (2000) proposed a method of finding the most likely type of variogram model by estimating the nonparametric derivative of the sample variogram values.

Zimmerman and Zimmerman (1991) compared seven estimators that can be used for estimating the parameters of a variogram. They also examined the estimates as a function of the parameters of the underlying variogram to get an indication of sampling scale effects. Their results showed that the performance of all methods was similar. They suggested that the loss of accuracy by using simple methods such as OLS and WLS is small compared to the extra computational burden of using more complex methods such as ML and REML. Cressie (1991, p. 100) discussed these results and questioned the robustness of the ML and REML methods to departures from the Gaussian assumption to which they are closely tied. Pardo-Igúzquiza (1998), however, suggested that the ML method is applicable to data that do not conform to a multivariate Gaussian distribution although this may be difficult to verify in practice.

There is a distinction between the theoretical variogram, and the local variogram (Journel and Huijbregts, 1978, p. 192). The theoretical variogram represents all possible realisations of a process while the local variogram is obtained by sampling exhaustively only within the area of interest. For a description of the general correlation structure of a variable the theoretical variogram is required, while for interpolation between observations in a study area, the local variogram is required. This chapter focuses on how the sampling scale affects the estimated variogram parameters of the theoretical variogram. Specifically, the aim of the chapter is to identify the biases and random errors in variogram parameters that result from the choice of a sample scale triplet (spacing, extent, and support) that is not commensurate with the scales of the underlying "true" soil variability. The correlation structure has been chosen to be simple, assuming normally distributed values, stationarity, and absence of noise to focus on the base case. More complicated distributions are beyond the scope of this chapter. The variogram parameters are estimated by the WLS and ML methods and are compared with the variogram parameters used for generating the random fields, along with the corresponding non-parametric estimates.

3.2 Methods

3.2.1 Generating Random Fields and Sampling

The data in this chapter consist of random fields generated in the same way as in Chapter 2. A large number of random fields have been generated with a prescribed correlation structure. The sampling scale effects have then been analysed by sampling from these generated fields. This provides an opportunity to analyse repeated realisations from a homogenously distributed variable with a known correlation structure.

The random field represents the spatial distribution of the variable of interest and each realisation represents one possible spatial distribution of such a variable. The random fields are assumed to be stationary, i.e., the statistical characteristics of the population of the random field (such as the mean and the variance) were assumed not to change with spatial location. It was also assumed that the univariate distribution of the random field conforms to a normal distribution with population mean μ and population variance σ^2. The last assumption necessary is that the spatial correlation structure of the population of the random field could be represented by an exponential variogram:

$$\gamma(h) = c_0 + c_1(1 - \exp(-(h/\lambda))) \tag{3.4}$$

where λ is the correlation length, h is the spatial distance between two points in the random field, and c_0 is the nugget effect, caused by micro scale variability and measurement errors, while $c_s = c_0 + c_1$ is the sill of the variogram, equal to the variance σ^2 of the variable.

The variogram describes the variance between two points separated by the distance h. If no micro-scale variation or measurement errors are present, the nugget is zero. The correlation length is a measure of the average distance over which a variable is correlated in space. Small correlation lengths indicate that the variable varies erratically over short distances while large correlation lengths indicate that the variable varies smoothly over short distances and variability starts to get significant at larger distances. The exponential variogram is consistent with a first-order autoregressive or Markov process (Webster and Oliver, 2001, p. 116) and is hence the simplest assumption one can make about the spatial variability of random fields.

The mean, variance and correlation length of the underlying variogram are the true population characteristics. No nugget effect was assumed, and two-dimensional random fields with the prescribed statistical characteristics were generated using the Turning Band Method (TBM) (Mantoglou and Wilson, 1981, 1982) on a square grid of 1024*1024 points with grid size of

Δx. 16 bands were used for generating the random fields. 1000 realisations were generated for each combination of the sampling scale triplet. All results later in this chapter are normalised with respect to the underlying random field. This corresponds to a random field with normal distribution N(0,1) and a correlation length λ=1. The realisations closely resembled the imposed correlation structure on average although, at small distances, the fields were a little too smooth. This will have a minor effect on some of the estimates as discussed later in this chapter.

3.2.2 Sampling

Dimensionless sampling scales L_S^*, L_E^* and L_A^* will be used also throughout this chapter, i.e. the spacing, extent and support scaled by the true correlation length λ of the population of the random field:

$$L_S^* = L_S / \lambda \tag{3.5}$$

$$L_E^* = L_E / \lambda \tag{3.6}$$

$$L_A^* = L_A / \lambda \tag{3.7}$$

In a first analysis the spacing and extent were varied jointly. From each of the 1000 realisations for each extent, 16, 100 and 1024 samples were taken. For a fixed number of samples the spacing increases linearly with the extent, similar to Eq. 3.1,

$$L_S^* = L_E^* / \sqrt{N} \tag{3.8}$$

where N is the number of samples. For these analyses the support was set as L_A^*=0, i.e. the samples from the grid were point samples. In a second analysis the support was varied, fixing the extent as L_E^*=10, with spacing being a function of the number of samples according to Eq. 3.8. Again, 16, 100 and 1024 samples were sampled. The effect of increasing support was represented by the aggregation of an increasing number of points from the generated field for each sample. The support areas were assumed to be quadratic. Two monitoring schemes where used in both analyses, sampling on a regular square grid and random sampling.

In all instances with $L_A > L_S$, measurements will be partly overlapping and the measurements closest to the border of the domain will also include values from outside what is defined as

the extent of the domain. This can be the case in remote sensing, where the support of a pixel can be larger than the spacing between the pixels. It was assumed that all sample values were error free. In a real world study, instrument errors will introduce additional uncertainty.

3.2.3 Estimation of Spatial Characteristics

3.2.3.1 Non-Parametric Estimates

From the N samples $z(\mathbf{x}_i)$ at locations $\mathbf{x}_i$ the sample mean ($\bar{z}$) and sample variance (s^2) were estimated as:

$$\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z(\mathbf{x}_i) \tag{3.9}$$

and

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N} (z(\mathbf{x}_i) - \bar{z})^2 \tag{3.10}$$

The sample variogram was estimated by the traditional estimator of Matheron (1965):

$$\hat{\gamma}(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \tag{3.11}$$

where $h = |\mathbf{h}|$ is the spatial lag between two points. The summation over the number of pairs $n(h)$ is within bins, which was chosen at logarithmical intervals. The integral scale of the population is defined as (Taylor, 1921, Russo and Jury, 1987, Western and Blöschl, 1999):

$$J = \int_0^{\infty} \left[1 - \frac{\gamma(h)}{\sigma^2}\right] dh \tag{3.12}$$

The sample integral scale was estimated from the sample variogram as:

$$\hat{J} = \sum_{i=1}^{N_0} \left[1 - \frac{\hat{\gamma}(h_{i-1}) + \hat{\gamma}(h_i)}{2s^2}\right] \Delta h_i \tag{3.13}$$

where N_0 is the bin number for which $\hat{\gamma}(h_{N_0}) = s^2$, that is the bin number where the sample variogram first intersects a horizontal line $\hat{\gamma}(h) = s^2$. If the variogram reaches the sample variance between two bins, the exact intersection point is found by linear interpolation. It was assumed $\hat{\gamma}(h_0) = 0$ and $\Delta h_i = h_i - h_{i-1}$. For the spatial correlation structure in Eq. 3.4 with no nugget effect, the theoretical integral scale is identical with the correlation length while for

different correlation structures the integral scale can be smaller or larger than the correlation length, depending on the shape of the variogram. All non-parametric estimates are shown for comparison purposes with their parametric counterparts. More details on the non-parametric estimates are given in Chapter 2.

3.2.3.2 Parametric Estimates

The variogram parameters were estimated by two different methods, assuming an exponential variogram model (Eq. 3.4). The first was the WLS method of Cressie (1985). This method optimises the objective function Φ dependent on the set of parameters θ by:

$$\Phi(\theta) = \sum_{i=1}^{N_b} |N(h_i)| \cdot \left[\frac{\hat{\gamma}(h_i)}{\gamma(h_i|\theta)} - 1 \right]^2 \tag{3.14}$$

where $N(h_i)$ is the number of pairs with separation distance h_i, N_b is the number of bins, $\hat{\gamma}(h_i)$ is the estimated sample variogram value for bin i, according to Eq. 3.11, and $\gamma(h_i|\theta)$ is the modelled variogram, given the parameter set θ. This method is simple and has two main advantages over OLS. First, bins with a higher number of pairs will get a higher weight. Second, the residuals close to $h = 0$, where the variogram values are small, will get higher weights because the errors are measured relative to the modelled variogram value. Three cases were examined. First, a one-parameter model with no nugget effect ($c_0 = 0$) was fitted, where it was assumed that the sill was equal to the sample variance s^2. Only the correlation length was fitted. In the second case, a two-parameter model with correlation length and sill was assumed, and the parameters fitted jointly. In the third case, the complete three-parameter model was fitted. The Shuffle Complex Evolution Method (SCE-UA) (Duan et al., 1992) was used for solving the optimisation problem. This method needs limits for the parameter space. To minimise the influence of these limits on the estimates, the possible parameter space had to include many orders of magnitude. It can in some cases be difficult to estimate the parameters in this case, especially if they are small. Therefore the parameters were logarithmised in the search procedure. This method makes it easier to find small parameter values, but at the same time increases the likeliness of estimating small parameters in cases where the optimum is flat.

The second method was the ML method, which maximises the likelihood of actually having observed the data set given the parameters of the variogram model. This method is based on

the assumption that the data conform to a multivariate Gaussian distribution. The joint probability density function of n experimental data can then be expressed as:

$$p(\mathbf{z}|\boldsymbol{\theta}) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\} \qquad (3.15)$$

where $\mathbf{z}$ is the $n\times 1$ vector of data, $\boldsymbol{\mu}$ is the $n\times 1$ vector of means, and $\mathbf{V}$ is the $n\times n$ variance-covariance matrix (Pardo-Igúzquiza, 1998). Separating the covariance matrix into a variance and a correlation matrix, $\mathbf{V}=\sigma^2\mathbf{Q}$, the Negative Log Likelihood Function (NLLF) can be defined as:

$$L'(\hat{\boldsymbol{\beta}},\hat{\sigma}^2,\boldsymbol{\theta}|\mathbf{z}) = \frac{n}{2}(\ln(2\pi)+1-\ln(n))+\frac{1}{2}\ln|\mathbf{Q}|+\frac{n}{2}\ln\left[(\mathbf{z}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{Q}^{-1}(\mathbf{z}-\mathbf{X}\hat{\boldsymbol{\beta}})\right] \qquad (3.16)$$

where $\hat{\beta}$ is the ML estimator of the drift parameters. The NLLF is minimised through an iteration procedure. Cressie (1991, p. 91) and Pardo-Igúzquiza (1998) give a more detailed description of the method, especially how to estimate the mean vector and the variance. A method implemented under the statistical environment R (R Development Core Team, 2004) by Ribeiro and Diggle (2001) was used.

The ML method was used for examining the same three cases as with the WLS method (one-, two-, and three-parameter models). The computational demand of ML estimation rapidly increases with the number of samples. This method was therefore used only to fit variogram parameters to 50 of the sample sets for each sampling scale for 1024 samples, while variogram parameters were fitted to all 1000 sample sets for 16 and 100 samples. The increasing demand of computational power is related to a particular advantage of ML estimation (Lark, 2000), i.e., the ML method estimates the parameters directly from the cloud of squared differences of the variable between the points. No predefined lag classes are hence needed.

To reduce the effect of outliers and confine the search to parameters within a reasonable parameter space, the variogram parameters were allowed to vary within certain bounds for the WLS and ML methods. In the WLS method the estimated nugget was allowed to vary between zero and the estimated sample variance; in the ML method there was no upper limit for the nugget. In both methods, the estimated sill was allowed to vary between 0 and 10 times the estimated sample variance, and the estimated correlation length was allowed to vary between 0 and 10 for $L_E^* < 1$. For larger extents, the upper limit was set to 10 times the extent of the data set, the lower limit to $L_S^*/50$. The lower limit of the correlation length will affect

the estimates of the correlation length, but in practice this is irrelevant since a fitted correlation length smaller than two percent of the spacing of the samples, would not be acceptable.

3.2.4 Presentation of Results

From each realisation for a given combination of the scale triplet, the non-parametric characteristics (mean, variance and integral scale) were estimated by Eqs. 3.9, 3.10, 3.13. The variogram parameters were estimated by the WLS and ML methods. The median was chosen as the representative value of all estimates as the distributions of most parameters were strongly skewed, and plotted the 25 percent and 75 percent quantiles as error bars. The error bars then represent the uncertainty of the estimates one encounters if only one sample set (i.e., one realisation) is available. With 1000 realisations, the estimation error of the median is small. The error bars are only shown for every third sampling scale combination and the estimates from the ML and WLS methods are slightly shifted for clarity of presentation.

Figure 3.2 shows an example of the sample variograms estimated from 10 realisations and samples from a 10*10 grid. The extent was $L_E^* = 5.6$, corresponding to a spacing of $L_S^* = 0.56$ and the support was $L_A^* = 0$. Although the extent is considerably larger than the correlation length ($\lambda^* = 1$), some of the sample variograms still appear non-stationary. Figure 3.2 also indicates the parametric estimates of sill (vertical dashed line) and correlation length (horizontal dashed line) from a two-parameter model, and the non-parametric estimates of sample variance (vertical solid line) and integral scale (horizontal solid line), together with their error bars, as calculated from the complete set of 1000 realisations.

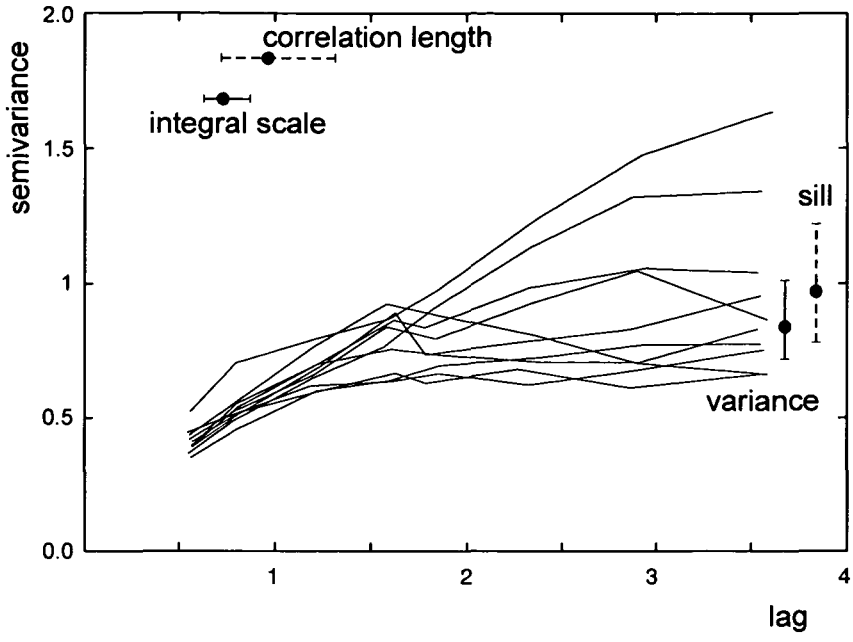


Figure 3.2 Sample variograms obtained from 100 samples with extent $L_E^* = 5.6$ and support $L_A^* = 0$ from 10 realizations of the random field. Horizontal bars represent estimates of the integral scale (solid) and the correlation length (dashed) from a two-parameter model, and the vertical bars represent the variance (solid) and sill from a two-parameter model (median with 25 percent and 75 percent quantiles as error bars).

The estimates are plotted in non-dimensional form scaled by the (true) population characteristics. The non-parametric estimates of the spatial correlation structure (variance and integral scale) are:

$$s^{2*} = \frac{s^2}{\sigma^2} \tag{3.17}$$

$$\hat{J}^* = \frac{\hat{J}}{\lambda} \tag{3.18}$$

and the parametric estimates (i.e., the parameters of the variogram model – nugget, sill, and correlation length) are:

$$\hat{c}_0^* = \frac{\hat{c}_0}{\sigma^2} \tag{3.19}$$

$$\hat{c}_s^* = \hat{c}_0^* + \hat{c}_1^* = \frac{\hat{c}_0 + \hat{c}_1}{\sigma^2} \tag{3.20}$$

$$\hat{\lambda}^* = \frac{\hat{\lambda}}{\lambda} \tag{3.21}$$

As spacing and extent are related through the number of samples, the results for spacing and extent are first presented jointly, followed by the results for the support.

3.3 Results

3.3.1 Spacing and Extent

3.3.1.1 One-Parameter Model

Figure 3.3 shows the results when only correlation length has been fitted to the sample variograms from gridded samples. The green lines relate to the WLS method, the blue lines to the ML method. Unbiased estimates of the correlation length should produce medians of $\hat{\lambda}^* = 1$. For comparison, the red lines show the non-parametric estimates (i.e., the integral scale). For small extents, the WLS estimates are strongly negatively biased for all sample sizes. In these cases, the sample variograms increase monotonously with distance. The sill has been fixed equal to the sample variance, but the estimated correlation lengths increase with extent. The WLS estimates for small extents are almost equal to the estimates of the integral scale, but the uncertainty is slightly larger. The ML estimates for small extents are negatively biased for 16 samples, close to unbiased for 100 samples, and positively biased for 1024 samples. This is related to a tendency of the ML method of estimating the variogram considerably flatter than the WLS method for small extents, which is because the ML method appears to put more emphasis on the short distance part of the variogram and the TBM generated fields are slightly too smooth. As the sill is fixed equal to the variance, the estimated correlation lengths for small extents are larger than those estimated by the two other methods. This effect increases with an increasing number of samples and, while the estimated correlation length is considerably underestimated for 16 samples, it is overestimated for 1000 samples. The small biases of the ML method for 100 samples are hence believed to be due to a compensation of two biases rather than an indication of a generally better performance than the other methods.

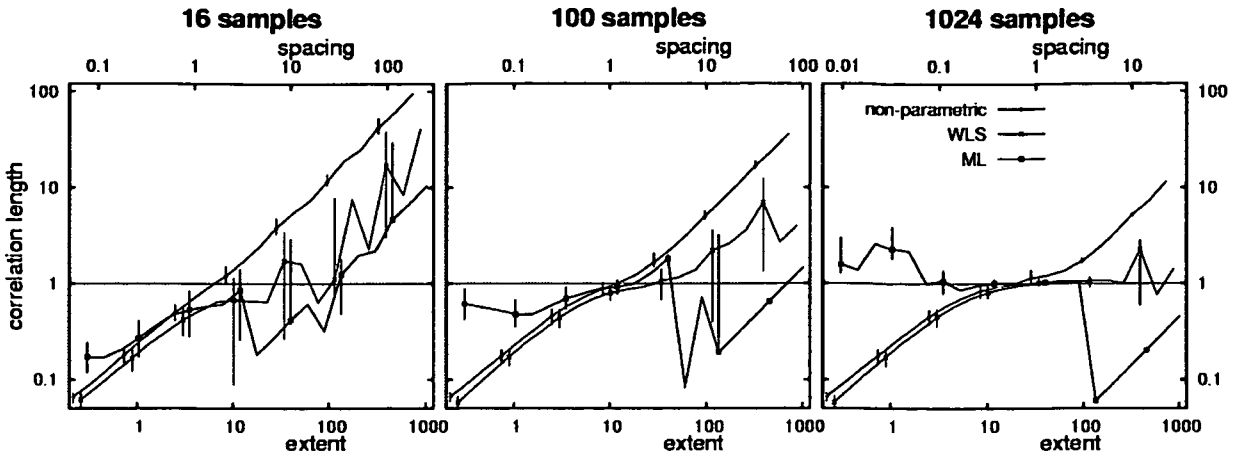


Figure 3.3 Effect of spacing L_S^* and extent L_E^* on the estimated correlation length $\hat{\lambda}^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue) and integral scale $\hat{J}^*$ (red lines) for gridded sampling (median with 25 and 75 percent quantiles as error bars). One-parameter model.

There is only a short range of sampling scales where both the WLS and the ML methods are able to produce close to unbiased estimates of the correlation length. Figure 3.3 indicates that the estimates are unbiased for $L_E^* > 8$ and $L_S^* < 3$. The range hence increases with an increasing number of samples. With 16 samples, the estimates are always biased. Unbiased estimates can only be achieved when the biases from small extent and large spacing cancel each other out. With 100 samples, the range of unbiased estimates is less than one order of magnitude for both the WLS and the ML methods. With 1024 samples, both methods can achieve unbiased estimates over more than one order of magnitude of the sampling scales relative to the scale of the underlying process.

For large spacings $L_S^* > 3$, the WLS estimates are positively biased. The estimates tend to fluctuate between the lower limit allowed and the spacing, and the uncertainty is large. In most cases the sample variograms do not exhibit any correlation (a steep increase to the sill at very short lags), so the WLS method will estimate the correlation length as a value between zero and the lag of the bin with the shortest lag, which has a lag close to the spacing. From the 1000 realisations, a large number of the estimates are equal to the lower limit. If this includes more than 500 of the estimates, the ensemble median is also equal to the lower limit. The ensemble medians from the ML method are close to the lower limit for large spacings. The parametric estimates of the correlation length for the WLS and ML methods are generally less

biased than the non-parametric estimates (i.e., integral scale), but the uncertainty is much larger.

Figure 3.4 shows the results for random samples as opposed to the gridded samples in Fig. 3.3. When extent is small, the results are similar to the results for gridded sampling. This is because the biases and uncertainties are controlled by the extent in this case. For larger extents, and hence larger spacings, most estimates are less biased than the estimates from gridded samples. For 16 samples, the estimates from both the WLS and ML methods increase with increasing extent. With 100 samples, the scale range of unbiased estimates is significantly larger than for the gridded case (approximately between $L_E^* > 8$ and $L_S^* < 10$), but the ML method gives unbiased estimates for larger spacings than the WLS method. The upper limit seems to increase slightly with an increasing number of samples. This is because there will be more sample pairs constraining the short part of the variogram, both increasing the resolution and the certainty of the estimates. The parametric estimates of the correlation length are again less biased than the non-parametric integral scale and the uncertainties are larger. The ML method has the largest uncertainties. With 1024 samples the WLS and ML estimates of the correlation length are close to unbiased over a large range, and the uncertainty is small. This also indicates that the TBM method preserves the correlation structure of the underlying random field.

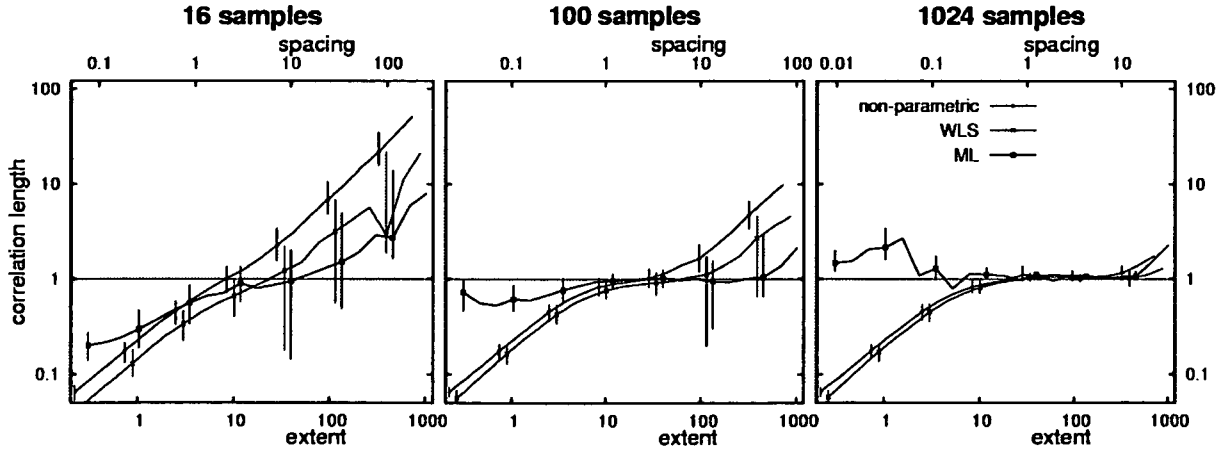


Figure 3.4 Effect of spacing L_S^* and extent L_E^* on the estimated correlation length $\hat{\lambda}^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue) and integral scale $\hat{J}^*$ (red lines) for random sampling (median with 25 and 75 percent quantiles as error bars). One-parameter model.

3.3.1.2 Two-Parameter Model

Figure 3.5 presents the results when sill and correlation length have been simultaneously fitted to the empirical variograms from gridded samples. For large extents the sill converges to unity, which indicates that the TBM method preserves the variance of the underlying random field. For small extents, the WLS method significantly overestimates both the sill and the correlation length. When extent is small, it is apparently not possible to separately estimate the sill and the correlation length. The derivative of the variogram is:

$$\frac{d\gamma}{dh} = \frac{c_s}{\lambda}\exp(-h/\lambda)$$

When the extent is small, the exponential part will be close to 1 and the derivative becomes independent of the lag. The parameter estimation methods can then only identify the ratio between the sill and the correlation length, c_s / λ.

The ML estimates of the sill for small extent are negatively biased and close to the sample estimates of variance, but they are slightly less underestimated from 100 samples and 1024 samples. As the ML estimates of the sill are close to the estimated variance, the ML estimates of the correlation length are similar to the results from the one-parameter model (Fig. 3.3). For large extents, the sill is estimated close to the true sill although for a small number of samples the uncertainty is larger than that of the sample variance. For extents slightly smaller than 1, the estimated sill and correlation lengths from the WLS and ML methods are less biased than the estimated variance and integral scale, except the ML estimates from 16 samples. The uncertainty is, on the other hand, much larger, even for the WLS estimates from 1024 samples. For large extents and spacings, the estimates of the sill are unbiased with small uncertainties, and the estimates of the correlation lengths are similar to the one-parameter model of Fig. 3.3. The estimates of correlation length are practically unbiased for $L_E^* > 2$ and $L_S^* < 3$. This means that the two-parameter model is able to provide unbiased estimates for smaller extents than the one-parameter model for which the limits were $L_E^* > 8$ and $L_S^* < 3$.

Figure 3.6 shows the same estimates as Fig. 3.5 but for a random sampling scheme. For extents $L_E^* < 5$, the sampling strategy does not appear to influence the estimates. For large extents and spacings, the figure shows the same effect as for a one-parameter model in Fig. 3.4 with a wider range of unbiased estimates than from gridded sampling.

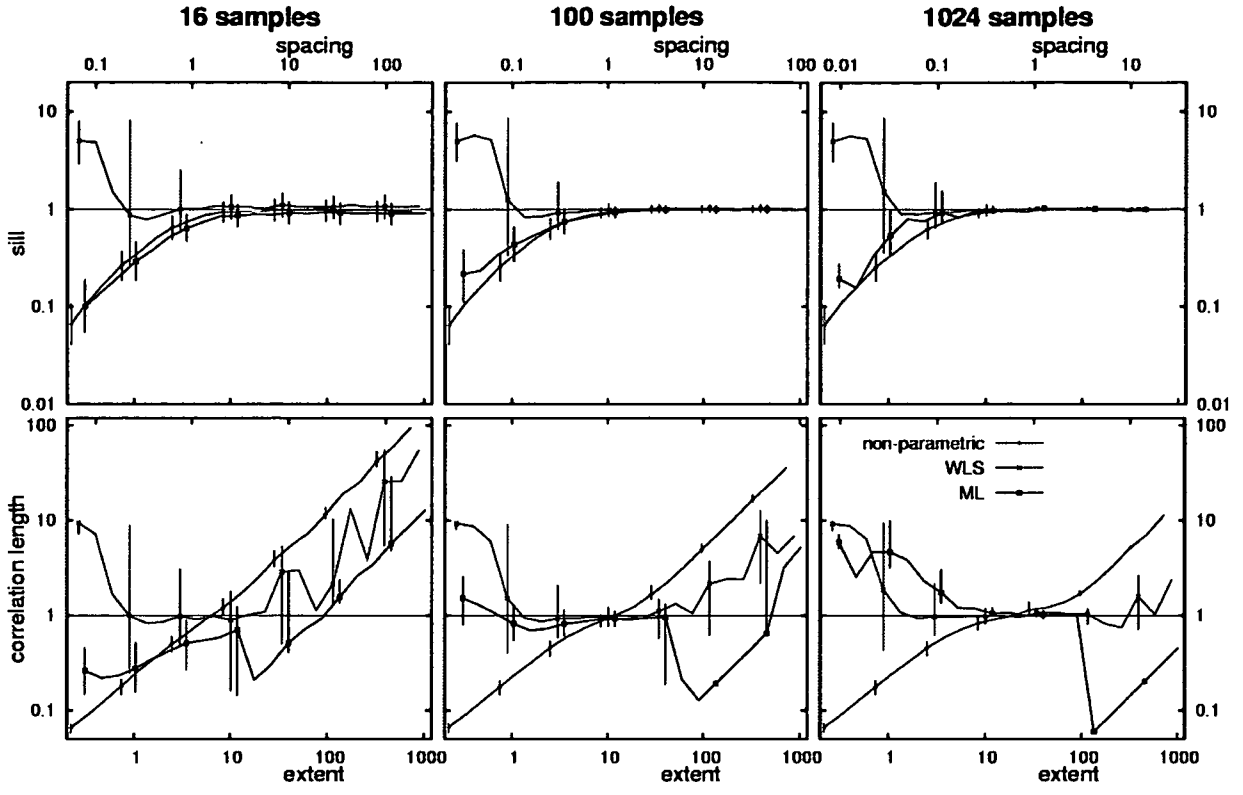


Figure 3.5 Top panels: Effect of spacing L_S^* and extent L_E^* on the estimated sill $\hat{c}_s^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue) and the sample variance s^{2*} (red). Bottom panels: Effect of spacing L_S^* and extent L_E^* on the estimated correlation length $\hat{\lambda}^*$ (WLS green, ML blue) and the integral scale $\hat{J}^*$ (red). Gridded sampling (median with 25 and 75 percent quantiles as error bars); two-parameter model.

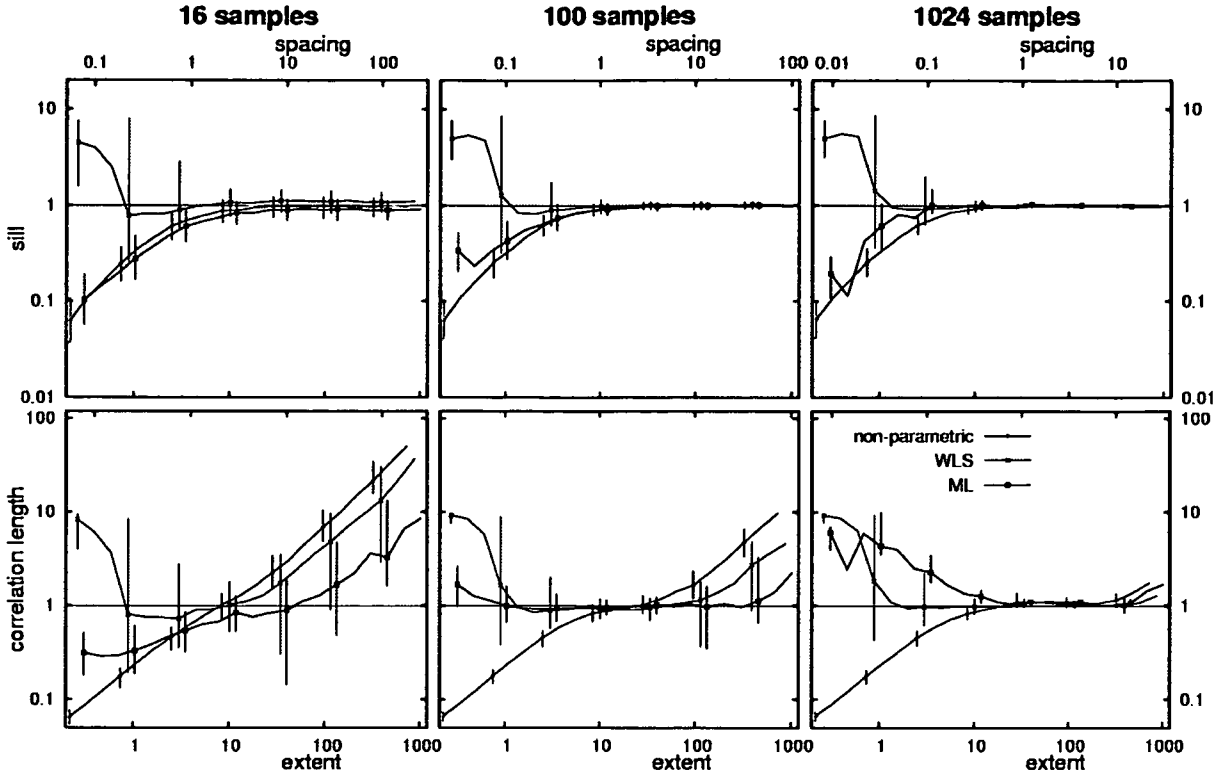


Figure 3.6 Top panels: Effect of spacing L_S^* and extent L_E^* on the estimated sill $\hat{c}_s^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue) and the sample variance s^{2*} (red). Bottom panels: Effect of spacing L_S^* and extent L_E^* on the estimated correlation length $\hat{\lambda}^*$ (WLS green, ML blue) and the integral scale $\hat{J}^*$ (red). Random sampling (median with 25 and 75 percent quantiles as error bars); two-parameter model.

3.3.1.3 Three-Parameter Model

Figure 3.7 shows the results of the three-parameter variogram model for gridded sampling. Unbiased estimates should produce median nuggets of $\hat{c}_0^* = 0$, median sills of $\hat{c}_s^* = 1$, and median correlation lengths of $\hat{\lambda}^* = 1$. Figure 3.7 top indicates that there is close to zero nugget effect for short spacings $L_S^* < 1$. For larger spacings, the estimated nugget increases for all sample sizes. The nugget estimated by the ML method is larger than that of the WLS method and approaches the underlying variance for very large spacings. The estimates of the sill in the middle panels of Fig. 3.7 indicate that, for small spacings ($L_S^* < 1$), there are only small differences between this sill and the estimates from the two-parameter variogram model, shown in Fig. 3.5. The estimates from 16 samples are always positively biased and are associated with larger uncertainties than the estimates from the two-parameter model. For

large extents, the bias is small, but the uncertainties of the WLS and ML estimates are considerably higher than the uncertainties of the sample variance for 16 and 100 samples. In the case of 1024 samples, the uncertainties are only slightly larger than those of the sample variance. The ML method slightly underestimates the sill for 16 samples while the WLS method slightly overestimates the sill. The ML estimates have significantly larger uncertainty than the ML estimates for 16 samples, and slightly larger uncertainties for estimates from 100 samples. Figure 3.7 bottom indicates that, for small extents, the estimates of the correlation lengths are similar to the two-parameter estimates shown in Fig. 3.5, except for a larger bias in the case of 16 samples. For large spacings $L_S^* > 3$, the estimates are positively biased, similar to the estimates from the one- and two-parameter model, but the uncertainties are larger. It appears to be impossible to fit a three-parameter variogram model to 16 samples with a reasonable level of accuracy. For 100 and 1024 samples, reasonable fits over a scale range of $L_E^* > 2$ and $L_S^* < 3$ are possible with the WLS and ML methods, but the uncertainty at the upper limit ($L_S^* = 3$) is much larger than for a two-parameter model.

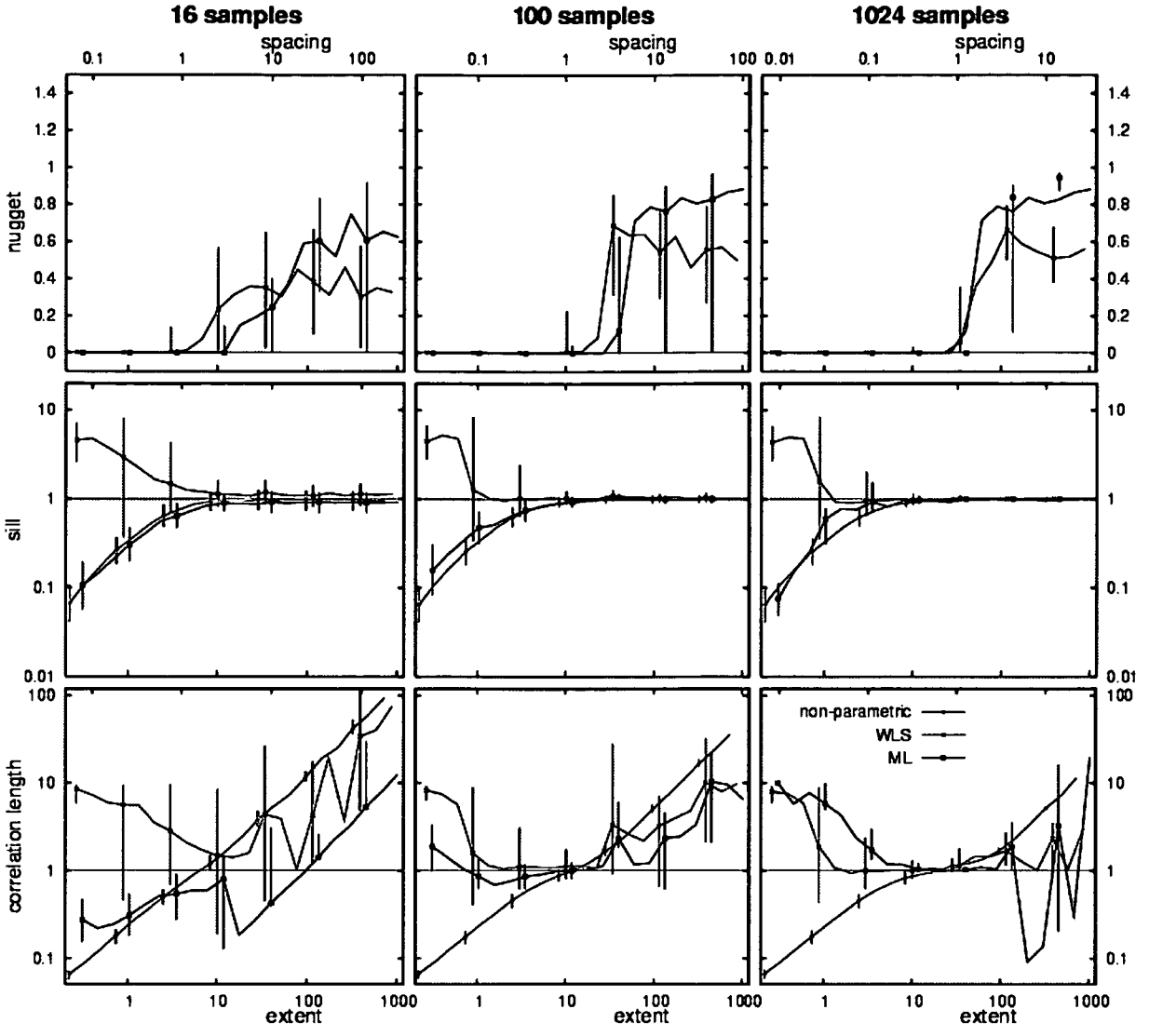


Figure 3.7 Top panels: Effect of spacing L_S^* and extent L_E^* on the estimated nugget $\hat{c}_0^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue). Middle panels: Effect of spacing L_S^* and extent L_E^* on the estimated sill $\hat{c}_s^*$ (WLS green, ML blue) and the sample variance s^{2*} (red). Bottom panels: Effect of spacing L_S^* and extent L_E^* on the estimated correlation length $\hat{\lambda}^*$ (WLS green, ML blue) and the integral scale $\hat{J}^*$ (red). Gridded sampling (median with 25 and 75 percent quantiles as error bars); three-parameter model.

Figure 3.8 shows the same estimates as Fig. 3.7, but for a random sampling scheme. The results are generally similar to those in Fig. 3.7 but differ for large spacings. Figure 3.8 top indicates that the biases in the nugget first appear for spacings larger than in the gridded sampling case. Estimates of the sill are similar to the estimates from the gridded sampling scheme, but the correlation lengths can be estimated without bias over a wider range of

sampling scales than for gridded sampling. For large spacings, the uncertainty appears to be slightly smaller than for gridded sampling. For large spacings, estimates of the correlation lengths from 100 and 1024 samples show similar biases as the integral scale but the uncertainties are larger.

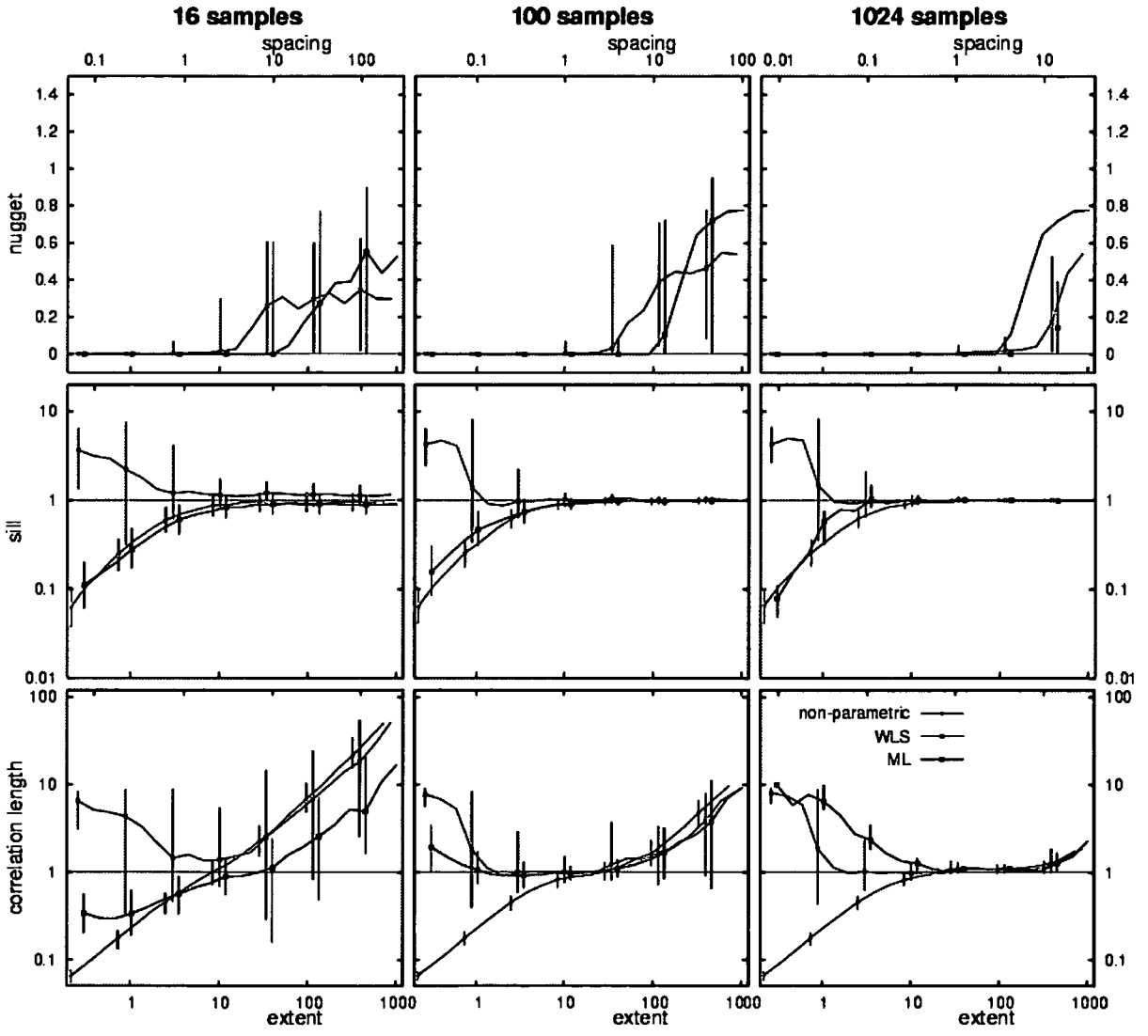


Figure 3.8 Top panels: Effect of spacing L_S^* and extent L_E^* on the estimated nugget $\hat{c}_0^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue). Middle panels: Effect of spacing L_S^* and extent L_E^* on the estimated sill $\hat{c}_s^*$ (WLS green, ML blue) and the sample variance s^{2*} (red). Bottom panels: Effect of spacing L_S^* and extent L_E^* on the estimated correlation length $\hat{\lambda}^*$ (WLS green, ML blue) and the integral scale $\hat{J}^*$ (red). Random sampling (median with 25 and 75 percent quantiles as error bars); three-parameter model.

3.3.2 Support

3.3.2.1 One-Parameter Model

Figure 3.9 shows the results for the case where the support is large compared to the underlying correlation length, using a one-parameter model and gridded samples. The vertical line in each panel shows $L_S^* = L_A^*$. An extent of $L_E^* = 10$ has been used in all cases. The figure indicates that large supports generally increase the correlation length and the integral scale. This is related to the smoothing effects of aggregation, which reduces the variability over short distances. For 100 and 1024 samples, the estimates from the ML and WLS methods are quite similar as long as $L_A^* < L_S^*$ with little bias. For larger supports, the biases increase. The WLS estimates of the correlation length from 100 and 1024 samples are slightly less biased than the estimates of the integral scale. Both the WLS and the ML methods underestimate the correlation lengths for small supports from 16 samples, as the spacing is large relative to the underlying correlation length. The estimates in this case are similar to the results for the same spacing and extent in Fig. 3.3 ($L_E^* = 10$, $L_S^* = 2.5$).

The ML estimates exhibit significant biases for large supports. This is because the variograms tend to monotonously increase with lag if the support is large. This is similar to the situation on the left side of Fig. 3.3, where also the ratio between sill and correlation length is estimated, rather than the sill and correlation length separately. Similar to what was noted earlier, the ML procedure has a tendency of underestimating this ratio. For a fixed sill, the correlation length is overestimated, leading to significant biases for 100 samples, and even more so for 1024 samples as shown in Fig. 3.9.

Random sampling (not shown here) gives similar results to those shown in Fig. 3.9, with some differences for 16 samples, as spacings are also large in this case. For small supports, the biases and uncertainties are similar to the results for the same spacing and extent in Fig. 3.4.

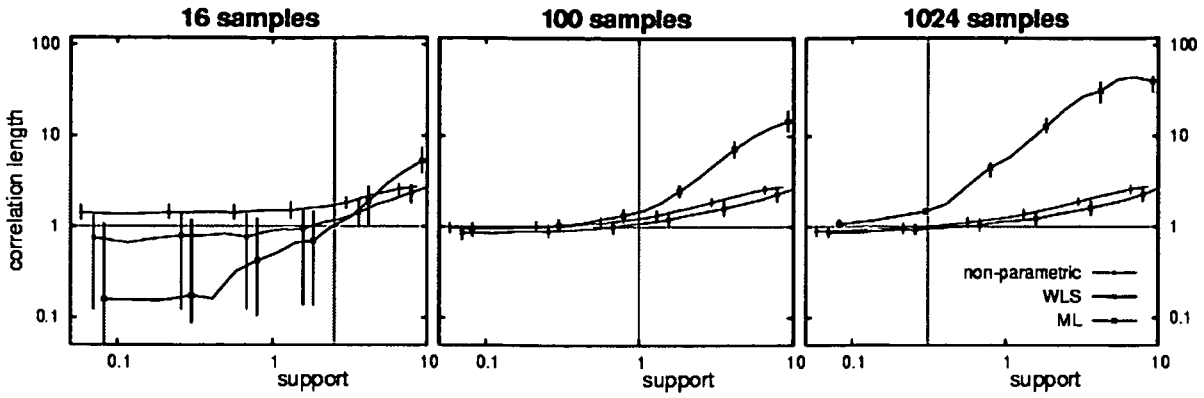


Figure 3.9 Effect of support L_A^* on the estimated correlation length $\hat{\lambda}^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue) and integral scale $\hat{J}^*$ (red) for gridded sampling (median with 25 and 75 percent quantiles as error bars). Vertical line in each panel shows $L_S^* = L_A^*$. One-parameter model. $L_E^* = 10$, spacing L_S^* according to Eq. 3.8.

3.3.2.2 Two-Parameter Model

Figure 3.10 shows the results for the case where the support is large compared to the underlying correlation length using a two-parameter variogram model and gridded sampling. The estimates of sill and variance (top panels in Fig. 3.10) decrease with increasing support, except for some estimates of the sill from the WLS method. The ML estimates of the sill are close to the estimates of the sample variance. For large supports, the variogram will appear non-stationary and it is again only possible to estimate the ratio between sill and correlation length. The WLS estimates of the sill then increase. For yet larger supports, the effect of reduced sample variance due to smoothing reduces the estimated sill.

The second row of panels shows the estimates of correlation length and integral scale. For small supports, the results of all methods are similar to the results of the one-parameter model. For large supports, the ML estimates are still similar to the estimates from the one-parameter model. The estimates from the WLS method, however, increase much faster with increasing support than those from the one-parameter model, which is again related to the fact that only the ratios between sill and correlation length can be estimated in this case. The WLS method estimates larger sills than the ML method, similar to the case of small extents. For very large supports, the WLS method overestimates correlation lengths more significantly than the ML method.

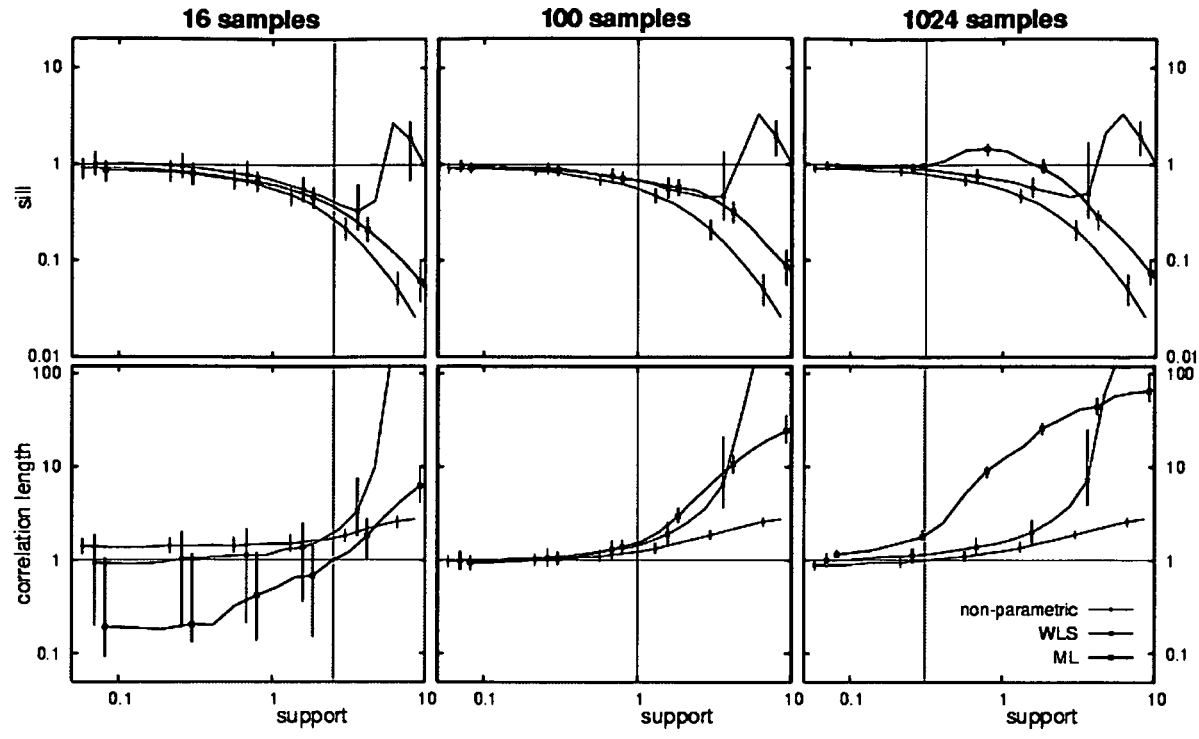


Figure 3.10 Top panels: Effect of support L_A^* on the estimated sill $\hat{c}_s^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue) and the sample variance s^{2*} (red). Bottom panels: Effect of support L_A^* on the estimated correlation length $\hat{\lambda}^*$ (WLS green, ML blue) and the integral scale $\hat{J}^*$ (red). Gridded sampling (median with 25 and 75 percent quantiles as error bars); two-parameter model. $L_E^* = 10$, spacing L_S^* according to Eq. 3.8.

3.3.2.3 Three-Parameter Model

Figure 3.11 shows similar results as Fig. 3.10 but for a three-parameter model. Biases in the nugget are only estimated in the case of 16 samples. This is because of the relatively large spacing of $L_S^* = 2.5$. The WLS estimates of the nugget have both higher median values and higher uncertainty than the ML estimates. For 100 samples and small support there is some uncertainty in the nugget, which again is related to the spacing ($L_S^* = 1$). The second row of panels of Fig. 3.11 shows the estimates of the sill. The results are similar to the results from the two-parameter model, except that the uncertainty is higher, especially for estimates from 16 samples. Figure 3.11 bottom shows the estimates of the correlation length together with the integral scale. For 100 and 1024 samples, the results are similar to the results of Fig. 3.10, except for small supports where the uncertainties are larger, especially for the WLS estimates.

For 16 samples, the ML method estimates the correlation length close to the lower limit, while the WLS method strongly overestimates the correlation length.

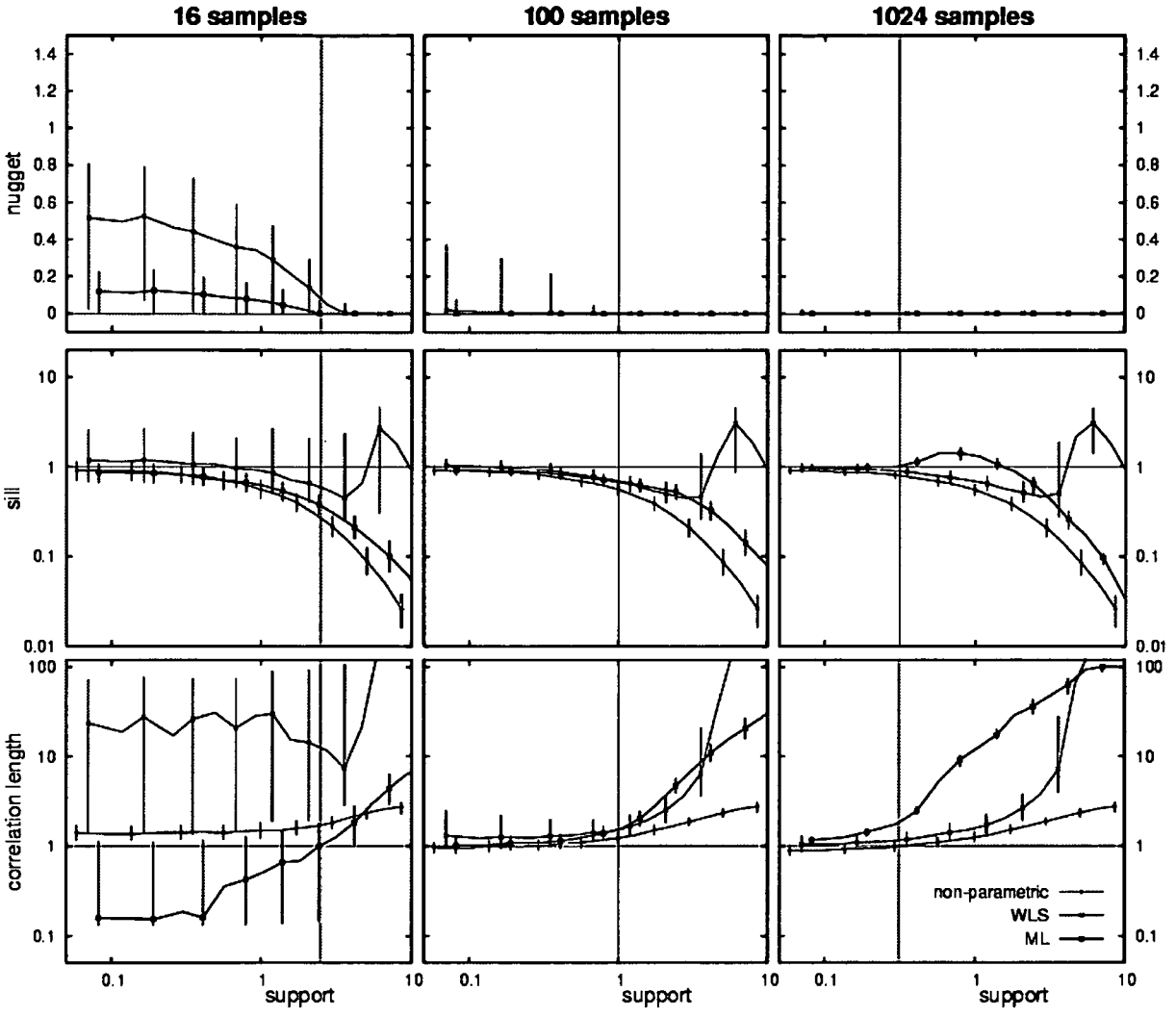


Figure 3.11 Top panels: Effect of support L_A^* on the estimated nugget $\hat{c}_0^*$ (Weighted Least Squares (WLS) green and Maximum Likelihood (ML) blue). Middle panels: Effect of support L_A^* on the estimated sill $\hat{c}_s^*$ (WLS green, ML blue) and the sample variance s^{2*} (red). Bottom panels: Effect of support L_A^* on the estimated correlation length $\hat{\lambda}^*$ (WLS green, ML blue) and the integral scale $\hat{J}^*$ (red). Gridded sampling (median with 25 and 75 percent quantiles as error bars); three-parameter model. $L_E^* = 10$, spacing L_S^* according to Eq. 3.8.

3.3.3 Comparison with Expected Biases

Western and Blöschl (1999) presented analytical expressions for the expected variance and integral scale as a function of extent, spacing and support. These are shown in Fig. 3.12a (thin

black lines) together with the WLS estimates of the variogram parameters of a two-parameter model for 100 gridded samples. Figure 3.12b shows the corresponding values for the ML estimates. The estimates of integral scale in both figures are close to the analytical expectations, while the parametric estimates of the correlation lengths differ from these expectations for both the WLS and ML methods. The summary also nicely demonstrates that, depending on what component of the scale triplet is changed, the effect on the sill and the correlation length will differ.

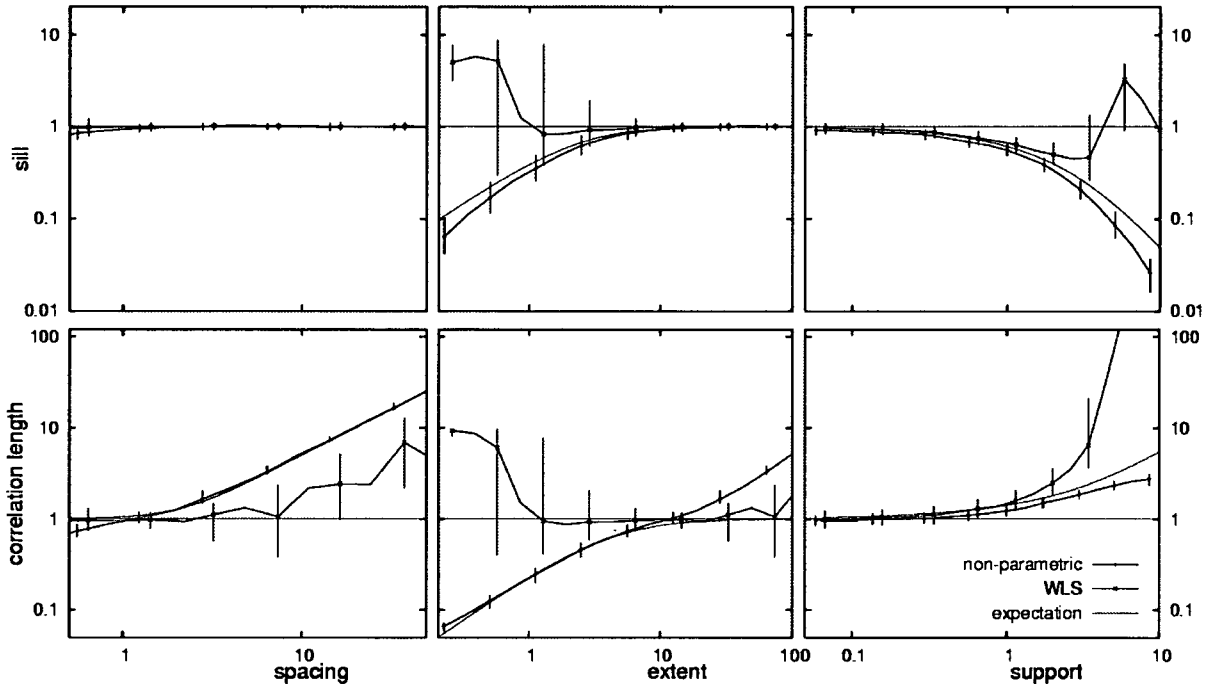


Figure 3.12a Estimated sill and correlation length from the Weighted Least Squares (WLS) method (green), together with variance and integral scale (red). The thin black lines correspond to the analytical expectations. Two-parameter model; 100 gridded samples.

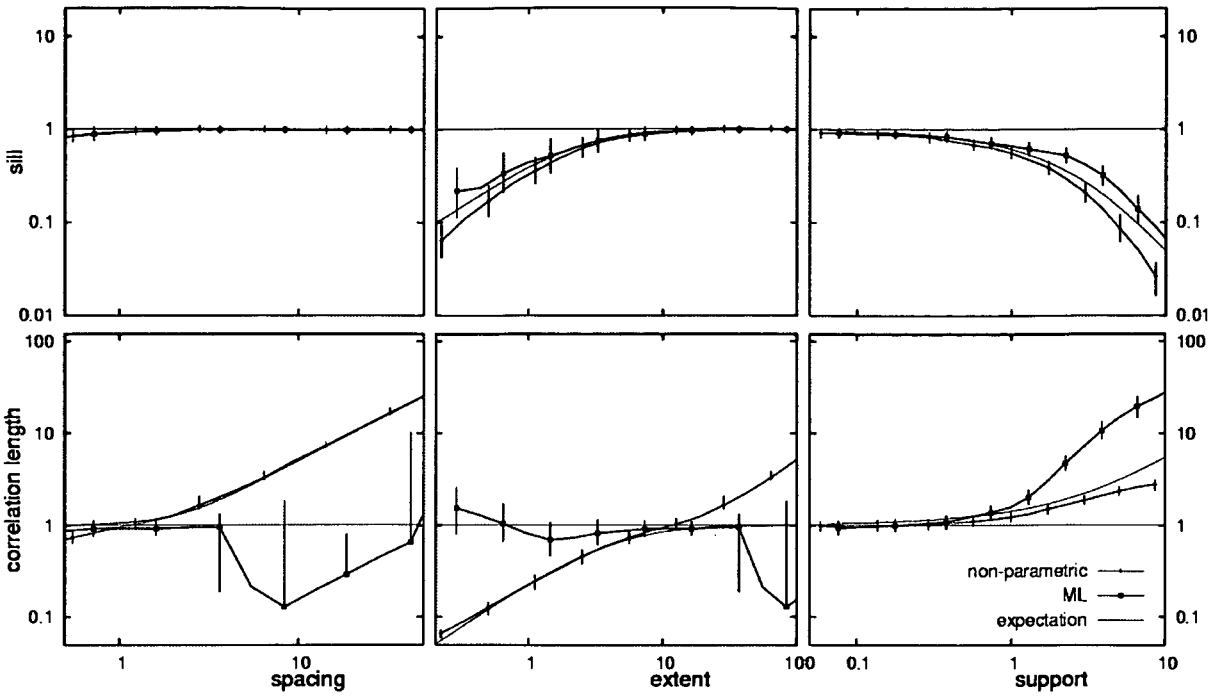


Figure 3.12b Estimated sill and correlation length from the Maximum Likelihood (ML) method (blue), together with variance and integral scale (red). The thin black lines correspond to the analytical expectations. Two-parameter model; 100 gridded samples.

3.4 Discussion and conclusions

There are two purposes of estimating a sample variogram and fitting a variogram model to the sample variogram. The first is to understand and characterise the spatial correlation of the variable of interest for which a theoretical variogram of the population of the random field is needed. The second purpose is interpolating the variable of interest within the study area to points where measurements are unavailable, for which a local variogram may be sufficient. This chapter focuses on the first application.

The random variable examined in this chapter can be any variable of interest in soil hydrology. The most typical variables are soil moisture, different kinds of concentrations, such as concentration of nutrients, or physical properties of the soil, such as its thickness, the thickness of pedological layers, soil density, and texture. In addition, it is also possible to use geostatistical methods to characterise spatially the dynamical properties of the soil such as the hydraulic conductivities.

The simulations in this chapter indicate that sampling scales significantly influence the results by biasing the estimates. This is not surprising as similar patterns were found in analyses of real data such as those of Gelhar (1993) and Blöschl (1999) who showed that the estimated correlation lengths tended to be a function of the extent of the data. The parametric estimates are found to be less biased than their non-parametric counterparts, which is consistent with the findings of Russo and Jury (1987). While in many cases the biases of the parametric estimates are smaller than those of the non-parametric methods, their uncertainties are generally higher, especially for estimates of the correlation length. This uncertainty also increases with an increasing number of parameters. The number of samples is important in this context as would be expected. An increasing number of samples reduce the uncertainty considerably. Quite clearly, correlation lengths cannot be estimated with any degree of confidence from 16 samples, since either the spacing is too large or the extent is too small. The window of unbiased estimates where the sampling scales are commensurate with the scale of the underlying variability is non-existent in the case of 16 samples. For 100 samples, the variogram is based on almost 5000 sample pairs, but these might be highly correlated. This is described in methods of assessing the uncertainty of variograms (Ortiz and Deutsch, 2002; Pardo-Igúzquiza and Dowd, 2001). Because of the correlation, variogram parameters estimated from 100 points may still be highly uncertain. Webster and Oliver (1992) suggested that the number of samples should be at least 150, and preferably 225. This suggestion was related to estimates of the local variograms. It is much more difficult to make such a statement about estimates of the theoretical variogram that applies to the population of realisations rather than to one particular study area. The uncertainties not only depend on the number of samples but also on the sampling scale. The uncertainties in the variogram parameters are almost independent of the sample size when the extent is small, while they strongly depend on the sample size for large extents. The uncertainty of the parameters also differs vastly with the number of variogram parameters to be estimated. One-hundred samples give a small uncertainty for the correlation length when it is the only parameter to be fitted and the extent is small. One thousand and twenty four samples can be too few for properly estimating a three-parameter variogram model if the extent is large. The exact range of relatively unbiased and certain estimates depends on the estimation method, the number of samples, and the spatial sampling arrangement. This range is the window where the sampling scales are commensurate with the scale of the underlying variability that would likely be used in a sampling setup. A large number of samples both increases the range of unbiased estimates and reduces the uncertainty within this range. The parametric estimates are

generally unbiased for non-dimensional extents $L_E^* > 2$ and non-dimensional spacings $L_S^* < 3$. Unbiasedness can also be achieved for larger spacings if the samples are taken at random locations. For a nested sampling arrangement one would also expect a wide range of unbiased estimates.

In order to make an *a priori* assessment of the suitability of a planned sampling setup, knowledge of the true correlation length or integral scale is required, as the analyses in this chapter are based on sampling scales normalised by the true correlation length of the underlying process. There are a number of options to obtain such estimates. The first is an exploratory sampling along a transect or in a nested sampling arrangement to obtain initial estimates of the correlation length. The second is to use values from the literature of similar studies. Third, process reasoning can be used to obtain an educated guess of the correlation lengths of the variables of interest in landscape soil processes. Seyfried and Wilcox (1995) proposed a framework of inferring correlation lengths from geological and physicals features that are important for hydrological modelling at different scales. Visual examination of outcrops and other soil related features in the landscape can be a valuable guide for obtaining an order of magnitude of the scale at which soil processes operate in a particular environment.

It is in some cases possible to correct for the bias effects of the scale triplet. If the spacing is small, the extent is large, and the support is small the biases will be small. As the spacing gets larger, the extent smaller, and the support larger, the biases will increase. If the components of the scale triplet do not differ too much from the unbiased case, estimates of the biases of the variance and the correlation length can be back-calculated from the results of the simulations study in this chapter. This is illustrated in Western and Blöschl (1999) and in Chapter 2 of this dissertation. The uncertainty of these back-calculation methods increases with increasing spacing, decreasing extent, and increasing support. For the case of large supports it has also been suggested to numerically back-calculate the point variogram by regularisation methods, which can deal with more general cases than addressed in this chapter (Mockus, 1998, Kyriakidis, 2004 and Chapter 5 and 6 of this dissertation). In this analysis it has been assumed that the variable of interest aggregates linearly, i.e., the value over a support area is simply the arithmetic average of the point values within this support area. Some of the variables of interest in soil hydrology do indeed aggregate linearly, such as soil moisture and soil thickness. Other variables such as hydraulic conductivities do not. More complicated aggregation rules exist for these parameters. These are dealt with in the abundant literature on up-scaling and effective parameters (e.g., Blöschl and Sivapalan, 1995; Hopmans and

Schoups, 2005). The support effects described in this chapter will for such cases only be an approximation to the real support effects.

The different estimation methods examined here have different characteristics. The non-parametric estimates are in most cases associated with the lowest uncertainties. They are on the other hand highly biased. The WLS method provides visually satisfying fits to the sample variograms. The uncertainty, however, tends to be large and increases with the number of variogram parameters to be estimated. The WLS method performs slightly better than the ML method for 16 samples. The ML method is computationally more demanding for large sample sizes because, for n data an $n \times n$ matrix has to be inverted for each iteration. The ML method gives better results for randomly distributed samples with large spacings, but does not seem to be significantly better than the WLS method for other sampling scale combinations. In addition, estimated variograms tend to be too flat for small extents. The results here are consistent with analyses performed by Zimmerman and Zimmerman (1991, p.90) who noted that "'krigers' will sacrifice little by using the easy-to-compute OLS [Ordinary Least Squares] or ... WLS ... estimators, rather than the REML [Restricted Maximum Likelihood] estimators". Lark (2000) also compared ML estimates with estimates from fitting a theoretical variogram to the sample variogram. He concluded that ML estimates were often advantageous for short extents but this was not the case for large extents and large nuggets. For a certain experimental setup, he showed that an ML estimate from 60 samples provided an equally good fit to the underlying variogram as a theoretical model fitted to the sample variogram using 90–120 samples. Lark (2000) also concluded that a potential advantage of the ML method is lost when the data are contaminated with outliers, as most real data sets are.

Different opinions exist in the literature on whether variograms should be estimated automatically or hand-fitted. Webster and Oliver (2001) argue for automatic calibration, but suggest that the estimated variogram should be visually examined as well. A number of authors have, however, provided arguments in favour of visually fitting (Ai-Geostats, 2004). Automatic fitting methods were used here because of the large number of variograms to be analysed, and to allow a more objective interpretation of the results. On the other hand, the more complex the correlation structure, the more difficult it will be to visually fit a variogram. Visually fitting a variogram with few parameters such as that of Eq. 3.4 in one dimension is a simple task, but is more complicated if nested variograms and more dimensions are analysed.

For small extents, typically where $L_E^* < 1$, strong trends will appear in the data. One could argue that for these cases a trend should be modelled explicitly and the geostatistical analysis

should be performed on the residuals. There are several reasons why it was chosen to estimate the variogram directly from the data. First, the interest was in the population characteristics, and apparent trends are indeed a feature of the population characteristics for short extents. Second, there is no objective way of ascertaining whether there is a trend in the data or the strength of it (Leuangthong and Deutsch, 2004). Third, it would raise a range of questions beyond the purpose of this chapter. Fourth, although in this chapter it was clear that the variable was non-stationary within the extent of the data when extents were short, this is not always the case for real data. This point has been nicely illustrated by Gelhar (1993, p. 295) who presents the same variogram plotted on linear-linear and log-log scales. The former looks stationary while the latter indicates that the variable is non-stationary. Di Federico and Neuman (1997) and Cintoli et al. (2005) suggest that the concept of truncated power variograms can be used to explain these sampling effects.

In vadose zone applications at the landscape scale, there are a range of variables that are of interest including soil water content, concentrations of chemical or biological tracers, as well as hydraulic soil characteristics. In examining these variables, it is a key concept that sampling always involves some sort of filtering (Cushman, 1984, 1987). It is important to take these filtering aspects into account when interpreting both observations and model estimates of these variables. The major implication of this study for soil hydrology is that sampling scales will likely have an important effect on the variogram parameters estimated from vadose zone data at the landscape scale. Sampling scales will introduce biases and random errors if these scales depart from the scales of the underlying process. Their magnitude is controlled by the sampling scale triplet relative to the scale of the underlying process.

Chapter 4. Characteristic space-time scales in hydrology

4.1 Introduction

Variability of hydrologic processes occurs over many orders of magnitude, from pore scale flow processes up to the global cycling of water and energy, from rain splash effects during less than a second up to changes in the hydrologic balance over centuries and more as a result of climate fluctuations and geomorphic processes. The wide range of variability complicates the scientific analysis of hydrologic processes, be it through theoretical analyses, measurements, or model studies. One concept of dealing with variability over many orders of magnitude is the notion of characteristic scales. The idea of a characteristic scale is that instead of dealing with a spectrum of lengths and times one adopts one typical length and time that is representative of a particular process. Often a characteristic scale is an order of magnitude figure, given as an integer power of ten, rather than a precise number.

While sister disciplines have an excellent track record of capitalising on the potential of characteristic scales, their use in hydrology has been rare, even though their merits are undisputed (Dooge, 1986, Seyfried and Wilcox, 1995, Blöschl, 1999). In fluid dynamics, atmospheric sciences and ecology, characteristic scales are widely used to tag processes for a number of purposes. Some of the most obvious merits are summarised below.

First, characteristic scales can be used for performing order of magnitude or "back of the envelope" analyses. Excellent examples in the context of mixing processes in open waters are given in Fischer et al. (1979) and a range of illustrative examples in the environmental sciences in general are given in Harte (1988). Second, characteristic scales may facilitate the formulation of simple relationships. For example, Prandtl's (1925) suggestion of a "mixing length" allowed him to determine the mean velocity distribution of a fluid near the wall with the aid of only one empirical constant. Third, characteristic scales are very powerful when used in connection with dimensional analyses and dimensionless numbers. The classical example is the Froude number, which is composed of two characteristic velocities. A good overview for boundary layer meteorology is given in Stull (1988) and a number of suggestions for how to apply dimensional analyses to hydrology are given in Dooge (1986).

Fourth, in a similar fashion, characteristic scales play a central role in similarity theories which are well developed in the atmospheric sciences (e.g. Haltiner and Williams, 1980), based on the early work of Charney (1948), and other disciplines such as soil physics (Miller and Miller, 1956). In hydrology, similarity theories are at a very early stage of development (Blöschl and Sivapalan, 1995). Fifth, characteristic scales are among the main input variables into behavioural or structure-imitating space-time models (e.g. Rodríguez-Iturbe, 1986, Koltermann and Gorelick, 1996). Sixth, characteristic scales provide a powerful tool for planning and interpreting measurements. In order to capture a particular process, the space-time scales of the measurement set up must be compatible with the characteristic scales of the processes one is attempting to sample. The case is excellently made by Smagorinsky (1974) for the example of ocean dynamics and an example in the context of the spatial distribution of soil moisture is given in Western and Blöschl (1999). Lastly, and maybe most importantly, the identification of characteristic temporal and spatial scales relevant to a prescribed problem can provide a common framework between disciplines (Blöschl, 2001), as excellently illustrated in Seyfried and Wilcox (1995) for mountain hydrology and Hatcher et al. (1987) for the case of coral reef systems.

In hydrology the various applications of characteristic scales are still in their infancy but it is hoped that deriving characteristic scales will foster the development of methods similar to those in the sister disciplines.

Characteristic scales can be defined in a number of ways, either as deterministic scales or as stochastic scales (Blöschl, 1996). Deterministic characteristic scales of hydrologic processes include storm size and storm duration (e.g. Austin and Houze, 1972); size of saturation areas; residence times and response times from catchments and hillslopes (Uchida et al., 2001); hillslope length and drainage density; and soil depth and aquifer depth. An alternative is to derive stochastics characteristic scales from a correlation analysis. Stochastics characteristic scales were first proposed by Taylor (1921) as a measure of the average correlation distance in turbulence and have since then been widely used for characterising space-time variability in a number of disciplines (e.g. Vanmarcke, 1983, Christakos, 2000).

In this chapter, the stochastic characteristic scales are examined based on correlation analyses in both space and time. The correlation can be represented by the variogram (see e.g. Webster and Oliver, 2001 for a practical assessment). If the variable under study is stationary in the mean the variogram will flatten out at large distances while for non-stationary variables the variogram will increase continually. Strictly speaking, a characteristic scale only exists for a

stationary variable and can be defined as the distance (or time) where the variogram flattens out. When deriving quantitative estimates of the characteristic scale from the variogram one can use one of various definitions of scale such as the integral scale (average distance of correlation), the range (the maximum distance of correlation), and the e-folding distance (the distance where the variogram value is $1-1/e$ of its maximum value) (Blöschl, 1996) which all give similar orders of magnitude of the characteristic scale. For non-stationary variables, approximate characteristic scales can be derived if the variogram becomes flatter with increasing distance. If the variogram increases as a power of distance a characteristic scale does not exist, but this is another important case, termed a fractal variogram. It is of the form:

$$\gamma(h)= ah^Z \tag{4.1}$$

where h is the lag (i.e. either distance or time lag), a is a parameter controlling the variance, and Z is the fractal power of the variogram. Z is a measure of the ratio of large-scale variability and small-scale variability. Fractal behaviour is an indication of long term persistence (or large spatial correlations) also referred to as the Hurst phenomenon (Klemeš, 1974, Kirchner et al., 2000). A variogram of the form of Eq. 4.1 can be considered self similar as it satisfies

$$\gamma(h) = \lambda^{-Z}\gamma(\lambda h) \tag{4.2}$$

where λ is a scale factor, i.e. lag dependent variance at one scale is a constant multiple of that at another scale, and this constant depends on the scale factor. Because of the self similarity property, fractal variability lends itself to similarity analyses (alternatively to the use of characteristic scales) and many of the other analyses that can be done with characteristic scales. It is therefore useful to ascertain whether characteristic scales exists, and if not, if the variogram is fractal and what the fractal power is. It is difficult to accurately determine the fractal power Z (Gallant et al., 1994) and to ascertain with certainty whether the variogram is fractal or not (Kirchner et al., 2000) so a reliable analysis must be based on a sufficiently large data set.

Characteristic time scales can be derived from a variogram analysis of time series. Characteristic space scales can be derived from a variogram analysis of spatial patterns. It is also of interest to examine the space-time behaviour of characteristic scales. Stommel (1963) was probably the first to introduce diagrams showing a schematic relationship between spatial and temporal process scales. He used the diagram for characterising ocean dynamics. Blöschl and Sivapalan (1995) adapted the idea to hydrology and based their diagram (their Figure 2)

on data from Orlanski (1975), Dunne (1978), Fortak (1982) and Anderson and Burt (1990) as well as heuristic reasoning. Their diagram has been widely used in the literature to illustrate the existence of characteristic space-time scales in hydrology in a qualitative way (e.g. Quattrochi, 1997, Wilby, 1997, Ambroise, 1999, Buchholz, 2000, Schulze, 2000, Niehoff, 2002). In this type of diagram two quantities are of main interest. The first is the characteristic velocity of a process defined as the ratio of characteristic lengths and times. The characteristic velocity is a measure of the scales over which measurements become decorrelated from one another and may be related to a number of physical causes. It may be related to the celerity with which a perturbation (such as a weather front) moves through a system. It may correspond to particle velocities (such as typical vertical velocities in convective cells), and it may be related to the internal dispersion and disruption by smaller-scale perturbations. Blöschl and Sivapalan (1995) and Blöschl et al. (1995) suggested that, typically, characteristic velocities are on the order of 10 m/s for rain processes, 1 m/s for channel flow, 0.1 m/s for saturation excess overland flow and less than 10^{-4} m/s for groundwater flow. The second quantity is the slope of the space-time traces on a double logarithmic plot, i.e. the scaling exponent z in

$$T = a\,L^{z} \tag{4.3}$$

where T is time scale L is space scale and a is a parameter. In Blöschl and Sivapalan's (1995) diagram z is about 0.9 for most processes which implies slightly increasing velocities with scale. They argued that these may result from a slight tendency of channel flow velocities to increase with catchments scale (Leopold and Maddock, 1953) and the behaviour of response times for catchments of different sizes (Anderson and Burt, 1990). It is interesting to compare the diagram of Blöschl and Sivapalan (1995) to a quantitative analysis of space-time rainfall variability of Foufoula-Georgiou and Vuruputur (2000) based on radar images of individual rainstorms. They examined fluctuations of precipitation in time and calculated standard deviations in space for different aggregation levels in space, and derived space-time pairs of constant standard deviations. To these pairs they fitted a relationship of the form of Eq. 4.3 and obtained values of z = 0.5-0.6, i.e. significantly smaller values than those of Blöschl and Sivapalan (1995).

While characteristic scales are usually defined for *point* processes in the literature it is also possible to examine characteristic scales of aggregated processes by treating runoff from catchments of a given size as a process associated with a particular characteristic scale. Provided runoff can be assumed to be a spatially homogeneous process the characteristic

scale will be intrinsic to one particular catchment size but may be different for different catchment sizes, similar to the case of increasing support in Chapter 2. It is interesting to examine the effect catchment size has on characteristic scales, i.e. the effect of upscaling from small to large catchments. As one moves up in catchment scale, more and more spatial rainfall variability tends to get averaged out due to aggregation effects which will be reflected in a reduction of the variance of runoff and a tendency for increased characteristic scales (Western and Blöschl, 1999). As one moves up in catchment scale, it is also likely that temporal variability decreases and characteristic temporal scales increase because of longer flow paths. However, as with other hydrologic characteristics such as extreme precipitation and floods, it is not quite clear what controls this variance reduction and the changes in characteristic scales as a function of catchment size (Blöschl and Sivapalan, 1997, Sivapalan and Blöschl, 1998). It is important to unravel these controls in order to shed light on the space-time scaling behaviour of catchments and to better understand the role of characteristic scales for catchments of different size.

In the light of the above discussion the objectives of this chapter are twofold. The first is to estimate characteristic space and time scales for the main variables of the hydrological cycle (precipitation, runoff, groundwater level, soil moisture) or, alternatively, identify their fractal characteristics if no characteristic scales exist. The second is to examine and understand the effect catchment size has on characteristic scales for the case of precipitation and runoff. Both objectives are addressed by extensive variogram analyses of hydrologic data.

4.2 Data

Most of the data used in this chapter stem from a comprehensive hydrographic data set of Austria. Austria has a varied climate with mean annual precipitation ranging from 500 mm in the eastern lowland regions up to about 3000 mm in the western alpine regions. Runoff depths range from less than 50 mm per year in the eastern part of the country to about 2000 mm per year in the Alps. Potential evapotranspiration is on the order of 600-900 mm per year. Mean daily runoff data for the period 1971-1997 were used. Stream gauges with short records and catchments with significant anthropogenic effects and lake effects were excluded from the data set which gave a total of 654 stream gauges available for the analysis. The runoff series were divided into three classes according to catchment size – small (3 - 70 km^2), medium (70

- 250 km²) and large (250 - 130 000 km²) (Figure 4.1b). The analyses were carried out for each class separately. Some of the catchments are nested. Initial test analyses showed that when excluding nested catchments from the data set, the variograms changed only very slightly, therefore nested catchments were retained in the analysis. Daily precipitation data from 991 stations for the period 1971-1997 were used (Figure 4.1a). In order to be able to examine spatial aggregation effects, for each day catchment precipitation was calculated by external drift kriging interpolation of the point data for a total of 756 catchments. Catchment precipitation series were again divided into three size classes. In addition, one series of precipitation from the Frankenfels precipitation station with hourly data during a period of 7 years was used to look at the effect of temporal aggregation. Weekly groundwater levels were used. Short records were excluded which left a total of 3539 groundwater stations for the analysis. These data stem from porous aquifers that are mainly located in the valley regions (Figure 4.1c).

A similarly comprehensive soil moisture data set is not available in Austria. For comparison and to complement the Austrian data, soil moisture data from Australian and New Zealand catchments have been used. The climate in these catchments is not radically different from the Austrian catchments, at least from the lower parts of Austria. The Australian and New Zealand sites have 800-1600 mm mean annual precipitation, 0-600 mm mean annual runoff depth and potential evapotranspiration on the order of 800-900 mm per year. Also, hillslope scales do not vary much from the Austrian landscape, so the characteristic scales should be similar within the accuracies of the overall analysis. The soil moisture data stem from five small catchments, Tarrawarra, Satellite Station, Point Nepean, Claydens, and Carrans (Western and Grayson, 1998, Woods et al., 2001). The sizes of these catchments range from 0.1 to 1.5 km². The soil moisture data consist of separate temporal and spatial data sets of depth-average volumetric soil moisture in the 0-30 cm layer. The temporal data were sampled at 19 different locations with 30 min intervals over three years, using Campbell Scientific CS615 sensors. The spatial data have been sampled at between 285 - 589 points in each catchment on a regular grid using Time Domain Reflectometry. These patterns were collected on between six and eight occasions for each catchment. Table 4.1 summarises the data used in this chapter.

Table 4.1 Data series used in this chapter.

Data type	Size class	Size range (km^2)	Median size (km^2)	Number of stations	Extent of domain (km)	Time resolution	Period
Point precipitation	-	Point	Point	991	700	Daily sum	1971-1997
Catchment precipitation	Small	0.5-80	36	252	700	Daily sum	1971-1997
	Medium	80-273	135	252	700	Daily sum	1971-1997
	Large	277-131000	818	252	700	Daily sum	1971-1997
Point precipitation	-	Point	Point	1	-	Hourly sum	1988-1994
Runoff	Small	2.8-74	36	218	700	Daily average	1971-1997
	Medium	74-252	128	218	700	Daily average	1971-1997
	Large	253-131000	701	218	700	Daily average	1971-1997
Groundwater level	-	Point	Point	3539	700	Weekly instantaneous	1966-1998
Soil moisture time series	-	Point	Point	19	-	Half hour instantaneous	1998-2001
Soil moisture patterns	-	Point	Point	5*app. 400	0.3-1.4	6-8 occasions, instantaneous	1998-2000

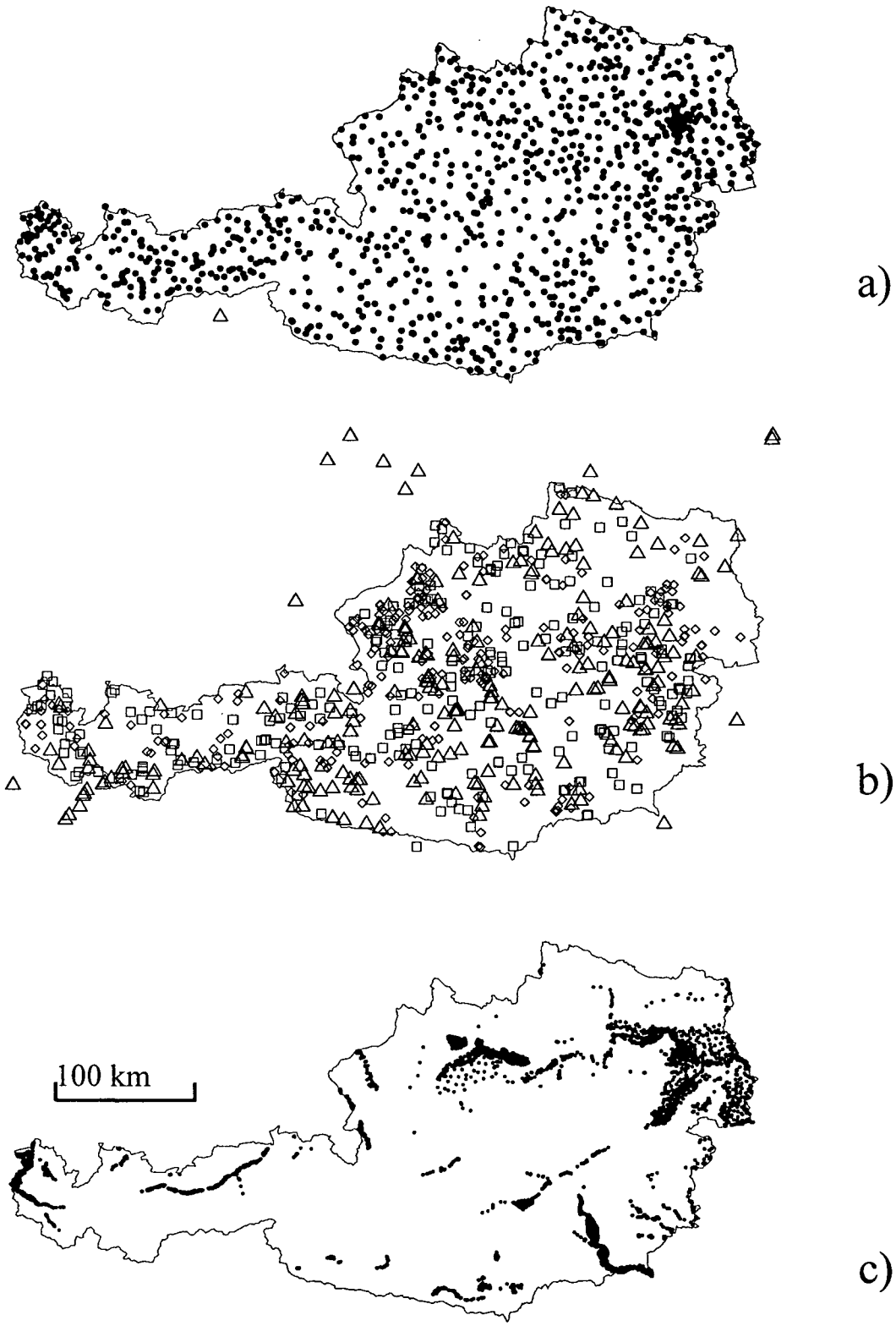


Figure 4.1 Network of measurement stations in Austria used in this chapter. (a) Precipitation gauges; (b) centroids of gauged catchments (small catchments shown as diamonds, medium sized catchments shown as squares, large catchments shown as triangles); (c) Boreholes for groundwater levels.

4.3 Method

4.3.1 Transformation of data

When performing a variogram analysis, it is useful to first transform the data as, ideally, the analysis should be done on a random variable that exhibits little periodicity with a distribution close to normal (e.g. Sokal and Rohlf, 1995). A number of transformations were therefore examined to render the data more suitable to variogram analysis. For runoff and precipitation, four steps of transformation were examined, each of them being a transformation of the previous one:

- Untransformed data (specific runoff ($m^3s^{-1}km^{-2}$)), precipitation (mm/day))
- Logarithms of data series
- Seasonal effects removed from the above series
- Mean of each of the time series subtracted from the above (hereafter referred to as case I)
- The above series divided by their standard deviation (hereafter referred to as case II)

The last transformation produces time series with zero mean and unity standard deviation. For groundwater levels, three steps of transformation were examined:

- Untransformed data (groundwater levels in m)
- Mean subtracted from time series to remove the elevation effects
- Seasonal effects removed from the above series (hereafter referred to as case I)
- Above series divided by their standard deviation (hereafter referred to as case II)

Again, the last transformation produces time series with zero mean and unity standard deviation. For soil moisture, two steps of transformation were examined:

- Untransformed data (soil moisture content in percent of total volume)
- Mean over 6-8 occasions of each sampling point subtracted from the spatial data to remove spatial organisation; Seasonal effects removed from the time series (hereafter referred to as case I)

- Spatial data (as above) divided by their temporal standard deviation for each sampling point; Time series (as above) divided by their temporal standard deviation (hereafter referred to as case II)

The last transformation produces spatial patterns of soil moisture for which each point has zero temporal mean and unity temporal standard deviation, and time series with zero mean and unity standard deviation.

For precipitation, the logarithmic transformation is not directly possible, as the logarithm of zero is undefined. Two possible alternatives were considered, treating zero precipitation as no data and padding zero values with the smallest measured precipitation (0.1 mm/day). The latter option was chosen as it provides information on the time scales of the interstorm periods. Initial test analyses indicated that the absolute gamma values will be higher when zero is treated as no data compared to padding them. This is because there are less pairs in time or space with equal (zero precipitation) values. However, the shape of the variogram did not change much. In the hourly data series zero precipitation was padded with 0.0041 mm/h which is equivalent to 0.1 mm/day. Zero runoff occurred only at a few instances and was treated as no data.

The seasonal components of precipitation, runoff and groundwater levels were estimated by a standard method of moving average estimation (MAE) as used for example by De Cesare et al. (2002). Firstly, the average values of each of the time series were calculated for each day of the year. These average values were smoothed by using a moving average procedure with a window of 21 days (precipitation and runoff) and 49 days (groundwater levels). The seasonally smoothed mean for a certain day of year was calculated as:

$$\overline{us(i)} = \frac{1}{(2 \cdot iavg + 1)} \sum_{j=-iavg}^{iavg} \frac{1}{n_{ij}} \sum_{k=0}^{n_{ij}-1} u(i + j + \mathrm{int}(365.25k)) \qquad (4.4)$$

where i is the day of the year, $iavg$ is the number of days in the past and future used in the moving averaging window, $2 \cdot iavg+1$ is the averaging window size, n_{ij} is the number of years with a measurement at day $i+j$ of the year, 365.25 is the average number of days in a year, leap-years included, and k is the year.

The seasonal components of the soil moisture time series were obtained by a Fourier fitting procedure. First the average value of each time series was calculated for each day and time of the year. These averages were then fitted by a two-harmonic Fourier curve. A Fourier curve was used in preference to the moving average described above due to additional variability

associated with the shorter time series and the strong link between the soil moisture seasonality and seasonality of potential evapotranspiration associated with changing global radiation, which follows a sinusoidal pattern.

Although the variograms change with the different transformations, they are mostly within an order of magnitude. For runoff and precipitation the data processed by the last two transformations exhibit little skewness and the deterministic or periodic parts in time and space have been removed. The focus will therefore be on these two cases (cases I and II) in the rest of this chapter. The groundwater level and soil moisture data exhibit only very minor skewness and so the logarithmic transformation was not needed. However, it was deemed appropriate to remove the deterministic or periodic parts in time and space, so again the focus in this chapter will be on cases I and II.

The difference between cases I and II for all data types is that in case II the temporal variance is always unity while this is not the case for case I. Both cases were examined as for some interpretations it is useful to keep the relative variances in space and time (e.g. for finding space-time pairs of scales) while for other interpretations (e.g. the shape of the variograms) normalisation is appropriate.

4.3.2 Variograms

The temporal variograms were calculated as the average of the temporal variograms of the individual time series:

$$\hat{\gamma}_t(h_t) = \frac{1}{\sum_{j=1}^{m} 2n_j(h_t)} \sum_{j=1}^{m} \sum_{i=1}^{n_j(h_t)} (u(\mathbf{x}_j, t_i + h_t) - u(\mathbf{x}_j, t_i))^2 \qquad (4.5)$$

where h_t is the time lag, $u(\mathbf{x}_j, t_i)$ is the value of the transformed variable at time t_i and spatial location $\mathbf{x}_j$ of station j, m is the number of data series, and $n_j(h_t)$ is the number of pairs with time lag h_t. For the longest time lags the data were not exhaustively sampled to reduce computation time. Similarly, the spatial variograms were calculated as the average of the spatial variograms of the individual points in time:

$$\hat{\gamma}_s(h_s) = \frac{1}{\sum_{i=1}^{m'} 2n_i(h_s)} \sum_{i=1}^{m'} \sum_{j=1}^{n_i(h_s)} (u(\mathbf{x}_j + \mathbf{h}_s, t_i) - u(\mathbf{x}_j, t_i))^2 \qquad (4.6)$$

where $h_s = |\mathbf{h}_s|$ is the space lag, $u(\mathbf{x}_j, t_i)$ is the value of the transformed variable at time t_i and spatial location $\mathbf{x}_j$ of station j, m' is the number of time steps, and $n_i(h_s)$ is the number of pairs with space lag h_s.

A variogram model with four parameters was fitted visually to each sample variogram. The model is a combination of a fractal model and a Weibull type model:

$$\gamma(h) = ah^b(1-\exp(-(h/c)^d)) \tag{4.7}$$

where h is lag (in either space or time) and a, b, c and d are the parameters. On a double logarithmic plot this variogram model consists of two straight lines with a curve in between. It is therefore possible to determine separately the slopes of the variogram for short lags and long lags. The slope at short lags is $b+d$, while the slope at long lags is b. Both slopes can be considered as a fractal power of the variogram but for different lag ranges. The exponential part reaches a sill after a lag dependent on c, so c is a parameter controlling the space (or time) scale. a controls the overall level (or variance) of the variogram.

For stationary or approximately stationary variograms (where b is close to zero) a characteristic length or time scale can be defined. One common possibility in geostatistics is to use the range, usually taken as the lag where the variogram reaches 95% of its maximum value (Webster and Oliver, 2001). An alternative is to use the e-folding distance in space or time. The e-folding distance is defined as the distance (in space or time) where the variogram value is $1-1/e$ of its maximum value. e is the Eulerian constant (2.718). For an exponential variogram ($b=0$ and $d=1$) in Eq. 4.7 the e-folding distance is one third of the range. If the variogram is not stationary it is still possible to calculate the range and the e-folding distance from the data, but both will depend on the maximum lag over which the data have been sampled. As the estimates of the range are much more sensitive to the fitting procedure than those of the e-folding distance, particularly if the variograms are not strictly stationary, the e-folding distance is used as a measure of the characteristic space and time scales in this chapter. The e-folding distance was found as

$$\gamma(\text{e - fold}) = \left(1-\frac{1}{e}\right)\sigma^2 \tag{4.8}$$

where γ is the fitted variogram model in Eq. 4.7 and σ^2 is the sample variance in either space or time. The sample variance in time was found as the average variance of each of the

time series. The sample variance in space was found as the average spatial variance for each time step.

Some of the parameters of the variograms are expected to be associated with significant uncertainty even though for most instances the sample size is relatively large. Specifically, if either the small scale fractal power, $b+d$, or the large scale fractal power, b, are estimated from a short range of lags, one would expect them to be poorly defined (see e.g. Gallant et al., 1994, Katsev and L'Heureux, 2003). If the fitted variograms are not extrapolated much beyond the range supported by the data, a lack of identifiability will not result in significantly different variograms as alternative models are similar over that range, but it is still very useful to quantify the uncertainties associated with the estimated parameter values, particular if one is to interpret the estimated parameters in terms of their scaling behaviour. A split sample analysis was therefore performed where variograms of independent subsets of the data were compared. For each catchment size class of runoff, the entire data set of 218 catchments (Table 4.1) was split into five non-overlapping groups of about 44 catchments. In a similar way the time series of catchment precipitation, point precipitation and groundwater levels were split into five groups of approximately equal size. It was not possible to perform the split sample analysis for soil moisture and the hourly rainfall data because of the limited sample size. Variogram models (Eq. 4.7) were then fitted to the sample variograms of each group by an automatic fitting procedure based on weighted least squares of the logarithmic variograms. This procedure resulted in five estimates of each parameter value in each catchment size class. The variance of each parameter value between the five groups, σ_b^2, was calculated. If the five groups produced similar values of a particular parameter then σ_b^2 was small and hence the uncertainty of that parameter was small. Conversely, large σ_b^2 indicated large uncertainties. σ_b^2 can therefore be considered an estimate of the error variance of the parameter estimate from one group. The error variance associated with the parameter estimates of the entire data set will be smaller than that of each group. σ_e^2 from Eq. 4.9 was used as a measure of the error variance:

$$\sigma_e^2 = \frac{1}{m}\sigma_b^2 \tag{4.9}$$

where m is the number of groups, i.e. m=5. For the sample mean, Eq. 4.9 is strictly applicable as the samples are independent. For the parameters of the variogram (Eq. 4.7), the e-folding distance and the sample variance Eq. 4.9 is an approximation. Alternative group numbers

were examined, which all resulted in similar estimates of σ_e^2. Eq. 4.9 is therefore a reasonable approximation.

4.3.3 Regularisation – spatial aggregation

In order to study the spatial aggregation behaviour of precipitation and runoff, the sample variograms for catchment precipitation and runoff were compared to aggregated or block variograms calculated by regularisation from point variograms. The idea of regularisation is that the variogram will change its shape when one aggregates or filters the original data. Specifically, the overall level of the variogram (parameter a in Eq. 4.7) will decrease and the e-folding distance will increase with increasing aggregation area. Standard procedures for calculating the regularised variograms exist (e.g. Isaaks and Srivastava, 1989). Here it is used a simple procedure proposed by Sivapalan (1986) and used by Western and Blöschl (1999) and others which builds on the probability density functions (pdf) of the distances within the aggregation area to calculate the aggregated variogram:

$$\gamma_{agg}(h,a) = \int_0^{r\max} \gamma(r) \cdot f_2(r|h,a)dr - \int_0^{s\max} \gamma(s) \cdot f_1(s|a)ds \qquad (4.10)$$

where h is the lag, a is the side length of the aggregation area, r is the distance between two randomly chosen points in each of two aggregation areas A that are separated by a (centre-to-centre) distance h, r_{max} is the maximum of r, f_2 is the pdf of r, γ is the point variogram, s is the distance between two randomly chosen points in one aggregation area A, s_{max} is the maximum of s, f_1 is the pdf of s. f_1 and f_2 are functions of the geometry of the aggregation area. For a square aggregation area analytical expressions were found by Sivapalan (1986) and Ghosh (1951), and also presented in Chapter 2. Sivapalan (1986) showed that the aggregated variogram is only moderately sensitive to the shape of the aggregation area, so a square shape should also be applicable to natural catchments. The analytical expressions do not cover the case of overlapping squares (i.e. $h < a$) for which the integral was found by numerical integration. In calculating the regularised variograms for catchment precipitation and runoff a side length of the aggregation area $a = \sqrt{A}$ was assumed, where A is the median area of the catchments in each size class as given in Table 4.1. Point variograms derived from untransformed precipitation data was used, because the aggregation of precipitation by catchments is consistent with mass conservation, a requirement that is fulfilled by aggregating

the untransformed data but not fulfilled by the logarithmic data. The regularised variograms were then compared with variograms directly calculated from (untransformed) catchment precipitation for each of the size classes.

For runoff, point time series were not available as the minimum size of gauged catchments in the data set was a few square kilometres. A hypothetical point runoff variogram was therefore back-calculated in the following way. It can be assumed that the point variogram is represented by Eq. 4.7. For a given parameter set the regularised variograms for the three catchment size classes were obtained in a similar fashion as for precipitation. The point variogram parameters were then optimised by fitting the regularised runoff variograms to the respective sample runoff variograms for all three catchment size classes simultaneously. One would expect a simultaneous fit to all three variograms to yield a more robust estimate of the point variogram than a fit to a single variogram. The objective function to be minimised consisted of a weighted sum of the squared differences of the logarithms of the sample variograms and the regularised variograms.

4.3.4 Space – time links

In order to examine the space-time behaviour of characteristic scales the spatial variograms were related to the temporal variograms. Specifically, the main assumption here was to assume that space-time scale pairs characteristic of a process are those scales where the spatial variogram value is equal to the temporal variogram value, i.e.

$$\gamma_t(T) = \gamma_s(L) \tag{4.11}$$

where T is time scale, L is space scale, and the γ_t and γ_s are the temporal and spatial variograms, respectively. The space and time scales were then plotted against each other, which gave traces of space-time characteristic scales. If the trace plots as a straight line in a double logarithmic plot, space and time scales are related by a power law (Eq. 4.3). The ratio $V = L/T$ is a characteristic velocity, which is a measure of the scales over which measurements become decorrelated from one another. It may be related to a number of physical causes such as the celerity with which a perturbation moves through a system, to particle velocities or to the effects of internal dispersion. Two points separated by a spatial lag L are able to see the same perturbation only if the length of the perturbation is at least L. Two points separated by a time T are only able to see the same perturbation if it is moving with a

velocity less than $V = L/T$. The correlation in time and space is then related to the possibility that two points separated by a time T or a distance L are able to see the same perturbation. As a natural system consists of a spectrum of small to large perturbations, this possibility will decrease with increasing lags in space and time as the number of perturbations visible to both points is decreasing. The gamma values of the variogram can in a similar way be seen as related to the number of perturbations not visible to both of the points. This applies to both stationary and non-stationary processes. Equating the gamma values in time and space for different gamma values can then be seen as finding the average velocity for a perturbation of size L. It is easy to illustrate this for precipitation. A front can be seen as a large scale perturbation in time and space, while a thunderstorm is a smaller scale perturbation. The temporal and spatial variograms of precipitation reflect the possibility of seeing these perturbations. Precipitation is also one of the main perturbations for the other processes examined in this chapter. Additionally, evaporation and snow deposition and melting modulate the dynamics. These perturbations are then filtered by the sub-processes at and under the surface until they reach the catchment outlet where the water flux is actually measured, removing the small scale components of the perturbations. The variogram still shows the possibility of seeing a perturbation at points separated in time or space, while the velocity found from the method explained above shows the velocities of the filtering processes. Foufoula-Georgiou and Vuruputur (2000) derived space-time pairs of characteristics scales from iso-lines of constant standard deviations for different aggregation levels in space of rainfall fluctuations. This is very similar to Eq. 4.11 as the variogram can be interpreted as the lag dependent variance of the fluctuations of a variable. The difference is that Foufoula-Georgiou and Vuruputur used variance as a function of aggregation area in space and as a function of lag in time while in this chapter the variance was examined as a function of lag in both space and time.

Spatial and temporal counterparts of the variograms for each data type were compared, e.g. the spatial variograms of runoff from small catchments to the temporal variograms of runoff from small catchments. The exception is hourly precipitation where spatial variograms were unavailable, so the temporal hourly variogram was compared with the spatial variogram from daily point precipitation. The rainfall regime is similar which justifies this comparison. However, because of the different time resolutions one would expect some bias to be introduced.

4.4 Results

4.4.1 Precipitation

Spatial and temporal variograms of point precipitation and catchment precipitation for three size classes are shown in Figure 4.2a (case I) and b (case II). The temporal variogram for hourly precipitation is also included. The variograms of cases I and II are very similar. The only significant difference is the overall level of the variograms which is a result of the normalisation in case II. Tables 4.2 and 4.3 (non-italic numbers) give the fitted variogram parameters (b, d) for cases I and II, respectively, as well as the associated e-folding distances and variances. The italic numbers in Tables 4.2 and 4.3 are the standard deviations σ_e of the estimation errors of the parameters.

Figure 4.2 shows that the temporal variograms are stationary and the e-folding distance of all daily variograms (point, small, medium and large catchments) is about one day. This means that very little temporal correlation exists beyond one day. The variogram for hourly precipitation has a significantly shorter e-folding distance (only 6 hours). Given that the station used for the hourly variogram has a similar rainfall regime as the daily stations, this difference suggests that the e-folding distance from the daily variograms may be biased as a function of the temporal support (see section 2.2.4.1). The daily sums of precipitation are unable to resolve the shorter time scale fluctuations. On closer examination of the daily variograms for catchments of different sizes (Tables 4.2 and 4.3) there appears a slight tendency for the e-folding distance to increase with catchment size.

In contrast, the spatial variograms for precipitation are all non-stationary. They appear to consist of two fractal parts. At large scales, the fractal power (i.e. the slope in the double logarithmic plot) is on the order of 0.2 (b in Tables 4.2 and 4.3) while at small scales the fractal power varies between 0.5 and 1.55 ($b+d$ in Tables 4.2 and 4.3). There is a clear difference between the spatial variograms of point precipitation and those of catchment precipitation of the different size classes. Point precipitation has the smallest slope at small scales ($b+d = 0.5$) while catchment precipitation for large catchments has the steepest slope at small scales ($b+d = 1.55$). The other catchment size classes are in between. These differences with catchment size translate into differences in the e-folding distances which range from about 45 km (point precipitation) to about 65 km (catchment precipitation for large catchments). This difference in the e-folding distances is larger than the estimation standard deviation, so one would expect it to be significant. There is also a significant decrease of the

spatial variance and hence the overall level of the variograms with catchment size (Tables 4.2 and 4.3). All of these changes of the spatial variograms with catchment size are related to aggregation effects as will be discussed later in this chapter in the context of regularisation. The temporal variance is significantly larger than the spatial variance (Table 4.2).

Table 4.2 Parameters of the fitted space and time variograms as in Eq. 4.7 and variance (non-italic numbers). Transformation case I. *b+d* is the fractal power at small scales, *b* is the fractal power at large scales, σ^2 is the variance in either space or time, e-fold is the e-folding distance which is a measure of the characteristic scale in either space or time. The italic numbers are the standard deviations σ_e of the estimation errors of the parameters as of Eq. 4.9.

Process		Space				Time			
	Size	*b+d*	*b*	σ^2	e-fold (km)	*b+d*	*b*	σ^2	e-fold (days)
Precipitation	Point	0.50	0.15	1.67	47	0.85	0.00	3.41	0.83
		0.01	*0.02*	*0.01*	*1.4*	*0.01*	*0.00*	*0.01*	*0.00*
	Small	1.06	0.21	1.39	59	0.85	0.00	3.64	0.93
		0.06	*0.06*	*0.02*	*1.2*	*0.00*	*0.00*	*0.03*	*0.01*
	Medium	1.33	0.23	1.39	66	0.85	0.00	3.51	1.00
		0.07	*0.02*	*0.02*	*1.3*	*0.00*	*0.00*	*0.04*	*0.01*
	Large	1.54	0.25	1.14	67	0.85	0.00	3.32	1.04
		0.09	*0.06*	*0.04*	*3.2*	*0.01*	*0.00*	*0.03*	*0.01*
	Hourly	-	-	-	-	0.65	0.00	3.25	0.3
		-	-	-	-	-	-	-	-
Runoff	Small	0.58	0.08	0.30	18	0.61	0.01	0.45	13
		0.12	*0.04*	*0.01*	*3.3*	*0.01*	*0.01*	*0.02*	*1.0*
	Medium	0.68	0.08	0.23	39	0.61	0.01	0.34	17
		0.08	*0.03*	*0.02*	*2.8*	*0.01*	*0.01*	*0.02*	*1.0*
	Large	0.93	0.08	0.15	62	0.66	0.01	0.23	18
		0.04	*0.03*	*0.00*	*1.9*	*0.02*	*0.01*	*0.01*	*1.2*
Groundwater		0.62	0.12	0.25 m^2	8.3	0.82	0.12	0.29 m^2	211
		0.09	*0.04*	*0.01 m^2*	*1.0*	*0.01*	*0.01*	*0.02 m^2*	*22*
Soil moisture		0.5	0.2	6.5	0.068	0.99	0.04	1.0	9.2
		-	-	-	-	-	-	-	-

Table 4.3 Parameters of the fitted space and time variograms as in Eq. 4.7 and variance (non-italic numbers). Transformation case II. *b+d* is the fractal power at small scales, *b* is the fractal power at large scales, σ^2 is the variance in either space or time, e-fold is the e-folding distance which is a measure of the characteristic scale in either space or time. The italic numbers are the standard deviations σ_e of the estimation errors of the parameters as of Eq. 4.9.

Process		Space				Time			
	Size	*b+d*	*b*	σ^2	e-fold (km)	*b+d*	*b*	σ^2	e-fold (days)
Precipitation	Point	0.50	0.15	0.48	45	0.85	0.00	1.00	0.80
		0.01	*0.01*	*0.00*	*0.9*	*0.01*	*0.00*	*0.00*	*0.00*
	Small	1.06	0.21	0.38	53	0.85	0.00	1.00	0.90
		0.04	*0.05*	*0.01*	*0.9*	*0.00*	*0.00*	*0.00*	*0.01*
	Medium	1.33	0.23	0.40	63	0.85	0.00	1.00	0.95
		0.04	*0.02*	*0.01*	*1.1*	*0.00*	*0.00*	*0.00*	*0.01*
	Large	1.55	0.25	0.34	65	0.85	0.00	1.00	1.02
		0.09	*0.06*	*0.01*	*2.3*	*0.01*	*0.00*	*0.00*	*0.01*
	Hourly	-	-	-	-	0.65	0.00	1.00	0.29
		-	-	-	-	-	-	-	-
Runoff	Small	0.47	0.06	0.68	28	0.61	0.01	1.00	16
		0.05	*0.03*	*0.02*	*2.6*	*0.01*	*0.01*	*0.00*	*1.1*
	Medium	0.57	0.06	0.66	42	0.61	0.01	1.00	19
		0.11	*0.03*	*0.02*	*3.7*	*0.01*	*0.01*	*0.00*	*0.8*
	Large	0.77	0.07	0.63	59	0.61	0.01	1.00	20
		0.08	*0.03*	*0.01*	*1.5*	*0.03*	*0.01*	*0.00*	*0.9*
Groundwater		0.76	0.11	0.81 m^2	6.5	0.85	0.15	1.00 m^2	145
		0.12	*0.03*	*0.01 m^2*	*1.1*	*0.01*	*0.00*	*0.00 m^2*	*5.0*
Soil moisture		0.5	0.2	0.13	0.062	0.99	0.04	1.00	11
		-	-	-	-	-	-	-	-

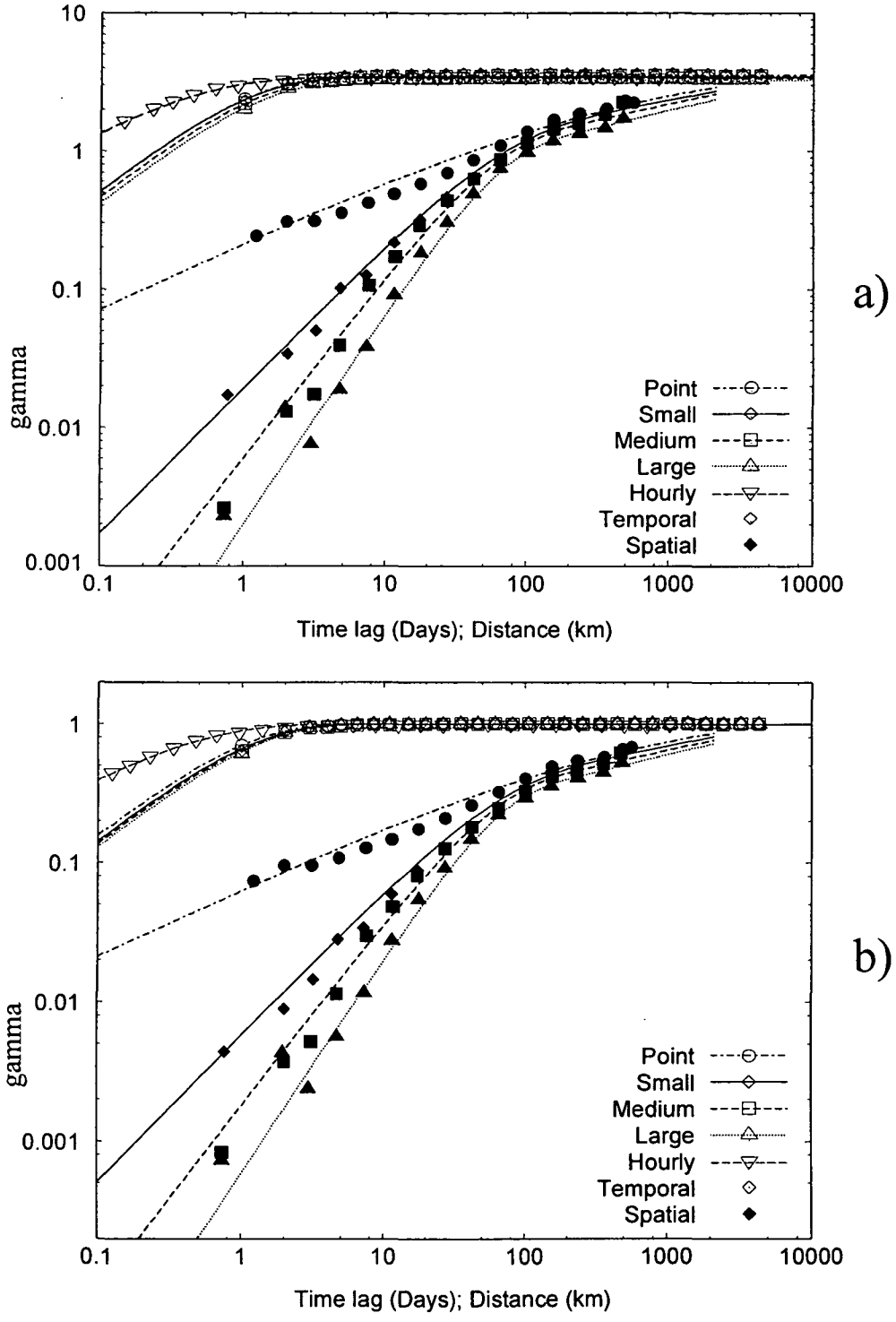


Figure 4.2ab Variograms of precipitation. Spatial sample variograms shown as full symbols and temporal sample variograms shown as open symbols. Point precipitation shown as circles, small catchments as diamonds, medium sized catchments as squares, large catchments as upward pointing triangles. Temporal variogram of hourly precipitation shown as downward pointing triangles. Fitted variograms are shown as lines: point precipitation as dashed-dotted lines, small catchments as solid lines, medium catchments as dashed lines, large catchments as dotted lines, hourly temporal variogram as a long dashed line. (a) shows case I and (b) shows case II (see text).

4.4.2 Runoff

Figure 4.3a (case I) and b (case II) shows spatial and temporal variograms of runoff. Unlike the variograms for precipitation there are some differences between cases I and II, although the main difference is the overall level of the variograms which is a result of the normalisation in case II. The resulting parameters for cases I and II are given in Tables 4.2 and 4.3, respectively.

The temporal variograms for different catchment sizes are almost stationarity (b is very small) and the e-folding distances are about two to three weeks. There is a slight tendency for the e-folding distance to increase with catchment size which appears to be significant given that the estimation errors are small (Tables 4.2 and 4.3). This implies that larger catchments tend to respond more slowly to precipitation forcing than smaller catchments. This is certainly consistent with streamflow analyses in various climates (e.g. Melone et al., 2002). The main difference of the temporal variograms of the different size classes is their overall level or temporal variance. The temporal variance significantly decreases with catchment size, from 0.44 for small catchments to 0.23 for large catchments. This means that the variability of runoff depth strongly decreases with catchment size. Again, this is consistent with common experience as extreme events (floods and droughts) certainly tend to become more moderate with catchment size if measured in terms of runoff depths or runoff per unit catchment area.

The spatial variograms (Figure 4.3a and b) for runoff are all non-stationary and appear to consist of two fractal parts although the large scale part is not as well defined as with precipitation (Figure 4.2). At large scales, the fractal power is about 0.07 (b in Tables 4.2 and 4.3). The associated estimation errors are relatively large (about 0.03). At small scales the fractal power varies between about 0.6 and 0.9 (case I) and 0.5 and 0.8 (case II) ($b+d$ in Tables 4.2 and 4.3) and is better defined than at the large scales. There is a clear difference between the spatial variograms for the catchments of the different size classes with the larger catchments exhibiting larger fractal power or steeper slopes at small scales than the smaller catchments. These differences with catchment size translate into significant differences in the e-folding distance which range from about 20 km (small catchments) to about 60 km (large catchments). These differences in the e-folding distances are larger than those for catchment precipitation where the e-folding distance increased by about 20% when moving from small to large catchments while the e-folding distance tripled in the case of runoff. These changes with catchment size are clearly related to aggregation effects. It appears, however, that catchment runoff exhibits a stronger aggregation behaviour than catchment precipitation. This

will be discussed later in this chapter in the context of regularisation. Similarly to the temporal variograms, the overall level of the spatial variograms or spatial variance decreases significantly with catchment size, from 0.30 for small catchments to 0.15 for large catchments (case I) which is, again, related to aggregation.

The temporal variance is larger than the spatial variance (Table 4.2) but in comparison with precipitation the difference is smaller. It is also interesting that the spatial variogram of runoff varies by one and a half orders of magnitude (from 0.05 to 1 in Figure 4.3b) while the spatial variogram of catchment precipitation varies by almost three orders of magnitude (from 0.001 to 1 in Figure 4.2b). This suggests that the variability of runoff increases less rapidly with spatial scale than the variability of catchment precipitation which is consistent with the smaller fractal powers of runoff.

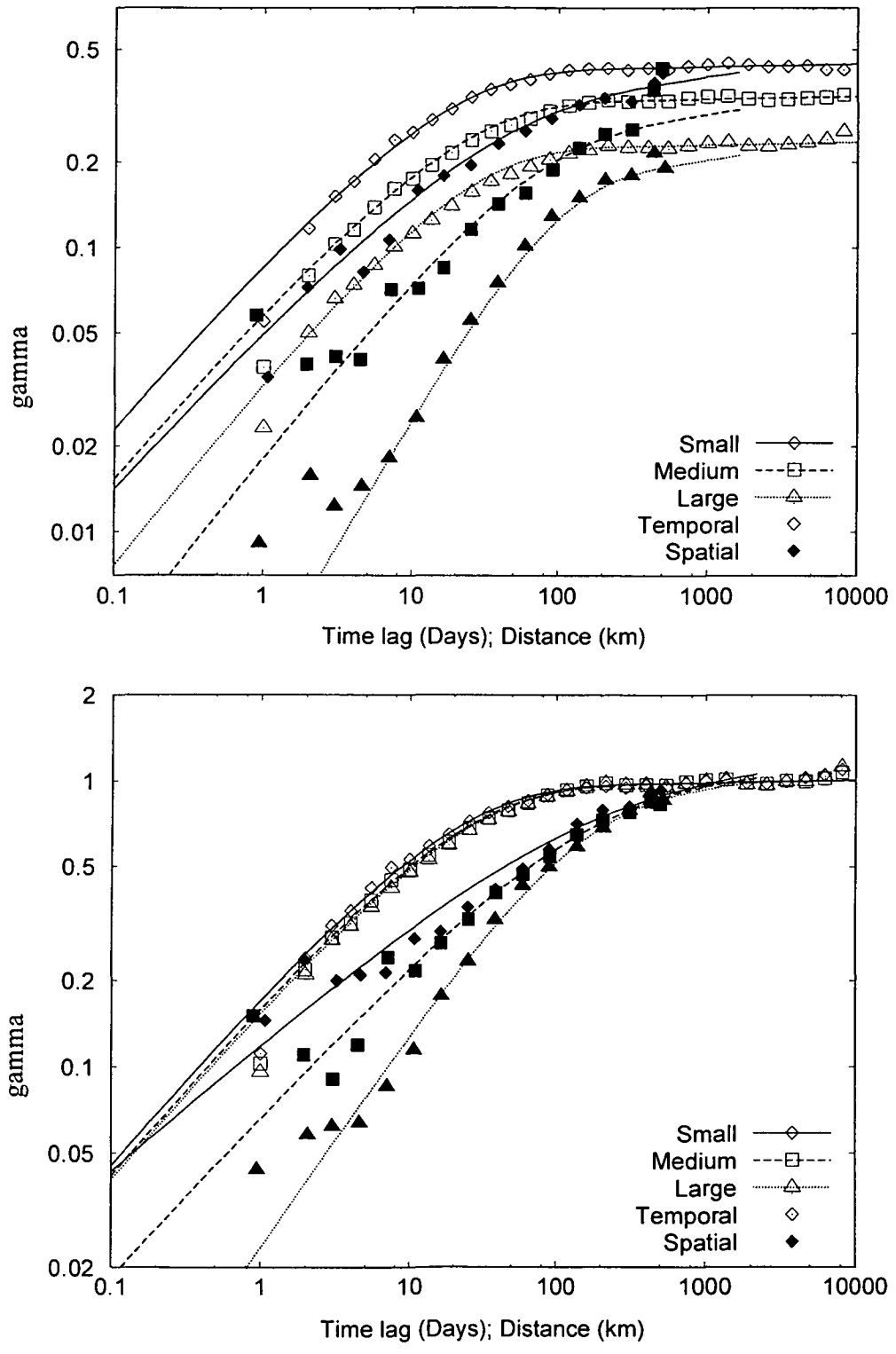


Figure 4.3ab Variograms of runoff. Spatial sample variograms shown as full symbols and temporal sample variograms shown as open symbols. Small sized catchments shown as diamonds, medium sized catchments as squares, large catchments as triangles. Fitted variograms are shown as lines: small catchments as solid lines, medium catchments as dashed lines, large catchments as dotted lines. (a) shows case I and (b) shows case II (see text).

4.4.3 Groundwater levels

Figure 4.4a (case I) and b (case II) shows spatial and temporal variograms of groundwater levels and the resulting parameters are given in Tables 4.2 and 4.3. The two cases are similar with the exception of the fluctuation of the spatial variogram at large space lags for case I. These fluctuations are an indication of spatial periodicity but disappear when normalising the data series (Figure 4.4b). This periodicity is most likely related to the geographic position of porous aquifers in Austrian valleys, some of which are regularly spaced (Figure 4.1).

The temporal variograms are clearly non-stationary and consist of two fractal parts. At large scales, the fractal power is about 0.1 (b in Tables 4.2 and 4.3) while at small scales the fractal power is about 0.7 ($b+d$ in Tables 4.2 and 4.3). The e-folding distance in time is on the order of half a year. This very long characteristic time scale is to be expected as groundwater is of course a much more slowly varying process than precipitation and runoff. The temporal non-stationarity is also a reflection of the presence of long time scale fluctuations.

The spatial variograms of groundwater are also non-stationary, as seen in Figure 4.4. The shapes of the spatial variograms are similar to the shapes of the spatial variograms of catchment precipitation and runoff with two fractal parts. However, the turning point appears to occur at smaller space scales (about 5 km, Figure 4.4b) than with catchment precipitation and runoff (about 100 km, Figures 4.2b and 4.3b). This translates into a relatively small e-folding distance in space of only about 7 km in the case of groundwater levels (Tables 4.2 and 4.3). This means that groundwater level fluctuations in space are a relatively small scale process as compared to precipitation and runoff. However, the large scale fractal power is somewhat larger than that of runoff, i.e. groundwater levels are slightly more non-stationary than runoff in space.

The temporal variance is slightly larger than the spatial variance (Table 4.2) but in comparison with precipitation and runoff the difference is much smaller. This, again, is related to the very slow dynamics of groundwater systems relative to other components of the hydrologic cycle.

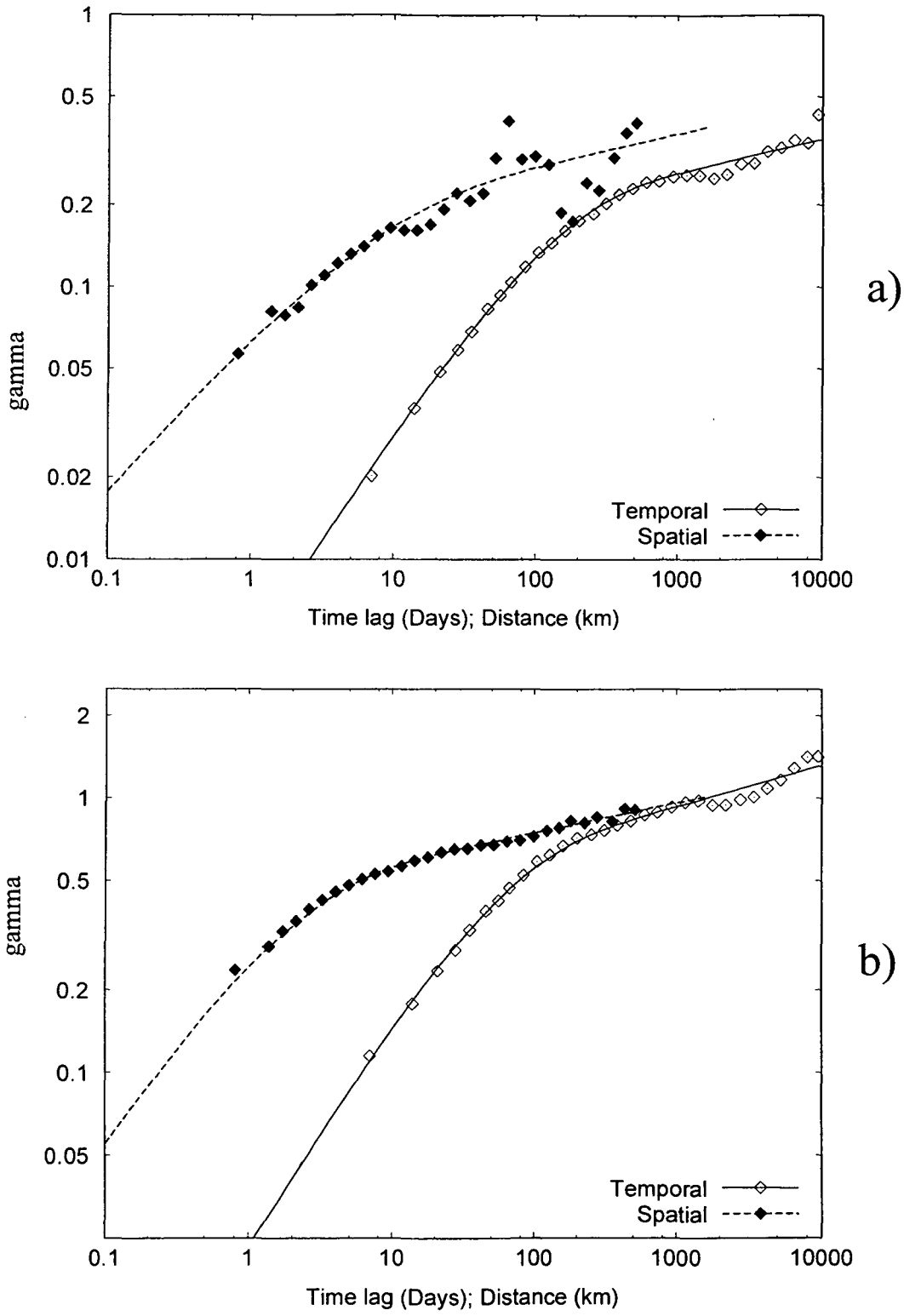


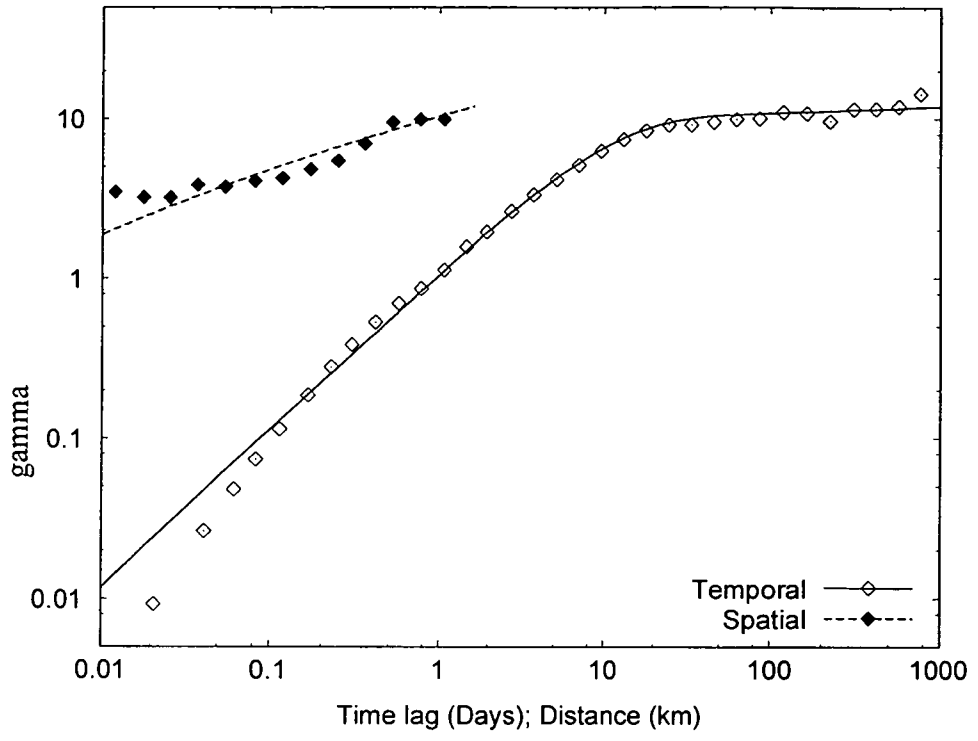
Figure 4.4ab Variograms of groundwater levels. Spatial sample variograms shown as full symbols and temporal sample variograms shown as open symbols. Fitted variograms are shown as lines. (a) shows case I and (b) shows case II (see text).

4.4.4 Soil moisture

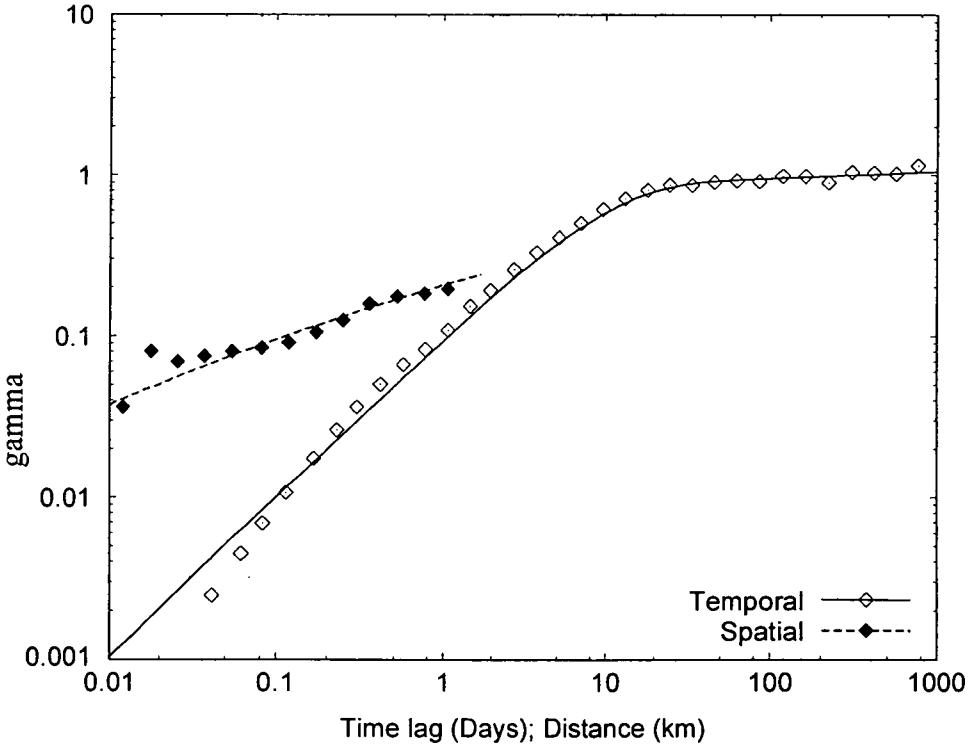
Figure 4.5a (case I) and b (case II) shows spatial and temporal variograms of soil moisture and the resulting parameters are given in Tables 4.2 and 4.3. The two cases are similar apart from the overall levels.

The temporal variograms are approximately stationary. The e-folding distance in time is on the order of one to two weeks. This is similar to the characteristic time scales found for runoff. This similarity is not surprising as the time scale analysis of runoff tends to mainly capture processes between events, such as moisture redistribution and drainage, as mentioned above, which are also those processes captured by the soil moisture data.

The spatial variograms of soil moisture appear non-stationary; however, the maximum extent does not extend far beyond the hillslope scale and there is little increase in variance for the three points with the largest separations. This suggests that there is significant uncertainty about how the variability might change at larger scales. The moisture patterns appear to exhibit a single fractal behaviour without a characteristic scale, although the fit is not very well defined by the data points. It should be noted that random measurement error is about equivalent to $\gamma=2$ in the spatial case, and this may artificially increase the spatial variogram for smaller lags. The fractal power (i.e. the slope in the double logarithmic plot) is on the order of 0.5 ($b+d$ in Tables 4.2 and 4.3). The slope at larger scales (b) is associated with some uncertainty as it has been extrapolated beyond the data points. The spatial variograms cover only two orders of magnitude in space, which is significantly less than for the other variables examined here. The temporal variance is slightly larger than the spatial variance (Table 4.2), as is the case for groundwater levels, although they are not strictly comparable because of the different extents over which the data have been collected (Table 4.1).



a)



b)

Figure 4.5ab Variograms of soil moisture. Spatial sample variograms shown as full symbols and temporal sample variograms shown as open symbols. Fitted variograms are shown as lines. (a) shows case I and (b) shows case II (see text).

4.4.5 Regularisation – spatial aggregation

The spatial aggregation behaviour of precipitation and runoff has been analysed in more detail by making comparisons with regularised variograms. Figure 4.6a shows the results for the case of precipitation. The full circles are the spatial variogram of point precipitation similar to that in Figure 4.2 but based on untransformed data. To this data-derived variogram, a variogram model was fitted, shown as a dashed-dotted line in Figure 4.6a. The regularisation procedure (Eq. 4.10) was then applied to this fitted point variogram to calculate aggregated or block variograms. The size of the aggregation area (a in Eq. 4.10) used was 6, 11.6 and 28.6 km for small, medium and large catchments, respectively, which are the square roots of the median catchment sizes as shown in Table 4.1. The three aggregated variograms so estimated are shown as solid, dashed and dotted lines in Figure 4.6a. For comparison, the variograms that have been directly estimated from catchment precipitation are shown as symbols. The full diamonds, squares and triangles are the spatial variograms of catchment precipitation for small, medium and large catchment classes, respectively, all based on untransformed data. As can be seen from Figure 4.6a, the regularised variograms fit closely to their counterparts directly derived from catchment precipitation. This suggests that precipitation aggregates linearly in space and can be represented by simple aggregation concepts as would be expected.

Figure 4.6b shows the results for the case of runoff, all based on untransformed data. As point runoff data were not available, a hypothetical point runoff variogram was back-calculated by fitting the regularised runoff variograms to the sample variograms of the three catchment size classes as described in section 4.3.3. The three sample variograms are shown as full diamonds, squares and triangles and the three regularised variograms fitted to them are shown as solid, dashed and dotted lines, corresponding to small, medium and large catchment classes, respectively. Figure 4.6b suggests that the regularised variograms provide a reasonable and consistent fit to all three sample variograms. It is now interesting to examine the back-calculated variogram of hypothetical point runoff shown as a dashed-dotted line in Figure 4.6b. The variogram parameters (Eq. 4.7) of this point variogram are $b+d = 0.58$ and $b = 0.017$, and the variance and the e-folding distance are about $\sigma^2 = 0.009$ mm^2 d^{-2} and e-fold = 0.7 km, respectively. This implies that hypothetical point runoff varies over very short space scales, much shorter than catchment runoff, and that its variance is much higher than that of catchment runoff (σ^2 = 0.00190, 0.00085, and 0.00048 mm^2 d^{-2} for the three catchment size classes of untransformed data in Figure 4.6b). The point variogram for runoff is much flatter than that for precipitation and the decrease of variance (as represented by the overall level of the variograms in Figure 4.6b) with catchment size is larger than for the case of precipitation (Figure 4.6a).

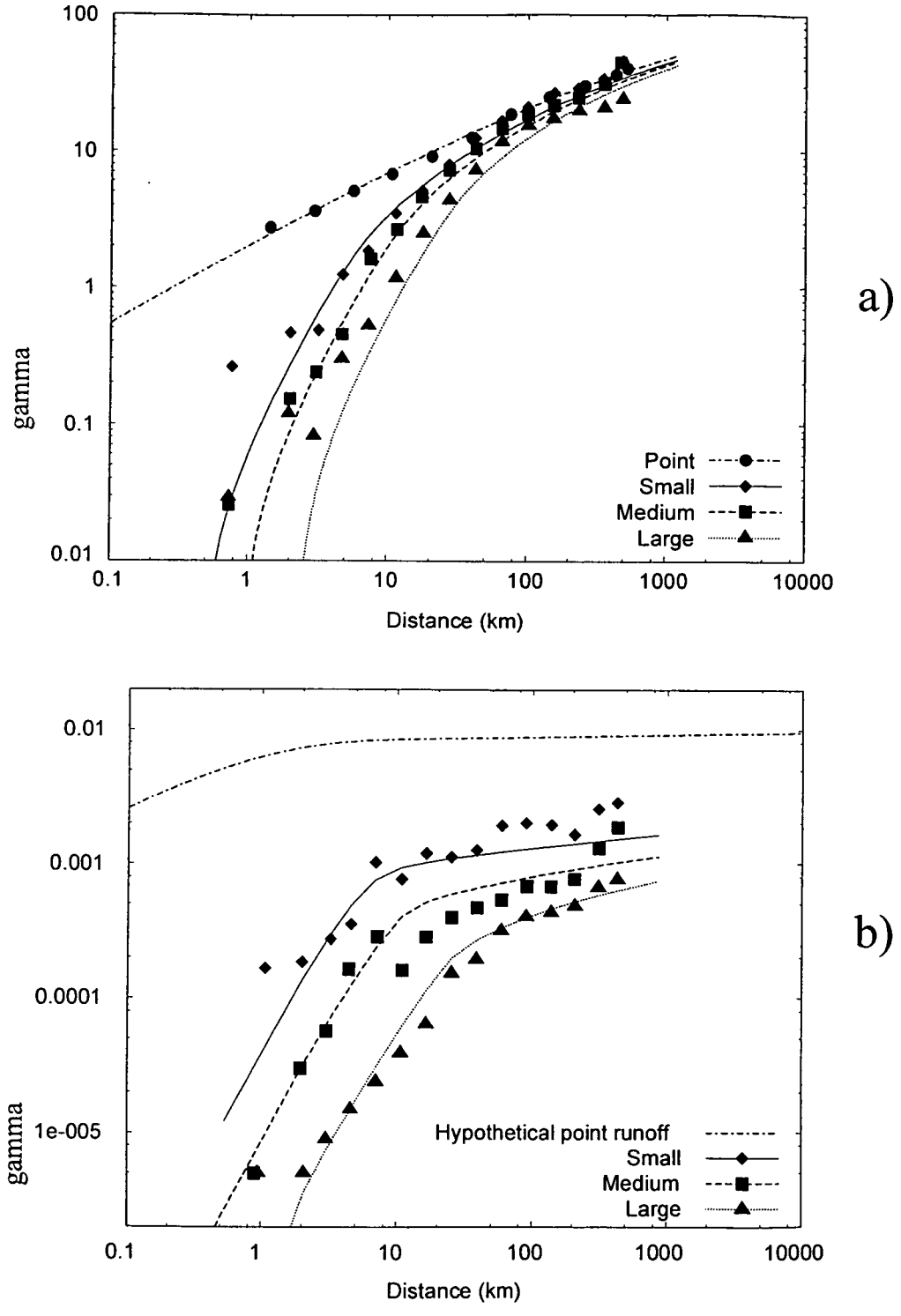


Figure 4.6ab Comparison of regularised spatial variograms with sample variograms for precipitation (untransformed daily precipitation) (a), and runoff (b). In (a) the dashed-dotted line is a fit to the point sample variograms (circles) of precipitation. The other lines have been derived from regularisation and fit the sample variograms of catchment precipitation (symbols) well. In (b) the dashed-dotted line relates to hypothetical point runoff and has been back-calculated by fitting the regularised catchment variograms (other lines) to the catchment sample variograms (symbols).

4.4.6 Links between space and time scales

The characteristic space-time scales have been derived as those scale pairs where the spatial variogram value is equal to the temporal variogram value (Eq. 4.11). The variograms for case I for each of the variables considered have been used (Figures 4.2a, 4.3a, 4.4a, 4.5a), as the relative variances in space and time should be reflected in the scale pairs, so normalisation by the variance is not appropriate. The space-time scale traces are shown in Figure 4.7. Each line shows the trace for one of the data types representing one process. The traces have been slightly extrapolated beyond the range covered by the data points (Figures 4.2a, 4.3a, 4.4a, 4.5a) but only to an extent deemed reliable given the fit of the variogram models to the variograms estimated from the data. The box in Figure 4.7 shows the minimum spacing and the maximum extent of the available precipitation and runoff data, both in space and time, i.e. 1 km to 700 km and 1 day to 30 years. The box thus represents the space-time range over which the precipitation and runoff data are reliable without extrapolation. As all of the precipitation data types lie outside the box, the precipitation space-time scales obtained here need to be interpreted with caution. However, the runoff traces are well within the reliable range. The traces in Figure 4.7 have been superimposed on the figure taken from Blöschl and Sivapalan (1995) representing schematic relationships between spatial and temporal process scales in hydrology. It is now interesting to compare whether the traces obtained from the variogram analysis fit to the schematic. In order to assist with the interpretation two quantities have been derived from the space-time traces. The first is the characteristic velocity $V = L/T$ where T is time scale and L is space scale. The characteristic velocities are a measure of the position of the traces in Figure 4.7 with high velocities plotting on the bottom right hand side of the figure and low velocities plotting on the top left hand side of the figure. The second quantity is the scaling exponent z as in Eq. 4.3. The scaling exponent is equivalent to the slope of the traces in Figure 4.7 as it is a double logarithmic plot. The scaling exponents have been obtained by fitting Eq. 4.3 to the traces over the reliable scale range. The characteristic velocities for different space scales and the scaling exponents are summarised in Table 4.4. Values shown in brackets have been extrapolated and are less reliable than the other values. For comparison, the same quantities have been derived but for transformation case II to examine the effect of the transformation (Table 4.5). Most of the values are very similar although groundwater and soil moisture characteristic velocities are somewhat larger for case II. This is related to the relative magnitudes of the spatial and temporal variances which for

these two variables do change when normalised (Figures 4.4 and 4.5) while they do not for precipitation and runoff (Figures 4.2 and 4.3).

Table 4.4 Characteristic velocities (m/s) and scaling exponents z in Eq. 4.3 for case I. Numbers in brackets have been extrapolated beyond the data.

Process	Size	0.01 km	0.1 km	1 km	10 km	100 km	1000 km	z
Precipitation	Point	-	-	-	-	(3.1)	(10)	0.5
	Small	-	-	-	-	(3.7)	(12)	0.5
	Medium	-	-	-	-	(3.6)	(12)	0.5
	Large	-	-	-	-	(3.9)	(13)	0.5
	Hourly	-	-	-	(5.5)	12	27	0.6
Runoff	Small	-	-	(0.032)	0.040	0.069	(0.18)	0.8
	Medium	-	-	(0.089)	0.075	0.082	(0.20)	1.0
	Large	-	-	-	(0.190)	0.096	(0.19)	1.3
Groundwater level (velocities*10^{-3})	-	-	-	0.39	0.68	0.80	(0.75)	0.8
Soil moisture (velocities*10^{-3})	-	0.063	0.20	0.28	-	-	-	0.5

Table 4.5 Characteristic velocities (m/s) and scaling exponents z in Eq. 4.3 for case II. Numbers in brackets have been extrapolated beyond the data.

Process	Size	0.01 km	0.1 km	1 km	10 km	100 km	1000 km	z
Precipitation	Point	-	-	-	-	(3.2)	(10)	0.5
	Small	-	-	-	-	(3.4)	(11)	0.5
	Medium	-	-	-	-	(3.5)	(12)	0.5
	Large	-	-	-	-	(3.9)	(13)	0.5
	Hourly	-	-	-	(5.3)	11	25	0.7
Runoff	Small	-	-	(0.023)	0.041	0.072	(0.030)	0.8
	Medium		-	(0.055)	0.066	0.081	(0.061)	0.9
	Large	-	-	-	(0.165)	0.100	(0.089)	1.2
Groundwater level (velocities*10^{-3})	-	-	-	0.56	1.1	4.3	(9.4)	0.8
Soil moisture (velocities*10^{-3})	-	0.31	1.2	4.9	-	-	-	0.4

For precipitation the traces of space-time scales in Figure 4.7 fall within those shown as fronts and squall lines in the schematic. There seems to be little difference between the traces for point precipitation and the traces for catchment precipitation for catchments of different size classes. For daily precipitation the characteristic velocities are on the order of 3 to 10 m/s which is slightly lower than what one would expect from the literature. There is likely some bias due to using daily precipitation sums, which can easily account for these differences. The trace for hourly precipitation has larger characteristic velocities than that of daily precipitation (about 5 to 30 m/s). Even though these traces need to be treated with caution as they stem from a combination of temporal hourly data and spatial daily data they seem to be less biased when compared with the schematic. The slopes or scaling exponents of the traces of precipitation are all about 0.5 (0.6-0.7 for hourly precipitation), which is lower than the 0.9 consistent with the schematic. This implies that the characteristic velocities increase significantly with scale. This might also be due to the use of daily precipitation data because

the greatest smoothing occurs at the shortest timescales, which is likely to result in a greater underestimation of velocities for the short time scales and thus a lower scaling exponent.

For runoff the space-time scale traces are slower than the scales shown for channel flow in the schematic and larger than the scales shown for subsurface stormflow and saturation excess overland flow. As suggested in Blöschl and Sivapalan (1995), their schematic shows individual processes. Runoff, however, is the combined result of runoff generation processes such as subsurface stormflow and saturation excess overland flow and runoff routing in the stream channel. The combined process will likely have the largest of the time scales of the component processes because runoff can be thought of as a system in series consisting of runoff generation and routing. Also, the runoff generation processes shown in the schematic are mainly related to individual events while the temporal variograms have been derived from continuous runoff data. This means that the variograms will also capture slower processes between events, such as moisture redistribution and drainage, and hence exhibit somewhat slower scales than shown for the events in the schematic. The characteristic velocities for runoff are on the order of 0.04-0.2 m/s which is slower than what one would expect for channel flow (1 m/s) and similar to what one would expect for subsurface stormflow and saturation overland flow. The traces for catchments of different size classes are significantly different in terms of their slope. The smallest catchments have the smallest scaling exponent (0.8), indicating increasing velocities with scale, while the largest catchments have the largest scaling exponent (1.3), indicating decreasing velocities with scale. The differences between catchment sizes are much larger for runoff than they are for precipitation.

For groundwater levels, the space-time scale traces are close to the scales shown for gravel aquifers in the schematic. As would be expected the characteristic velocities are much slower than those for the other processes and are on the order of $5 \cdot 10^{-4}$ m/s which is about 50 m/day. This is faster than what one would expect as an average flow velocity in all the aquifers from which the data have been taken. It is likely that the characteristic velocities here are related to the celerity with which a perturbation moves through a system. Pressure waves in groundwater systems move much (10 to 100 times) faster than flow velocities, particularly in confined aquifers, which explains the difference. The slope or scaling exponent of the trace of groundwater levels is about 0.8, which is slightly lower than the 0.9 consistent with the schematic. This implies that the characteristic velocities increase slightly with scale.

For soil moisture, the space-time scale trace plots directly on those shown for unsaturated flow in the schematic. The characteristic velocities range between $1 \cdot 10^{-4}$ and $3 \cdot 10^{-4}$ m/s which

are about 0.3 to 1 m/hour. The slope or scaling exponent of the trace is about 0.5, which implies that the characteristic velocities increase significantly with scale.

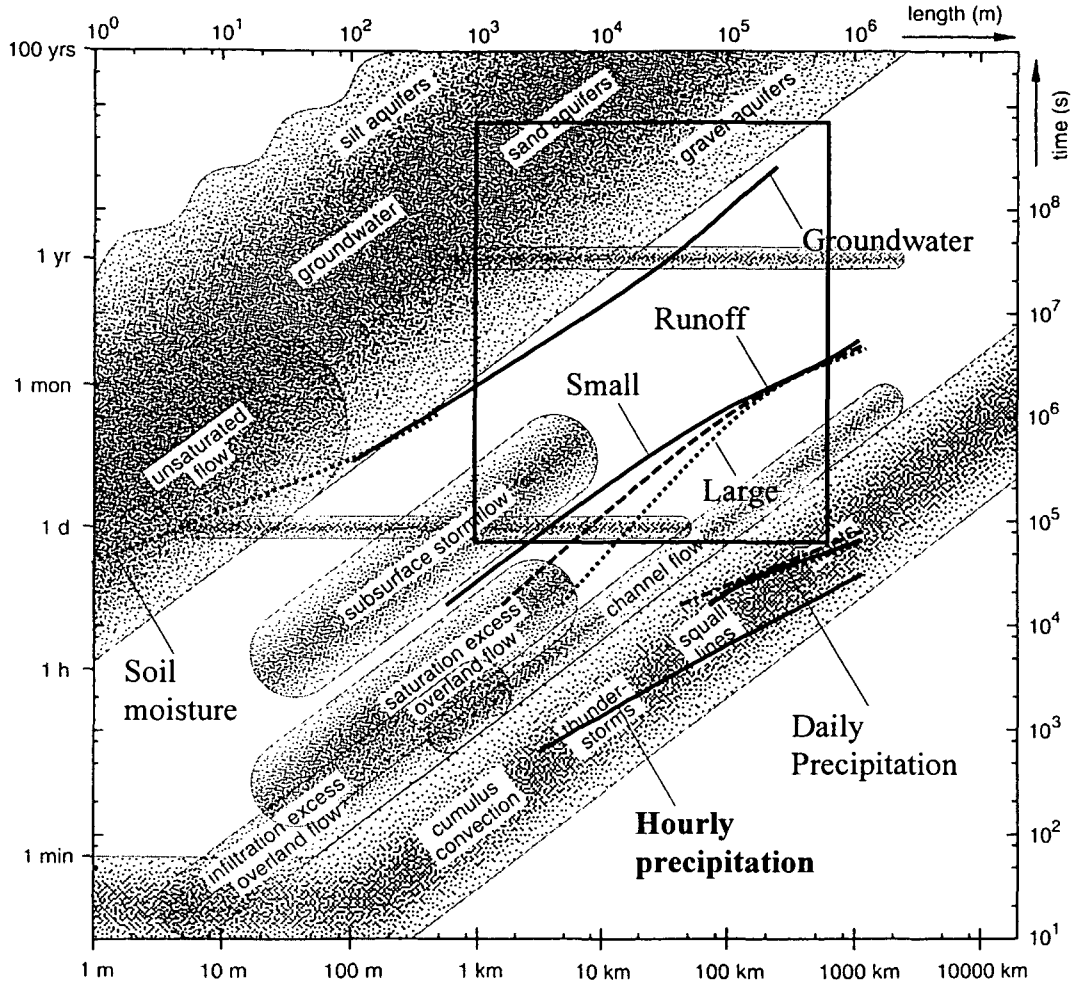


Figure 4.7 Schematic relationship between spatial and temporal process scales in hydrology (from Blöschl and Sivapalan, 1995) shown together with the space-time traces of characteristic scales obtained in this chapter. The rectangle indicates the minimum spacing and the maximum extent of the precipitation and runoff data, both in space and time.

4.5 Discussion and conclusions

4.5.1 Time scales

Characteristic time scales exist for both precipitation and runoff as the temporal variograms are stationary. Their e-folding distances are on the order of 1 day and two to three weeks, respectively. Soil moisture is only approximately stationary with a characteristic time scale of

about two weeks. Moving down into the subsurface the temporal variograms become even more non-stationary and the e-folding distance increases. The groundwater level variograms exhibit an e-folding distance of about half a year. This increase in the characteristic time scales can be thought of as a filtering effect as the water moves on the surface and into the subsurface which removes some of the short term fluctuations and imposes a longer memory on the time variations. This filtering, if linear, can be mathematically represented by a convolution operation of routing, both on the surface and in the subsurface, which will always increase the characteristic scale (e.g. Dooge, 1973). Even if most of the hydrologic processes involved here are known to be non-linear, qualitatively, linear system theory provides a nice framework for the increase in time scales as water flows through a series of compartments.

These filtering concepts can also be used to illustrate the increase in non-stationarity as one moves down into the subsurface. The fractal power of the variogram (b) can be seen as the ratio of large-scale variability and small-scale variability (e.g. Blöschl, 1996). While the small-scale variability of precipitation in time is very high, and practically makes it impossible to see the long time variability (b=0), some of the small scale variability is filtered out for the other processes. For runoff, b is very small but greater than zero, for soil moisture b is about 0.04, and for groundwater levels b=0.1. The slower processes are more affected by the long time variability of precipitation. Indeed, if one fits Eq. 4.7 to a variogram estimated from time series of annual precipitation totals (not shown here), b is about b=0.1 for the Austrian data used here. This similarity with the long term behaviour of groundwater levels is quite remarkable.

It is interesting to relate these results to the huge body of hydrological literature on time series analysis (see e.g. Fleming et al., 2002). For many variables, short records will indicate stationary behaviour although, as the record length increases, there is evidence for non-stationarity (see e.g. Kirchner et al., 2000), particularly if one examines extremes as in the classical paper of Hurst (1951). The stationary behaviour of precipitation is also partly due to the choice of examining continuous time series of precipitation. If the time variability within events is examined, most analyses find non-stationary behaviour, which can be well explained by different fractal concepts (e.g. Seed et al., 1999).

The effect of catchment size on the temporal e-folding distance of both precipitation and runoff is small. For precipitation there is a slight tendency for temporal e-folding distances to increase with catchment size (0.83 days for points, 1.04 days for large catchments). Daily data was used in this chapter and it is likely that the effect would be larger if higher resolution

precipitation data had been used. Although not analysed in the same detail as in this chapter, this is shown in Chapter 5. The precipitation analyses of Seed et al. (1999), for example, showed that the temporal e-folding distances for big grid elements (128*128 km) were significantly longer than for small grid elements (4*4 km) although their work focused on individual storms rather than continuous series. While the overall levels of the variograms (i.e. the variances) of runoff decrease dramatically with catchment size the increase in the temporal e-folding distance is relatively small. Apparently, the tendency of slower response times with increasing catchment area that one would expect, only slightly changes the relatively constant temporal variability of the precipitation forcing, as precipitation becomes runoff. This implies that the slowest process (i.e. that which controls the extent of temporal smoothing of precipitation) is related to generation of runoff rather than routing of runoff. This is supported by the similarity in e-folding timescales for runoff and soil moisture. It is likely that this behaviour would change when one moves to the very big catchments around the world in which routing times of months occur.

4.5.2 Space scales

Characteristic space scales do not exist for point precipitation as the spatial variograms are clearly non-stationary and approximately fractal. This is consistent with a substantial body of literature on the fractal behaviour of rainfall processes (e.g. Lovejoy and Schertzer, 1985) although some authors state that spatial rainfall variability may exhibit a more complex behaviour than that suggested by the power law in Eq. 4.1, e.g. multifractality (Gupta and Waymire, 1990). As one moves from point precipitation to catchment precipitation, a break appears in the variograms which is more pronounced for large catchments than it is for small catchments. Clearly, the averaging over a catchment area imposes a characteristic spatial scale on precipitation. These aggregation effects are consistent with spatial regularisation methods, i.e. simple linear aggregation or filtering in space. The main effect of the aggregation by catchments is to filter out the small scale variability (i.e. the variograms get steeper at short lags), to reduce the overall variance (i.e. the overall level of the variograms decreases), and to increase the e-folding distance. The spatial variograms of runoff are similar to those of catchment precipitation in that they exhibit a break, and consist of a steep part at short lags and a flatter part at large lags. Again, catchment size appears to impose a characteristic spatial scale. When moving from small catchments to large catchments the characteristic scales of runoff increase from about 20 km to about 60 km. The characteristic space scale of back-

calculated point runoff is much smaller (0.7 km) which is consistent with this catchment scale effect. These differences in the e-folding distances are larger than those for catchment precipitation where the e-folding distance increased by about 20% when moving from small to large catchments while the e-folding distance tripled in the case of runoff.

Catchment runoff exhibits stronger aggregation behaviour than catchment precipitation. Woods et al. (1995) performed a similar analysis of spatial variance of runoff although their catchment sizes were smaller (<10 km^2) than in the analyses presented here. They found that the variance between catchments decreased faster than with the inverse of catchment area. A decrease of variance with the inverse of catchment area would result if point runoff were an uncorrelated random field. They interpreted the faster decrease as an evidence of spatial organisation. An alternative interpretation is that their fast decrease is a result of increasing catchment response times with catchment scales, i.e. space-time aggregation effects. Also in the analyses presented here the spatial variance of runoff does decrease strongly with catchment size, but it is less than with the inverse of catchment area (e.g. $\sigma^2 = 0.3$ for a medium catchment size of 36 km^2 and $\sigma^2 = 0.15$ for a medium catchment size of 701 km^2, Table 4.2). Hypothetical point runoff may therefore be spatially correlated which is consistent with the results from the regularisation analysis.

It is also of interest to compare the spatial variograms of runoff obtained here with the theoretical variograms of Gottschalk (1993). The main difference between the variograms of Gottschalk and those derived here is that Gottschalk (1993) took into account the spatial network structure while this has not been done here. He suggested that the organisation of the landscape into catchments implies that the spatial runoff field is non-homogeneous and can be better described by a tree-structure than by a homogeneous random field. With this type of spatial structure, the variograms are no longer second order stationary but will change with location. Gottschalk (1993) assumed an exponential spatial variogram for hypothetical point runoff applicable to the average runoff from a number of events. He then plotted the correlation against distance along the river rather than lag as has been done here. When going from points to catchments the shape of the variograms changed in a similar fashion as shown in this chapter. The overall level of the variograms decreased (i.e. better correlations) and the spatial e-folding distances increased (i.e. correlations over longer distances) as catchment size increased. It would be interesting to apply the method of Gottschalk (1993) to the data used here to see how much the effects of non-homogeneity will change the spatial variogram shapes.

As one moves from catchment precipitation to runoff, the spatial e-folding distance decreases for the small catchments and it is still shorter for groundwater levels (59 km, 18 km and 8 km for catchment precipitation, runoff and groundwater levels, respectively, Table 4.2). This means that the characteristic spatial scales of the processes decrease as the water moves through the hydrologic system. This is likely to be the result of a superposition of the small scale variability of catchment and aquifer properties on the rainfall forcing. Both soils and aquifer characteristics tend to by highly variable at small scales (see Roth et al., 1990, Gelhar, 1993, p. 292). These smaller scales are then imposed on the spatial variability of the water fluxes.

It is interesting to contrast this reduction of spatial scales with the increase in temporal scales by catchments and aquifers discussed above. The main reason for this difference is that, in the time domain, the routing processes can be thought of as a convolution that will always increase the characteristic scales as the water flows on the surface and through the subsurface. This aggregation in time also effectively results in a smoothing of the precipitation forcing of the system in space as a consequence of the space-time correlations in the precipitation field. However, this space-time smoothing of forcing is dominated by a second spatial effect. The most important effect in the space domain is not a convolution that takes place for a given catchment area but rather it is the transformation of rainfall to rainfall excess and recharge which can be represented, for example, by a multiplicative or additive operation (see e.g. Woods and Sivapalan, 1999). In addition to this spatially variable point scale transformation, processes such as lateral flow can add spatial variability at small spatial scales, as is the case with soil moisture on hillslopes. If a large scale random field (precipitation) is combined with a small scale random field (soils) the resulting field will exhibit intermediate scales and this is what can be observed for runoff and groundwater levels. As one moves from small to large catchment runoff the spatial scales increase because of aggregation effects.

The e-folding distance found for soil moisture is much smaller than that for the other processes. This fits into the above reasoning of soil characteristics imposing short space scales. However, it should be noted that the overall extent of the data used here was much smaller than that of the other variables, which may bias the estimated characteristic scales towards smaller scales (e.g. Blöschl, 1999). It is also worth comparing the fractal shape of the spatial variograms found here to reviews of Western et al. (1998, 2002). Most of the variograms compiled by Western et al. suggest stationary spatial soil moisture variability with characteristic scales on the order of tens to hundreds of meters. It is possible that part of the

difference between the results of the analyses in this chapter and those of Western et al. (1998, 2002) is related to removing the spatial organisation of the soil moisture patterns here, as the focus was on the random part of the variability rather than on the deterministic part. The spatial organisation imposed by the landscape may exhibit a characteristic spatial scale (Western et al., 1999), which however will not be apparent if only the random component is examined. Also it should be recalled that the shape of the variograms at the largest lags suggests some uncertainty in the interpretation of stationarity. However, where intermediate extents of up to tens of km have been examined, fractal behaviour has been observed (Rodríguez-Iturbe et al., 1995). This may suggest that the underlying soil moisture variability is indeed non-stationary, which may however not appear in the variogram analysis because of the limited size of an individual case study. It should be noted that soil moisture is a bounded variable, so one would not expect the variogram to increase without bounds. Indeed Entin et al. (2000) found stationary behaviour at scales above hundreds of kilometres.

4.5.3 Links between space and time scales

The space-time traces of characteristic scales as derived from equating the variogram values of the spatial variograms and those of the temporal variograms fit well into the schematic diagram of space-time scales proposed by Blöschl and Sivapalan (1995). The characteristic velocities for precipitation and runoff are on the order of 10 and 0.1 m/s respectively. The characteristic velocities of both groundwater levels and soil moisture range between 10^{-4} and 10^{-3} m/s. There are order of magnitude differences between the different processes, and it is these one means to capture when characterising processes by characteristic scales as discussed in the introduction. There are some minor differences between the schematic and the traces found here which can all be explained on the basis of process reasoning and data characteristics. The traces for daily precipitation are slightly slower than those in the schematic, which may be related to biases imposed by using daily precipitation data. The traces for hourly precipitation are slightly faster than those in the schematic, which may be due to biases introduced by combining hourly temporal variograms and daily spatial variograms. For example, Seed et al. (1999) showed that hourly precipitation is less well correlated in space (shorter e-folding distances) than daily precipitation, so the assumption made here may overestimate the characteristic velocities slightly.

The characteristic velocities of runoff found here are somewhat slower than those suggested by the schematic of Blöschl and Sivapalan (1995). As suggested earlier in this chapter this is likely to be due to both the processes and the scales considered. The schematic focuses on single events while this chapter examined continuous series. The schematic also represents individual processes while this chapter examined runoff as the combination of a number of processes. The analysis in this chapter utilized daily flow data and most catchments were less than 1000km^2, thus the routing times for surface runoff would generally be less than a day. However the temporal scales of runoff are much longer than a day. This means that the temporal pattern of flow is dominated by the temporal behaviour of precipitation forcing and runoff generation processes at daily and longer time scales, not by the routing of surface runoff, in catchments of this size. Comparisons of the small, medium and large catchments show that neither the scales of the temporal forcing or the filtering of that forcing depend on catchment size. However there is a significant increase in the spatial scales as one moves from small to large catchments. This combined with the spatially independent temporal scales leads to the apparent increase in velocity with spatial scale.

In general there is a fundamental difference in how the landscape modifies the meteorological forcing in space and time. It removes temporal variability and adds spatial variability. The landscape can be conceptualised as a collection of parallel (vertical stores) that interact laterally to some extent via different flow paths, including overland and shallow subsurface flow paths determined by, and thus limited to the scale of, the topography as well as deeper groundwater flows through aquifers, typically at somewhat larger scales. There are also spatial differences in these stores due to variations in soil and vegetation properties. The stores act in time to filter out variance in the forcing, especially at smaller temporal scales.

These storage effects increase the characteristic time scale from about one day for rain to ten days for soil moisture and a hundred days for groundwater, which also becomes nonstationary as a consequence. The runoff response timescales are a mix of the soil moisture and groundwater response timescales because runoff is influenced by both soil moisture and groundwater. The soil moisture affects the partitioning of rainfall into infiltration and runoff, as well as the subsurface stormflow and interflow processes. The groundwater affects the baseflow. However spatial differences between the stores mean that this temporal filtering is spatially variable. Milly and Wetherald (2002) showed how storage in the saturated zone and in surface waters acts as a low pass filter that damps small scale variability in the time domain, but leaves large scale variability unmodified from that of runoff production. The

strength of the filter, in their paper quantified by an effective water residence time, varied widely from one catchment to another.

In addition there are lateral interactions between these stores. Some of these, such as subsurface lateral flow, also add to the spatial variability of the soil moisture field and hence the runoff field, by moving water from dry upgradient (hilltop) to wetter downgradient (valley) locations. The end result of this landscape variability is to add smaller scale spatial variability to the large scale spatial forcing, thus resulting in a mixing of the spatial scales of the landscape and of the forcing and a reduction in the characteristic spatial scales from about 50 km for precipitation to about 10 km for groundwater. There is little reduction in the spatial scale of runoff compared with rainfall because the data used here is an integral over the catchment area. Back-calculated characteristic space scales of hypothetical point runoff are indeed much smaller (about 1 km). Note that the characteristic spatial scale of soil moisture is difficult to compare to the other fields due to the different spatial extent of the data sets. These decreases in spatial scale are more than an order of magnitude less than the increases in temporal scale, thus the changes in process velocity that are discussed below result more from temporal effects that are related to residence times and flow velocities through the various stores than from spatial effects.

Most of the space-time traces of characteristic scales plot as straight lines or approximately straight lines in the double logarithmic diagram indicating that space-time scales are related by a power law of the form of Eq. 4.3. The scaling exponents z found here are somewhat flatter than those shown in the schematic of Blöschl and Sivapalan (1995) and there are significant differences between the processes examined here. It is reasonable to interpret the scaling exponents found here in the context of environmental transport processes. Taylor's (1938) hypothesis states that, instead of measuring spatial variability one can measure temporal variability, and then calculate spatial variability by a simple conversion factor, the characteristic velocity. Taylor's hypothesis holds if the diffusion of the process is significantly smaller than advection, i.e. if a perturbation moves sufficiently fast through the sensor that it does not change too much during the time it passes the sensor. One way to express this relationship is the Peclet number which is the ratio of the advective term and the diffusive term in, say the groundwater advection dispersion equation. If the Peclet number is significantly larger than unity, then advection dominates and one would expect Taylor's hypothesis to hold. In the context of the present analysis one can replace spatial variability by temporal variability if and only if the spatial and temporal variograms of a particular process

have the same shape. If they do have the same shape then z in Eq. 4.3 must be $z = 1$ as time scale and space scale must be proportional with a simple conversion factor, the characteristic velocity. One can argue that if $z = 1$ advection can be expected to dominate over diffusion. On the other hand if $z < 1$ then one would assume that diffusion plays a significant role. Because of $z < 1$ the characteristic velocity will increase with both space and time scales. This may be due to diffusive processes that come into play and contribute to changing the shape of a perturbation in addition to advection. The combined effect of advection and diffusion will give rise to larger characteristic velocities than advection alone.

In the light of this reasoning an interpretation of the z values found here is given. Precipitation gives a relatively small value of $z = 0.5$ implying, if one accepts the interpretation suggested above, that diffusion is quite important. For the space-time scales examined here this is likely, as the weather systems do change significantly within a day as they move over Austria, as evidenced by weather radar images. The value of z is also consistent with the range of $z = 0.5$-0.6, found from radar data analysis by Foufoula-Georgiou and Vuruputur (2000) even though they worked at a time scale of minutes rather than days as is the case here. For groundwater, z is significantly larger than the value for precipitation ($z = 0.8$), which may suggest that advection becomes relatively more important. All the groundwater level data used here stem from porous aquifers with significant lateral movement. Some of the aquifers are in mountain valleys and the water flow follows the topographic gradient. Other aquifers such as the Südliches Wiener Becken aquifer south of Vienna, again, has relatively high flow velocities which result from significant recharge in the upstream part adjacent to the mountains but very little recharge in the lower flatter part of it. Soil moisture on the other hand has a much smaller value of $z = 0.4$-0.5, suggesting that advection is relatively unimportant as compared to diffusion. This is consistent with field observations in some of these catchments, which indeed suggest that the lateral redistribution of soil moisture is limited to a relatively short time of the year (Grayson et al., 1997). For the case of catchment runoff, z differs with catchment size class. For the smallest size class $z = 0.8$, implying that advection is relatively more important than diffusion as compared to precipitation. As catchment size increases, so does z, implying that the importance of advection over that of diffusion increases even more. This can be related to the longer stream distances and stream routing times in larger catchments, which would suggest that they more strongly impose the characteristics of advection on the space-time variability of runoff than is the case in smaller catchments. One can also speculate that if one extrapolates the runoff traces in Figure 4.7 from large

catchments beyond small catchments to give a trace for hypothetical point runoff, this trace would be flatter, perhaps with a slope of z = 0.5-0.6. This slope is consistent with diffusion being quite important relative to advection, which is what one would expect for hypothetical point runoff. However, this reasoning is somewhat speculative as it does not explain the z values larger than unity for large catchment runoff.

Although the interpretations offered above are of a general nature it is believed that they may contribute to a better understanding of macroscale space-time hydrologic variability in the spirit of Dooge (1986).

4.5.4 Outlook

The main strength of the analyses presented in this chapter is that space and time variograms have been examined over many orders of magnitude, which is needed for a robust analysis of the presence of characteristic scales. This has been made possible by using a relatively comprehensive data set for each variogram. However, the robustness has been at the cost of lumping together a large number of stations for the temporal variograms, and a large number of time steps for the spatial variograms. This implies that many different sub-processes have been lumped together. Ideally one would like to derive characteristic scales for individual processes separately, in order to be able to link process and scale more closely. One example of this classification in the context of scale is given by Skaugen (1997) who classified precipitation events into two groups according to their spatial characteristics. He then calculated variograms for each of these groups separately. The groups had quite contrasting properties. One of the groups exhibited stationary variograms with characteristic scales on the order of 10 km while the other group exhibited non-stationary variograms with characteristic scales of 100 km or more. A process classification for runoff and catchment state has been suggested by Blöschl et al. (2000) which should be applicable to the problem studied here and will be pursued in future work.

Another obvious extension of the work in this chapter is to examine space-time aggregation effects. This is done in Chapter 5. Although it was possible to fit spatially aggregated runoff variograms to catchments of different size classes in a consistent way, a more complete approach would involve both spatial and temporal aggregation effects by making use of catchment size and catchment response times, respectively. For space-time regularisation

more elaborate space-time variogram models are needed such as those suggested by De Cesare et al. (2001) and Christakos (2000).

Some initial attempts at making use of characteristic scales of hydrologic processes as those derived in this chapter do already exist. One example is network planning and the examination of scale effects in the context of soil moisture measurements (Western and Blöschl, 1999). However, overall, hydrology is still at a very early stage as compared to other disciplines. It is hoped that the characteristic scales of hydrologic processes as derived in this chapter will contribute to fostering a more wide spread use of characteristic scales in hydrology.

Chapter 5. Catchments as space-time filters – a joint spatio-temporal geostatistical analysis of runoff and precipitation

5.1 Introduction

Geostatistical methods fall into two groups. The first focuses on the characterisation of spatial variability and is termed structural analysis. It provides a representation of the spatial structure of the variables of interest in terms of the variogram and sheds light on the continuity of the processes involved. In hydrology, structural analysis plays an important role in aquifer assessment and sampling design (e.g. James and Freeze, 1993). The second group of geostatistical methods consists of spatial estimation methods where the variogram obtained in the structural analysis step is used to estimate the variable of interest at locations where no measurements are available. Spatial estimation methods based on geostatistical concepts are widely used in many geosciences including subsurface hydrology (Renard et al., 2005).

In catchment hydrology, geostatistical concepts have been used more sparingly. This is because of the nested structure of catchments which makes geostatistical analyses more complicated as compared to the usual analysis of point samples or blocks. However, a number of recent studies have demonstrated that geostatistical methods can indeed account for the nested catchment structure. This applies to both the structural analysis step of understanding the spatial structure and the spatial estimation step of estimating variables such as streamflow at locations where no data are available. The latter addresses the ungauged catchment problem (Sivapalan et al., 2003). Based on the work of Gottschalk (1993a, 1993b) Sauquet et al. (2000) presented a spatial estimation method for annual streamflow. A similar spatial estimation method, termed Top-kriging, will be presented in Chapter 6 and 7. Reliable variograms are needed for applying this type of spatial estimation methods.

Runoff is a process that varies in both space and time. It is therefore appealing to extend the spatial estimation of Sauquet et al. (2000) and the spatial and temporal analyses in Chapter 4 to the spatio-temporal case, i.e. to analyse and estimate runoff as a function of both space and time. Spatio-temporal variograms are needed for this. At the same time, spatio-temporal

variograms of runoff may shed light on the nature of hydrological variability in space and time. Chapter 4 presents analyses of the effect of different catchment sizes on the spatial and temporal variograms of precipitation and runoff. The results indicated that variograms of observed runoff were consistent with variograms obtained by aggregating variograms of hypothetical point runoff. However, their study examined spatial and temporal variograms independently. It is likely that the spatial and temporal variabilities of runoff are related given that it takes longer for water to move through large catchments than through small catchments. Woods et al. (1995) analysed catchments in the range of 0.04-50 km^2 and found the variance of streamflow to decrease more strongly with catchment area than what would be expected for the spatial aggregation of a random process. Woods et al. (1995) noted that this may be due to the presence of organisation at large scales that is not present at small scales. In Chapter 4, it is suggested that this may be related to spatio-temporal aggregation effects instead.

In this chapter, runoff will hence be analysed in space and time jointly, including an examination of the joint spatial and temporal aggregation effects of runoff. As a central concept, the description of Woods and Sivapalan (1999) is adopted, where runoff from a catchment is represented as the convolution of the local runoff generated within the catchment within a time period. This is consistent with the filter concept from Chapter 4 where the catchment area is interpreted as the geostatistical support of the runoff measurements. In a joint spatio-temporal analysis the catchments then operate as space-time filters and runoff measurements are associated with both a spatial support (the catchment area) and a temporal support (the response time of the catchment).

The objective of this chapter is to analyse spatio-temporal variograms of runoff and examine the potential of estimating these from point variograms of runoff by spatio-temporal aggregation. For comparison and for testing the aggregation procedure, spatio-temporal variograms of precipitation are analysed as well. This study uses a similar data set as the one used in Chapter 4, but goes beyond the earlier study in two ways. First, the variograms were in Chapter 4 analysed in space and time separately while, here, a joint analysis is performed to shed light on the connection of space and time scale variability. Second, the analyses in Chapter 4 were on a data set of daily values while in this chapter, a data set of quarter hourly values is used. It is then possible to perform a more detailed analysis of the short term characteristics of runoff that are important for space-time connections.

5.2 Data

The data used in this chapter stem from a comprehensive hydrographic data set of Austria. Austria has a varied climate with mean annual precipitation ranging from 500 mm in the eastern lowland regions up to about 3000 mm in the western alpine regions. Runoff depths range from less than 50 mm per year in the eastern part of the country to about 2000 mm per year in the Alps. Potential evapotranspiration is on the order of 600-900 mm per year. Precipitation data from 991 stations for the period 1981-1997 were used in this study (Fig. 5.1a). 161 of the stations were recording rain gauges while the rest were daily raingauges. The daily records were disaggregated to a time step of 15 minutes based on the temporal patterns of the neighbouring stations (Merz et al., 2006). In order to be able to examine spatial aggregation effects, catchment precipitation was calculated for each time step by external drift kriging interpolation of the point data for a total of 579 catchments using topographic elevation as an auxiliary variable. The catchment precipitation series so obtained were divided into three size classes (Table 5.1). Runoff data from 591 catchments for the period 1971-2000 were used that all had a time resolution of 15 minutes. The catchments were subdivided into three classes according to catchment size – small (3 - 71 km²), medium (72 - 250 km²) and large (250 - 131000 km²) (Fig. 5.1b). Catchments smaller than 10 km^2, as well as catchments with short records, significant anthropogenic effects or lake effects were excluded from the data set. This resulted in a total of 488 stream gauges available for the analysis. Table 5.1 summarises the data series used in this chapter. The runoff data set consists of a total of $5 \cdot 10^8$ individual data values.

Table 5.1 Data series used in this chapter.

Data type	Size class	Size range (km^2)	Median size (km^2)	Number of stations	Extent of domain (km)	Time resolution (min)	Period
Point precipitation	-	Point	Point	991	700	15	1981-1997
Catchment precipitation	Small	3-71	35	193	700	15	1981-1997
	Medium	72-236	125	193	700	15	1981-1997
	Large	241-131000	670	193	700	15	1981-1997
Runoff	Small	10-71	42	142	700	15	1971-2000
	Medium	72-248	119	178	700	15	1971-2000
	Large	251-131000	605	168	700	15	1971-2000

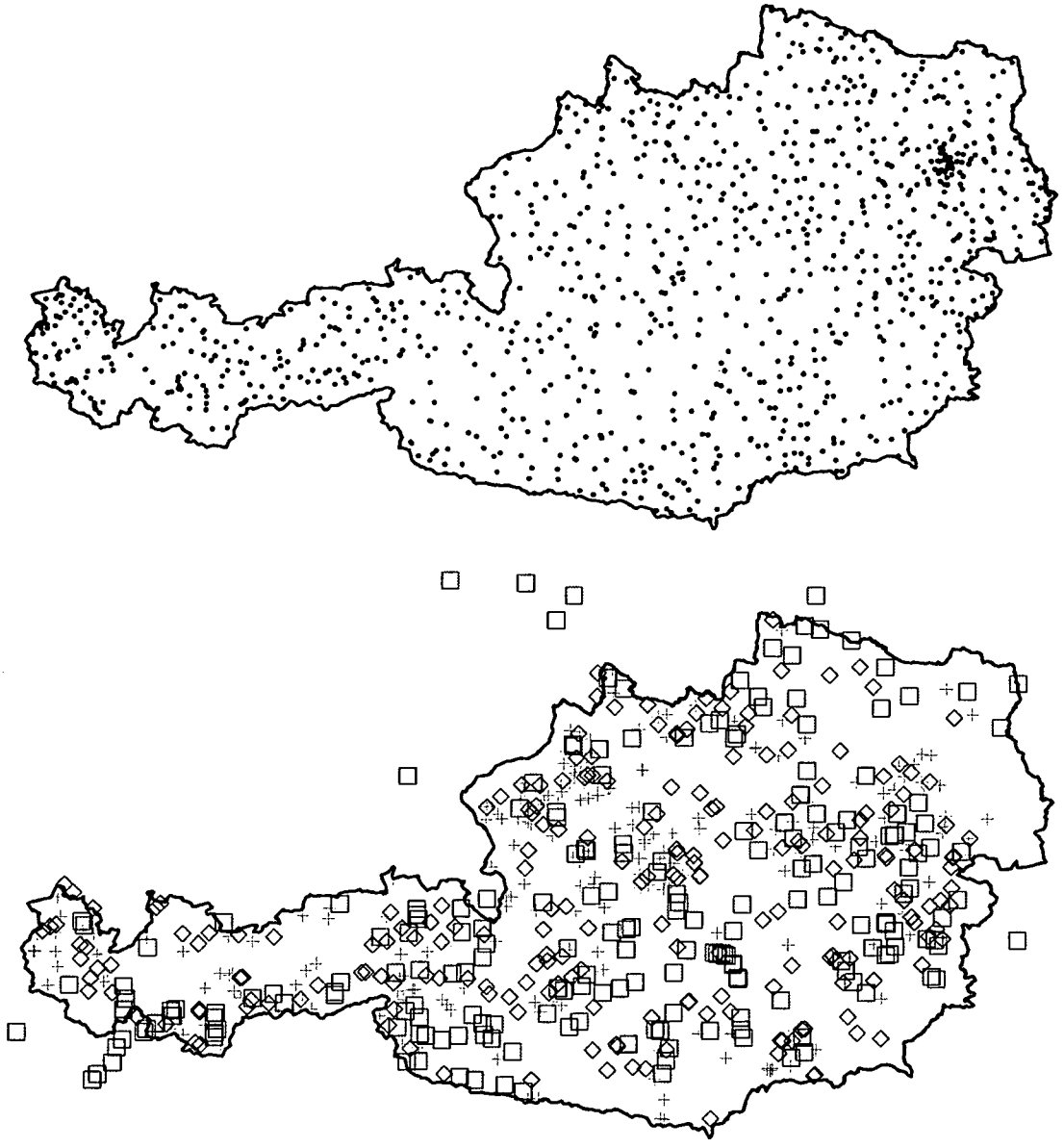


Figure 5.1 Network of measurement stations in Austria used in this chapter. Precipitation gauges (top); centroids of gauged catchments (bottom) (small catchments shown as plusses, medium sized catchments as diamonds, large catchments as squares).

5.3 Method

5.3.1 Spatio-temporal sample variograms

Spatio-temporal sample variograms were calculated from the runoff data separately for the three catchment size classes, and from catchment precipitation separately for the three catchment size classes as well as for point precipitation:

$$\hat{\gamma}_{st}(h_s,h_t)=\frac{1}{2\sum_{j=1}^{m(h_s)} n_j(h_t)}\sum_{j=1}^{m(h_s)}\sum_{i=1}^{n_j(h_t)}(z(\mathbf{x}_j+\mathbf{h}_s,t_i+h_t)-z(\mathbf{x}_j,t_i))^2 \qquad (5.1)$$

where $h_s=|\mathbf{h}_s|$ and h_t are the spatial and temporal lags, respectively, $z(\mathbf{x}_j,t_i)$ is precipitation or runoff at time t_i and spatial location $\mathbf{x}_j$ of station j, $m(h_s)$ is the number of pairs of stations with distance h_s, and $n_j(h_t)$ is the number of pairs of points in time with time lag h_t within a spatial or temporal bin. h_s was taken as the distance between the centres of gravity of the catchments for the cases of runoff and catchment precipitation and as the station distance for the case of point precipitation. The spacings of the bins were selected approximately logarithmically (with the exception of zero lags). The variograms of precipitation were calculated on the basis of precipitation intensity, those of runoff on the basis of specific discharge. The physical units of the precipitation and runoff variograms hence are ($mm^2 \cdot h^{-2}$) and ($m^6 \cdot km^{-4} \cdot s^{-2}$) with 1 $m^6 \cdot km^{-4} \cdot s^{-2}$ = 12.96 $mm^2 \cdot h^{-2}$. The space and time units used are kilometres and hours, respectively.

5.3.2 Spatio-temporal variogram models

Numerous spatio-temporal variogram models have been proposed in the literature. There are two types, separable and non-separable models. In separable models, the covariance can be factorised into two components, one component containing time lag only and the other containing space lag only. Rodríguez-Iturbe and Mejía (1974) presented an example of a separable model. Cressie and Huang (1999) proposed a series of non-separable models. De Cesare et al. (2001) and De Iaco et al. (2001) extended some of the earlier models into a product-sum model. Kyriakidis and Journel (1999) reviewed spatio-temporal variogram models and discussed advantages and disadvantages of different model types. Fuentes (2006) and Mitchell et al. (2005) proposed methods for testing if a process can be modelled by a

separable model. They noted that for some spatio-temporal modelling applications, the computational burden can be reduced considerably by using separable models. Cressie and Huang (1999), however, suggested that non-separable models are necessary for many natural cases.

Four models are compared in this chapter that are all non-separable: a spatio-temporal exponential model, a model proposed by Cressie and Huang (1999), the product-sum model (De Cesare et al., 2001, De Iaco et al., 2001), to all of which a fractal component is added (Eq. 5.8), as well as a pure fractal model. The exponential model is:

$$\gamma'_{1st}(h_s,h_t) = a_1(1-\exp(-((c_1h_t+h_s)/d_1)^{e_1})) \tag{5.2}$$

a_1 is the sill or the variance for infinite lag, c_1 is a scaling parameter for time, d_1 is a spatio-temporal correlation length and e_1 defines the slope of the short distance part of the variogram. The model is consistent with the Taylor hypothesis which assumes that a constant characteristic velocity exists, so space and time are interchangeable (Taylor, 1938 and Chapter 4 this dissertation).

Cressie and Huang (1999) derived a number of models from Bochner's theorem (Bochner, 1955). A number of them were tested and the focus in this chapter is on:

$$\gamma'_{2st}(h_s,h_t) = a_2\left(1-\frac{1}{(c_2h_t+1)^{(d+1)/2}}\exp\left\{-\frac{b_2^2h_s^2}{c_2h_t+1}\right\}\right) \tag{5.3}$$

a_2 is the sill, b_2 and c_2 are scaling parameters for space and time, respectively, and d is the spatial dimension.

The third model is the product-sum model which is derived from a covariance model that combines products and sums (De Cesare et al., 2001, De Iaco et al., 2001):

$$\gamma'_{3st}(h_s,h_t) = \gamma'_{3s}(h_s)+\gamma'_{3t}(h_t)-k\gamma'_{3s}(h_s)\gamma'_{3t}(h_t) \tag{5.4}$$

where k is a parameter. $\gamma'_{3s}(h_s)$ and $\gamma'_{3t}(h_t)$ represent the spatial and temporal variograms, respectively:

$$\gamma'_{3s}(h_s) = a_{3s}(1-\exp(-(h_s/d_{3s})^{e_{3s}})) \tag{5.5}$$

$$\gamma'_{3t}(h_t) = a_{3t}(1-\exp(-(h_t/d_{3t})^{e_{3t}})) \tag{5.6}$$

with parameters similar to Eq. 5.2. The product-sum model reduces to the separable model proposed by Rodríguez-Iturbe and Mejía (1974) for $a_{3s}=a_{3t}=1/k$. The three variogram models (Eqs. 5.2, 5.3, 5.4) are stationary, i.e., they are finite for infinite lags. It was in Chapter 4 showed that daily precipitation can be regarded as stationary in time, daily mean runoff is almost stationary in time, while neither of the processes can be regarded as stationary in space within the spatial extent of the data set used. The variograms were therefore modified to account for non-stationarity in both spatial and temporal directions. Although it was in Chapter 4 noted that runoff was almost stationary in time, a small non-stationary part was found to be necessary for the regularisation procedure in this chapter. For application in spatial (and spatio-temporal) estimation a variogram needs to be such that the variance of any linear combination Y of the variable z of the type $Y=\sum_{i=1}^{n} a_i z(\mathbf{x}_i,t_i)$ is equal to zero or positive. This requirement is fulfilled by Eqs. (5.2, 5.3, 5.4). If the variogram is non-stationary, the following condition has to be fulfilled:

$$Var(Y) = -\sum_i \sum_j a_i a_j \gamma_{st}(|\mathbf{x}_i - \mathbf{x}_j|, t_i - t_j) \geq 0 \qquad (5.7)$$

with $\sum_{i=1}^{n} a_i = 0$. $-\gamma_{st}(h_s,h_t)$ is then by definition said to be a "conditional positive definite function" (Journel and Huijbregts, 1978, Cressie and Huang, 1999). To ensure conditional positive definiteness of $-\gamma_{st}(h_s,h_t)$, it is common to specify the variogram as a sum or a product of models that are known to have this property.

Spatial and temporal fractal components that are positive definite have therefore been added to the three variogram models γ'_{st} of Eqs. 5.2, 5.3 and 5.4:

$$\gamma_{st} = \gamma'_{st} + a_s h_s^{\alpha} + a_t h_t^{\beta} \qquad (5.8)$$

where a_s and a_t are parameters that adjust the level of the fractal part, and α and β are the spatial and temporal fractalities, $0<\alpha<2$ and $0<\beta<2$. Although this model ensures conditional positive definiteness, the non-stationary part (Eq. 5.8) does not include space-time interactions.

In addition to the three variogram models, a pure fractal model was examined for comparison (Eq. 5.12 below). In summary, the following variogram models were used in this chapter:

Exponential model:

$$\gamma_{1st}(h_s,h_t) = a_1(1-\exp(-((c_1h_t+h_s)/d_1)^{e_1})) + a_sh_s^{\alpha} + a_th_t^{\beta} \quad (5.9)$$

Cressie-Huang model:

$$\gamma_{2st}(h_s,h_t) = a_2\left(1-\frac{1}{(c_2h_t+1)^{(d+1)/2}}\exp\left\{-\frac{b_2^2h_s^2}{c_2h_t+1}\right\}\right) + a_sh_s^{\alpha} + a_th_t^{\beta} \quad (5.10)$$

Product-sum model (using Eqs. 5.5 and 5.6):

$$\gamma_{3st}(h_s,h_t) = \gamma'_{3s}(h_s) + \gamma'_{3t}(h_t) - k\gamma'_{3s}(h_s)\gamma'_{3t}(h_t) + a_sh_s^{\alpha} + a_th_t^{\beta} \quad (5.11)$$

Fractal model:

$$\gamma_{4st} = a_sh_s^{\alpha} + a_th_t^{\beta} \quad (5.12)$$

5.3.3 Spatio-temporal regularisation

5.3.3.1 Concept of catchments as space-time filters

Measurements are strongly affected by the measurement scale. Blöschl and Sivapalan (1995) formulated the measurement scale as a scale triplet: the distance between measurements (spacing); the size of the region over which measurements are available (extent); and the area or volume that each measurement represents (support). A coherent studies of measurement scale effects on parametric and non-parametric estimates of spatial correlation was presented in Chapter 2 and 3, respectively. As the support increases, the variable of interest becomes increasingly smoother. Because of this, the variance (and hence the sill of the variogram) decreases and the correlation lengths increase.

In this chapter, the catchment area is interpreted as the spatial support of the runoff measurements and conceptualise local runoff as a point process following Woods and Sivapalan (1999). In a joint spatio-temporal analyses both the spatial and the temporal supports need to be taken into account. In this chapter, the response time of a catchment is therefore interpreted as the temporal support. Runoff at the catchment outlet is then assumed to be some sort of aggregated value of local runoff over the catchment area (spatial support) over the catchment response time (temporal support).

The concept starts with local runoff or rainfall excess, $R(x,y,t)$. To account for routing on the hillslopes and in the channels within the catchment, a weighting function $u(x,y,t)$ is introduced which allows to combine local instantaneous runoff into runoff at the catchment outlet, Q_i :

$$Q_i(t) = \iint_{A_i} \int_{t-T_i}^{t} R(x,y,\tau)u(x,y,\tau)d\tau dxdy \tag{5.13}$$

where A_i is the area of catchment i, T_i is the time interval that influences the output, x and y are the space coordinates, t is time and τ is the temporal integration variable. The weighting function $u(x,y,t)$ represents the routing processes within the catchment and varies in space. For example, runoff generated close to the outlet or close to the streams will reach the outlet faster than runoff generated further away. Also, $u(x,y,t)$ will be a function of catchment characteristics such as hill slope orientation, catchment slope and soil types. $u(x,y,t)$, for a certain point in space, also changes with time as the flow velocities change with changes in the catchment state. As an approximation, it is in this chapter assumed that, for a given catchment, the weighting function is constant within the integration limits both in space and time, $u = 1/T_i$. For a constant weighting function, Eq. 5.13 becomes a linear filter or a convolution integral. In time, the weighting function is equivalent to a unit hydrograph that is constant between 0 and T_i and zero elsewhere. In space, the weighting function is constant within the catchment area and zero elsewhere which is consistent with the assumptions of Sauquet et al. (2000) and in Chapter 4. The specific runoff at the catchment outlet (runoff divided by catchment area) then becomes:

$$q_i(t) = \frac{1}{A_i T_i} \iint_{A_i} \int_{t-T_i}^{t} R(x,y,t-\tau)d\tau dxdy \tag{5.14}$$

The runoff routing process is hence conceptualised as a linear space-time filter in this chapter. For simplicity, it is assumed that the filter kernel in space is a square with area A_i (catchment size), and in time the filter kernel is a block unit hydrograph with time base T_i as mentioned above. A simple relationship between catchment response time and catchment area is assumed:

$$T_i = \mu A_i^{\kappa} \tag{5.15}$$

where μ and κ are parameters to be estimated from the data. For $\kappa>0$ the response time increases with catchment size. Eq. 5.15 embodies the space-time connections of the rainfall-

runoff process from a linear filter perspective. Note that Eq. 5.15 applies to runoff. For comparison, catchment precipitation was also analysed. The same aggregation procedure was used in space for these data, but with a constant temporal support of T_i = 15 min, as consistent with the raingauge data.

In a geostatistical framework, the linear aggregation of Eq. 5.14 is represented by the second moments. A point variogram of runoff represents the second moment of local, instantaneous runoff. From the point variogram with zero support in space and zero support in time (i.e. instantaneous) one can estimate variograms that are valid for finite support areas and finite support times by a procedure that is usually referred to as regularisation (Journel and Huijbregts, 1978). Conversely, it is possible to back-calculate the point/instantaneous variogram from variograms based on finite supports (Chapter 4). The point variograms are the basis of spatial estimation methods such as those of Sauquet et al. (2000) and the ones presented in Chapter 6 and 7. In addition, the point variogram sheds light on the spatio-temporal structure of instantaneous runoff generated at the local scale.

5.3.3.2 Implementation

The variogram value, given a certain distance, represents the expected variance of a process within an extent equal to this distance. If a variable is linearly aggregated, each measurement is the average of the point process within the support of the measurement. Assuming that the variance of catchment runoff is both dependent on the spatial and temporal supports (A and T, respectively, dropping the index), for two catchments of equal size the spatial regularisation technique of (Cressie, 1991, p. 66) can be extended to:

$$\gamma_{st}(h_s|a,h_t|T) = \frac{1}{A^2T^2}\int_A\int_A\int_T\int_T \gamma_{st}(|\mathbf{r}_1+\mathbf{h}_s-\mathbf{r}_2|,|\tau_1+h_t-\tau_2|)d\mathbf{r}_1 d\mathbf{r}_2 d\tau_1 d\tau_2 - \frac{1}{A^2T^2}\int_A\int_A\int_T\int_T \gamma_{st}(|\mathbf{r}_1-\mathbf{r}_2|,|\tau_1-\tau_2|)d\mathbf{r}_1 d\mathbf{r}_2 d\tau_1 d\tau_2 \quad (5.16)$$

where $\gamma_{st}(r,\tau)$ is the spatio-temporal variogram of the instantaneous point process, $\mathbf{h}_s$ is the separation vector between two catchments (with space lag $h_s = |\mathbf{h}_s|$), h_t is the time lag and a is the side length of the square that approximates a catchment, i.e., $a = \sqrt{A}$. The catchment size A has been taken as the median catchment size for all catchments of a given size class (Table 5.1). Eq. 5.16 indicates that the regularised variogram value between two catchments

of size A with response time T is the variance integrated in time and space between the two catchments, minus the integrated variance within one catchment. This concept is illustrated in Fig. 5.2. Each catchment is visualised as a spatio-temporal "volume" separated by spatio-temporal distances.

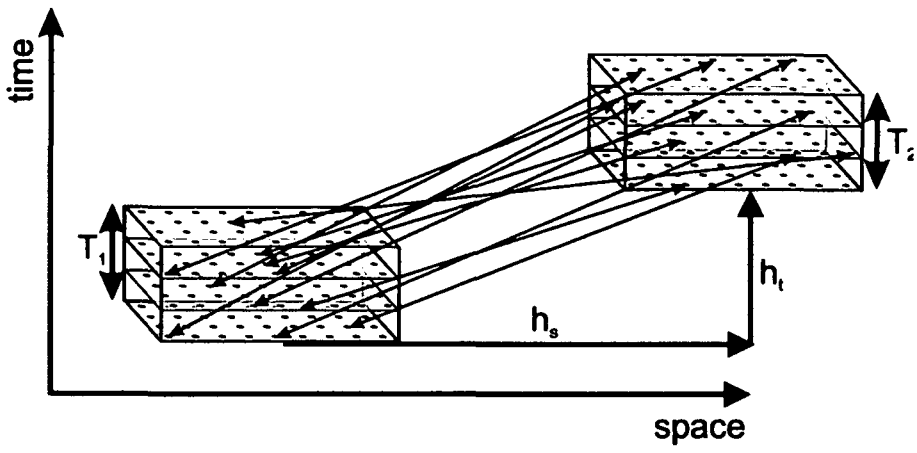


Figure 5.2 Schematic of variance estimation between two catchments 1 and 2 and a range of time lags. Thin arrows represent some of the spatio-temporal pairs of data points.

The number of integrals has been reduced here by using the distribution function of spatio-temporal distances within and between catchments in a similar way as Western and Blöschl (1999) and Chapter 4 but extended to space and time:

$$\gamma(h_s|a,h_t|T)=\int_{h_t-T}^{h_t+T}\int_{0}^{R\max}\gamma_{st}(r,\tau)f_{2st}(r|(h_s,a),\tau|(h_t,T))d\tau dr - \int_{-T}^{T}\int_{0}^{R\max}\gamma_{st}(r,\tau)f_{1st}(r|(h_s,a),\tau|(h_t,T))d\tau dr \quad (5.17)$$

$f_{1st}(r|a,\tau|T)$ is the probability density function (pdf) of distances in space and time within a catchment with spatial support a and temporal support T. $f_{2st}(r|(h_s,a),\tau|(h_t,T))$ is the pdf of distances in space and time between points in two catchments with a centre-to-centre distance h_s in space and h_t in time. $R_{\max}$ is a practical integration limit. The distances in space and time can be assumed to be independent, and then f_{1st} and f_{2st} can be separated into spatial and temporal parts:

$$\gamma_{st}(h_s|a,h_t|T)=\int_{h_t-T}^{h_t+T}\int_{0}^{R\max}\gamma_{st}(r,\tau)f_{2s}(r|(h_s,a))f_{2t}(\tau|(h_t,T))d\tau dr - \int_{-T}^{T}\int_{0}^{R\max}\gamma_{st}(r,\tau)f_{1s}(r|a)f_{1t}(\tau|T)d\tau dr \quad (5.18)$$

f_{1s} and f_{2s} are the pdfs in space which have been evaluated as in Western and Blöschl (1999) and in Chapter 2 and 3. f_{1t} and f_{2t} are the pdfs of the temporal distances, within and between catchments, respectively, which for a block unit hydrograph are:

$$f_{1t}(\tau|T)=\begin{cases}\frac{1}{T}\left(1-\frac{\tau}{T}\right) & \tau>0\\ \frac{1}{T}\left(1+\frac{\tau}{T}\right) & \tau\le 0\end{cases} \quad (5.19)$$

and:

$$f_{2t}(\tau|(h_t,T))=\begin{cases}(1-\frac{h_t}{T}+\frac{\tau}{T})/(h_t-\frac{h_t^2}{2T}+\frac{T}{2}) & 0\le\tau<h_t, 0<h_t<T\\ (1+\frac{h_t}{T}-\frac{\tau}{T})/(h_t-\frac{h_t^2}{2T}+\frac{T}{2}) & h_t\le\tau<h_t+T, 0<h_t<T\\ \frac{T-h_t+\tau}{T^2} & h_t-T\le\tau<h_t, h_t\ge T\\ \frac{T+h_t-\tau}{T^2} & h_t\le\tau\le h_t+T, h_t\ge T\end{cases} \quad (5.20)$$

5.3.4 Parameter estimation of variograms

The analyses are organised into two parts. In the first part, variogram models are fitted to the sample variograms of the small, medium and large catchment size classes independently (sections 5.4.1 and 5.4.2). In the second part, one point variogram model is fitted jointly to the three catchment size classes based on regularisation (sections 5.4.3 and 5.4.4).

First a version of the weighted least-squares (WLS) method (Cressie, 1985) was used to estimate the parameters of the variogram models by minimizing the objective function:

$$\Phi=\frac{1}{\sum_{i=1}^{ns}\sum_{j=1}^{nt}w(i,j)}\sum_{i=1}^{ns}\sum_{j=1}^{nt}w(i,j)\cdot\left[\frac{\hat{\gamma}_{st}(h_{si},h_{tj})}{\gamma_{st}(h_{si},h_{tj})}-1\right]^2 \quad (5.21)$$

where $\hat{\gamma}_{st}(h_{si}, h_{tj})$ is the sample variogram for one of the three catchment size classes or that of point rainfall (Eq. 5.1), $\gamma_{st}(h_{si}, h_{tj})$ is one of the variogram models (Eqs. 5.9-5.12), h_{si} and h_{tj} are the spatial and temporal lags, and *ns* and *nt* are the number of bins in space and time. $w(i,j)$ is the weight of each bin, with the indices i and j in spatial and temporal directions, respectively. The square root of the number of pairs in each bin was used as the weight, except that this weight was increased by a factor of 10 for $h_{si} = 0$ and $h_{tj} = 0$. These lags represent the marginal variograms in space and time. In a spatio-temporal estimation procedure, the marginal variograms will be important. As the bins on the margins only constitute approximately one tenth of the total number of bins in the spatio-temporal variograms, the increased weights balance the importance of the margins with the rest of the variogram. The SCEUA-method (Duan et al., 1992) was used to search for the best parameter set. The search was carried out ten times for each model type and catchment size class with different starting values, to reduce the probability of finding local minima. The variogram models associated with the smallest objective function of the ten trials are shown. The procedure was repeated for each catchment size class (including point precipitation), each variogram model and for precipitation and runoff separately.

In the second part, the parameters of a point variogram were estimated instead. For a certain point variogram, spatio-temporal variograms for the three catchment size classes were estimated by regularisation (section 5.3.3). These regularised variograms were jointly compared to the sample variograms of the three catchment size classes. The same objective function was used as above, but the summation was over all catchment size classes. Regularised variogram models associated with the smallest objective function of ten trials are shown. The procedure was repeated for each variogram model and for precipitation and runoff separately. The parameters κ and μ of Eq. 5.15 were also simultaneously fitted by this procedure, separately for each variogram model. The response time of the catchments is hence a result of the fitting procedure.

The scales of the diagrams of the spatio-temporal variograms are scaled linearly in terms of the bin spacing. As the bins have been selected approximately logarithmically (with the exception of zero lags) the axes are close to logarithmical.

5.4 Results

5.4.1 Separately fitted variograms of precipitation

The left column of Fig. 5.3 shows the spatio-temporal sample variograms of point and catchment precipitation, sorted by catchment size class. The total variance of precipitation is similar in time and space within the spatial and temporal extents of the data set (300 km, 1000 hrs shown here). The variogram values increase with increasing spatial and temporal distances which indicates the presence of spatial and temporal correlations as would be expected. There is a reduction in the variogram values as one moves from points to larger catchments which reflects the smoothing as a result of an increasing support. Columns two to five of Fig. 5.3 show the spatio-temporal variogram models that have been independently fitted to the sample variograms. For all models, with the exception of the fractal model, the visual fits are very good and the differences between the models are small. Fig. 5.4 shows the margins of the sample variograms and the fitted variogram models for precipitation. The margins of a spatio-temporal variogram are equivalent to the spatial and temporal variograms. The sample variograms are represented by points, while the fitted variograms are represented by lines. For a certain catchment size class, points and lines are of the same colour. All models, except for the fractal model, provide close fits. The shortest spatial lags show some differences between the models as this is where the models have been extrapolated beyond the data. Table 5.2 gives the values of the objective function for each variogram model and catchment size class as well as the average over the three size classes. The table also include the number of parameters to be fitted, including the two parameters of Eq. 5.15. The table indicates that the product-sum model can be best fitted to the sample variograms of precipitation. It should be noted that the product-sum model has the largest number of parameters, so the good fits may be both a result of a suitable model structure and the large number of degrees of freedom.

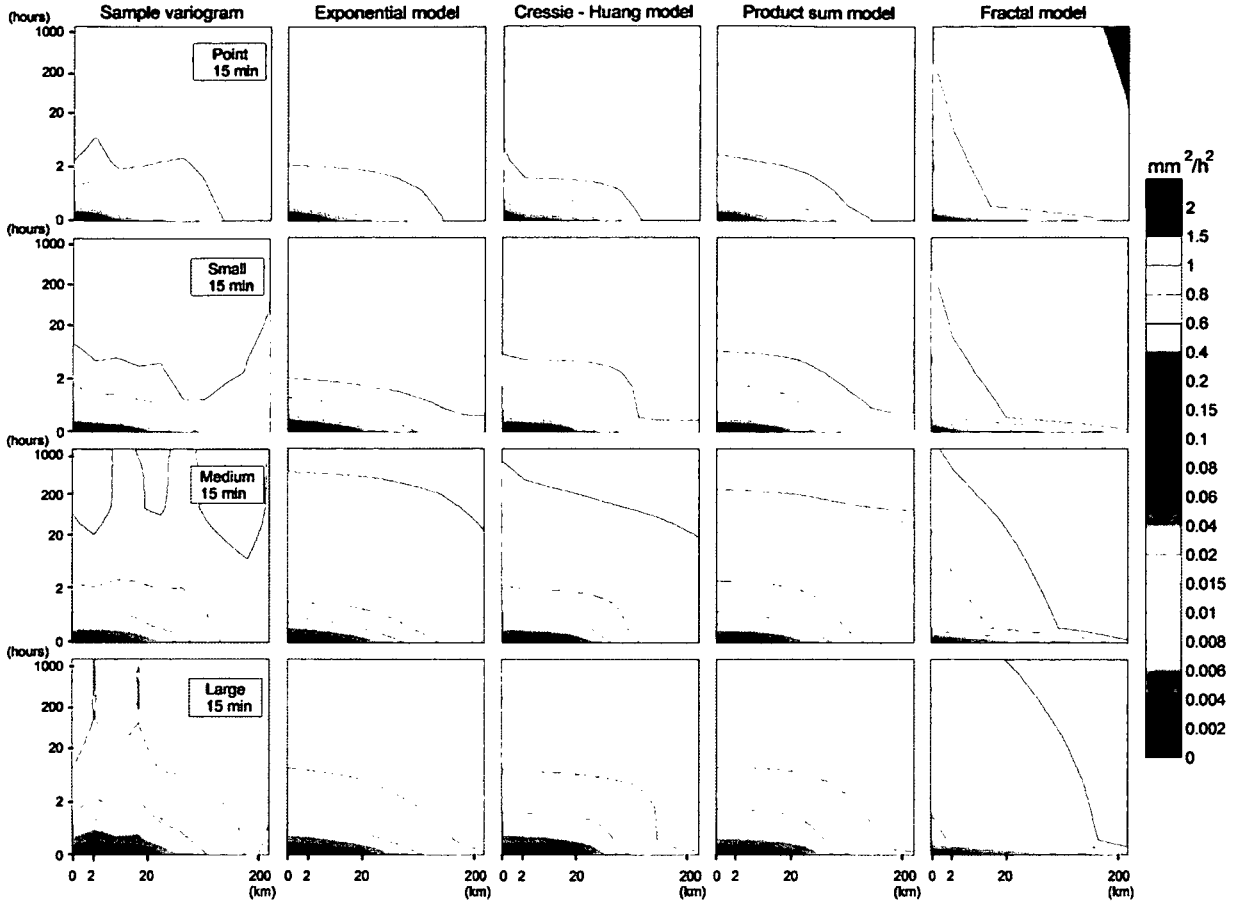


Figure 5.3 Spatio-temporal variograms of catchment precipitation. Sample variograms (left column) and independently fitted variogram models (columns 2-5). The rows relate to different catchment size classes (small, medium, large) including point precipitation in the top row. The horizontal axes are space lag, the vertical axes are time lag.

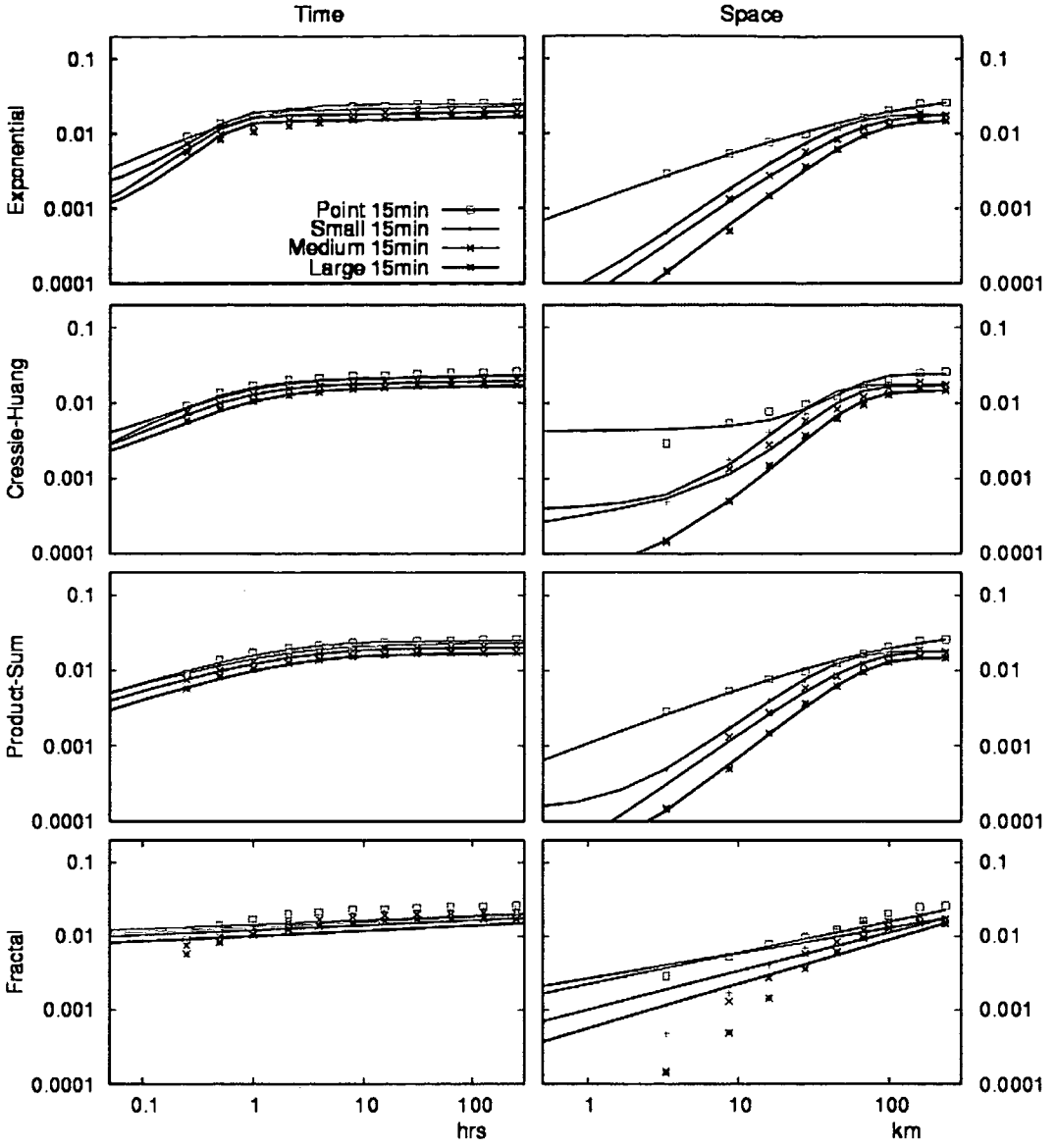


Figure 5.4 Temporal (left) and spatial (right) margins of the spatio-temporal variograms of catchment precipitation as in Fig. 5.3. Sample variograms are shown as points, fitted variogram models as lines.

Table 5.2 Objective function (Eq. 5.21) for the variogram models of precipitation fitted independently to spatio-temporal sample variograms for small, medium and large catchment size classes. Average refers to the average of the objective functions from the three catchment size classes. The number of parameters fitted consists of the parameters of the point variogram models and the two parameters of Eq. 5.15.

Variogram model	Point	Small catchments	Medium catchments	Large catchments	Average
Exponential model	0.0036	0.0132	0.0122	0.0123	0.0126
Cressie-Huang model	0.0121	0.0122	0.0125	0.0062	0.0103
Product-sum model	0.0036	0.0094	0.0062	0.0042	0.0066
Fractal model	0.0992	0.1285	0.1310	0.1307	0.1301

5.4.2 Separately fitted variograms of runoff

The left column of Fig. 5.5 shows the spatio-temporal sample variograms of runoff, sorted by catchment size class. The variograms indicate that there is a higher variance in space than there is in time within the spatial and temporal extents of the data set. The spatio-temporal variograms increase monotonously with spatial and temporal distances. There is a much stronger variance reduction effect between the variograms of the different catchment size classes than for precipitation. It is obvious that the catchment size has an efficient smoothing effect. Columns two to five of Fig. 5.5 show the variogram models that have been fitted separately for each catchment size class. All models can be fitted well to the sample variograms, with the exception of the fractal model, and the differences between the models are small. Fig. 5.6 shows the margins of the sample variograms and the fitted variogram models. The figure shows in more detail the much stronger variance reduction from smaller to larger catchments than that of precipitation. Table 5.3 gives the values of the objective functions for the fitted variogram models. The product-sum model offers a slightly better fit than the exponential and the Cressie-Huang models.

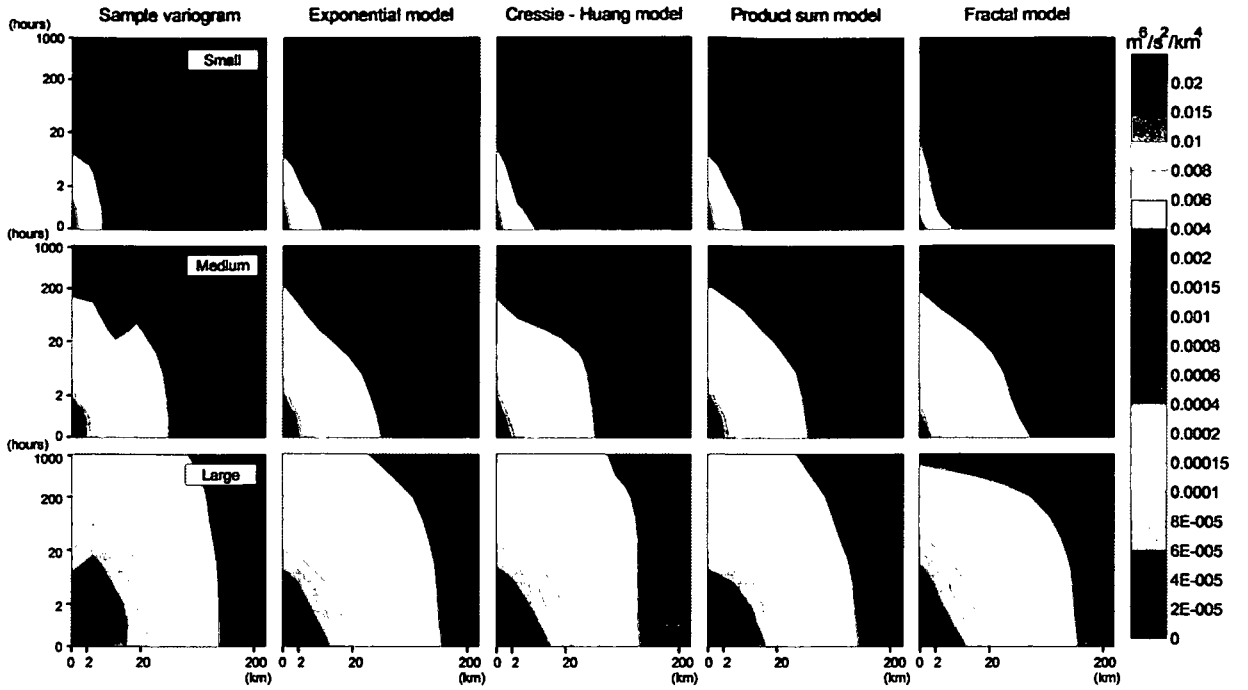


Figure 5.5 Spatio-temporal variograms of runoff. Sample variograms (left column) and independently fitted variogram models (columns 2-5). The rows relate to different catchment size classes (small, medium, large). The horizontal axes are space lag, the vertical axes are time lag.

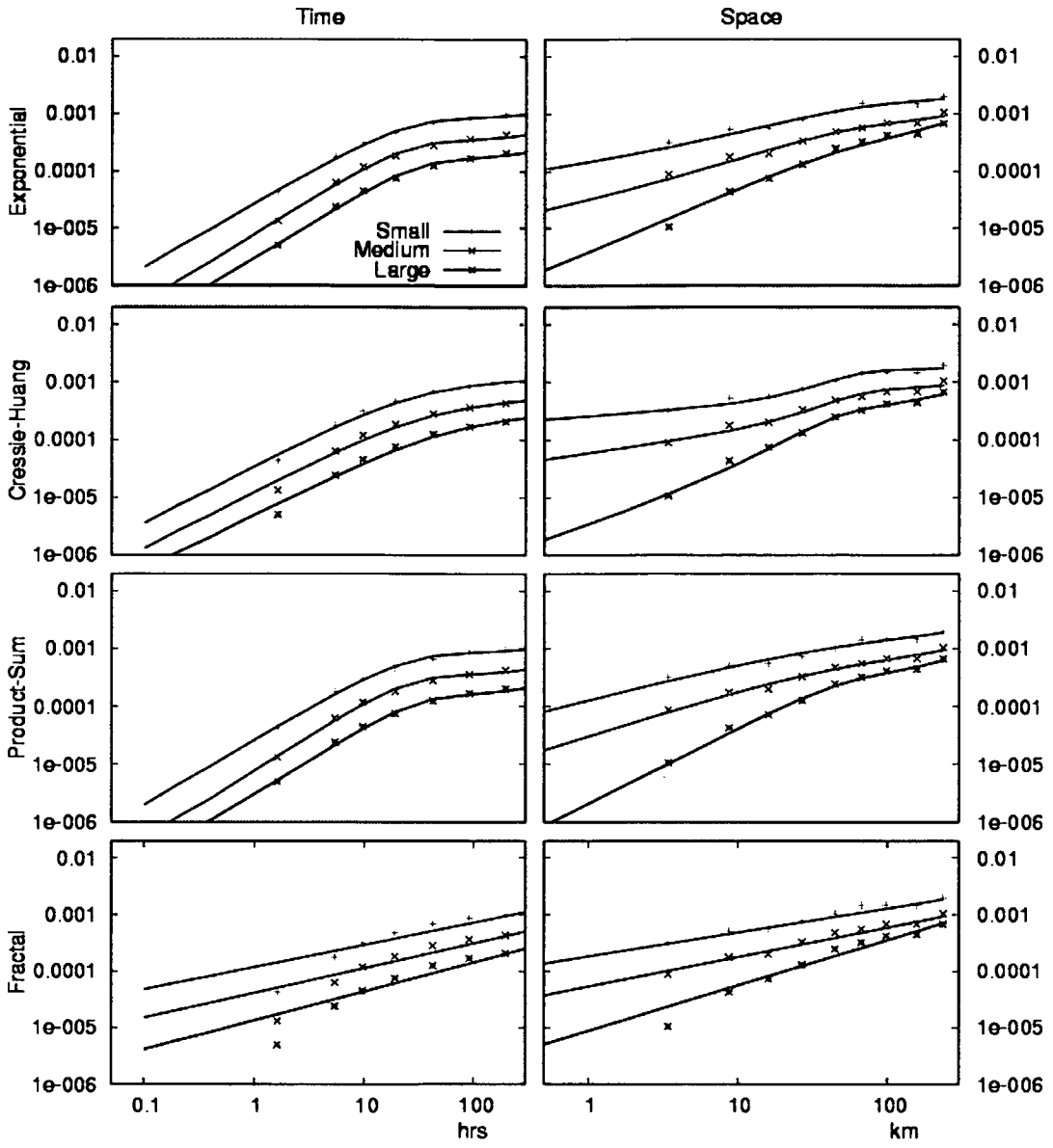


Figure 5.6 Temporal (left) and spatial (right) margins of the spatio-temporal variograms of runoff as in Fig. 5.5. Sample variograms are shown as points, fitted variogram models as lines.

Table 5.3 Objective function (Eq. 5.21) for variogram models of runoff fitted independently to spatio-temporal sample variograms for small, medium and large catchment size classes. Average relates to the average of the objective functions for the three catchment size classes.

Variogram model	Small catchments	Medium catchments	Large catchments	Average
Exponential model	0.0094	0.0095	0.0151	0.0113
Cressie-Huang model	0.0135	0.0209	0.0216	0.0186
Product-sum model	0.0082	0.0083	0.0110	0.0092
Fractal model	0.0851	0.0878	0.0982	0.0904

5.4.3 Jointly fitted variograms of precipitation

Fig. 5.7 shows the results of jointly fitting the variograms of precipitation to the three catchment size classes. The variograms in the left column are again the sample variograms. The letters on the left side of the figure relate to the respective rows and denote estimation (E), verification (V) and fitting (F). The sample variograms of rows three, four and five have been used for the fitting of the models in columns two to five. With the exception of the fractal model, there are only small visual differences between the fitted variograms. The sample variogram of row two (point data of precipitation with a temporal support of 15 minutes) can be used for verification. For this case, the differences between the models are slightly larger than for the fitting but the models are still rather close to the sample variogram, again with the exception of the fractal model. The top row of Fig. 5.7 shows the back-calculated point variograms valid for zero temporal and zero spatial supports, i.e. instantaneous point variograms. These do differ between the variogram models with the fractal and Cressie-Huang models giving larger sills than the other models. The exponential and product sum models are rather similar.

The margins of the variograms of Fig. 5.7 are shown in Fig. 5.8. The margins more clearly show that the overall fits are good to very good. The margins of the fractal model are less biased than the rest of the spatio-temporal variogram, especially along the spatial axis. As all catchment size classes have the same temporal support (15 min) the fractal model does not estimate any temporal variance reduction with increasing catchment size. The temporal

variograms indicate that the Cressie-Huang and product-sum models slightly underestimate the temporal variance of point precipitation with a temporal support of 15 min. The exponential model performs best on the margins.

Table 5.4 shows the values of the objective function for the fitted variogram models. The product-sum model offers a slightly better fit than the exponential and the Cressie-Huang models but for the verification case (point precipitation with temporal support of 15 min) the exponential model is the best model. The objective function for the goodness of fit (small, medium, large catchment sizes classes) is around 0.01 (with the exception of the fractal model) which is similar to the separate fitting (Table 5.2). This indicates that the regularisation is fully consistent with the catchment precipitation data. Note that the joint fitting (Table 5.4) has only one third of the free parameters of the separate fitting. This comparison tests the assumptions of regularisation in space, i.e., the approximation of the catchments by squares with an area equal to the median of each size class. It is clear that this approximation is sufficiently accurate for the purposes of regularisation. In the verification case (Table 5.4, right column) the errors are somewhat larger (0.01-0.03 depending on the model, excluding the fractal model) but in absolute terms this is still a small number.

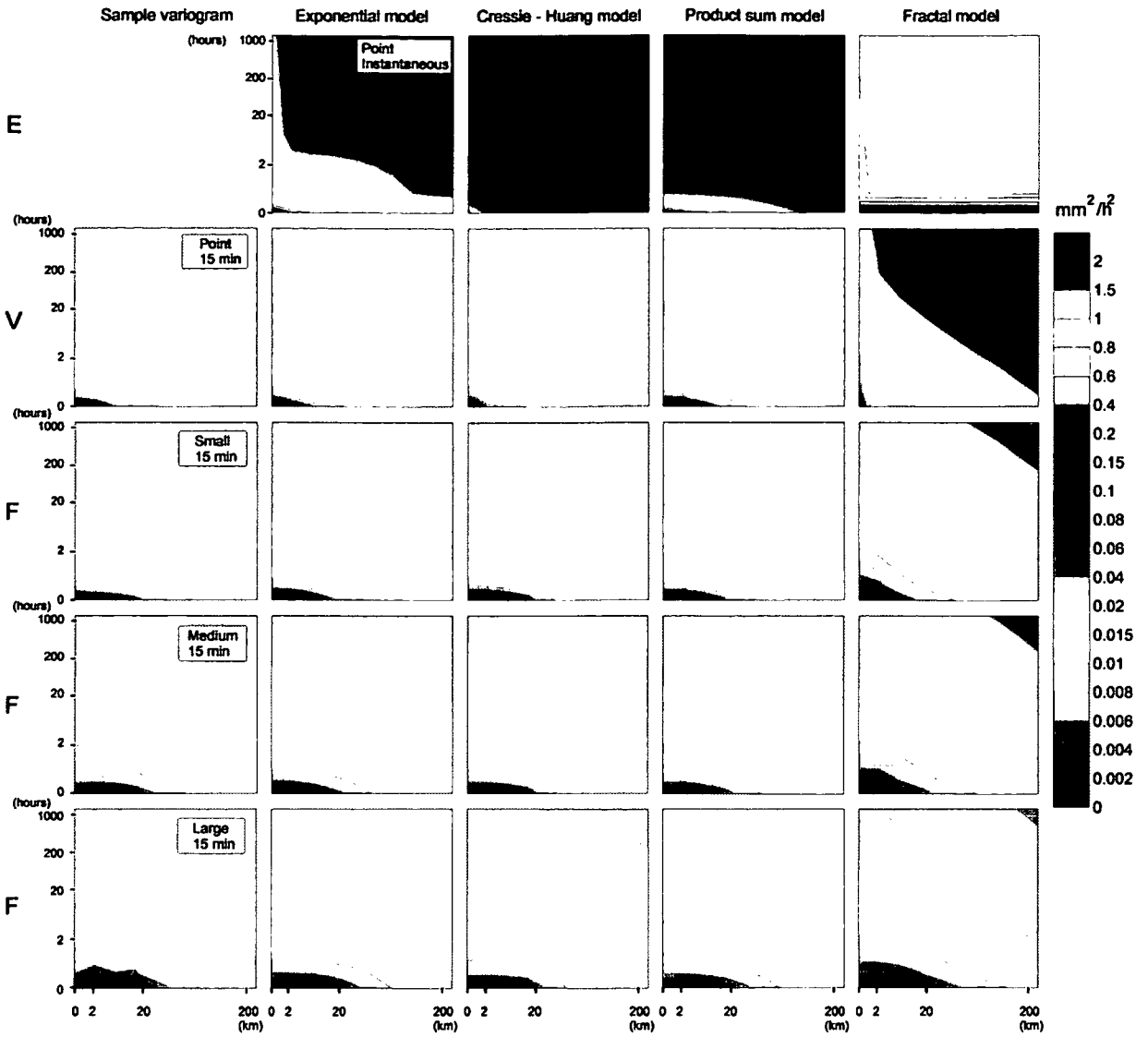


Figure 5.7 Spatio-temporal variograms of catchment precipitation. Sample variograms (left column) and jointly fitted variogram models (columns 2-5, rows 3-5). Row 2 (point precipitation, temporal support of 15 min) has not been used in the fitting and is used for verification. Top row shows the back-calculated variograms for zero spatial and temporal supports (instantaneous point precipitation). Letters "E", "V" and "F" stand for estimation, verification and fitting, respectively. The horizontal axes are space lag, the vertical axes are time lag.

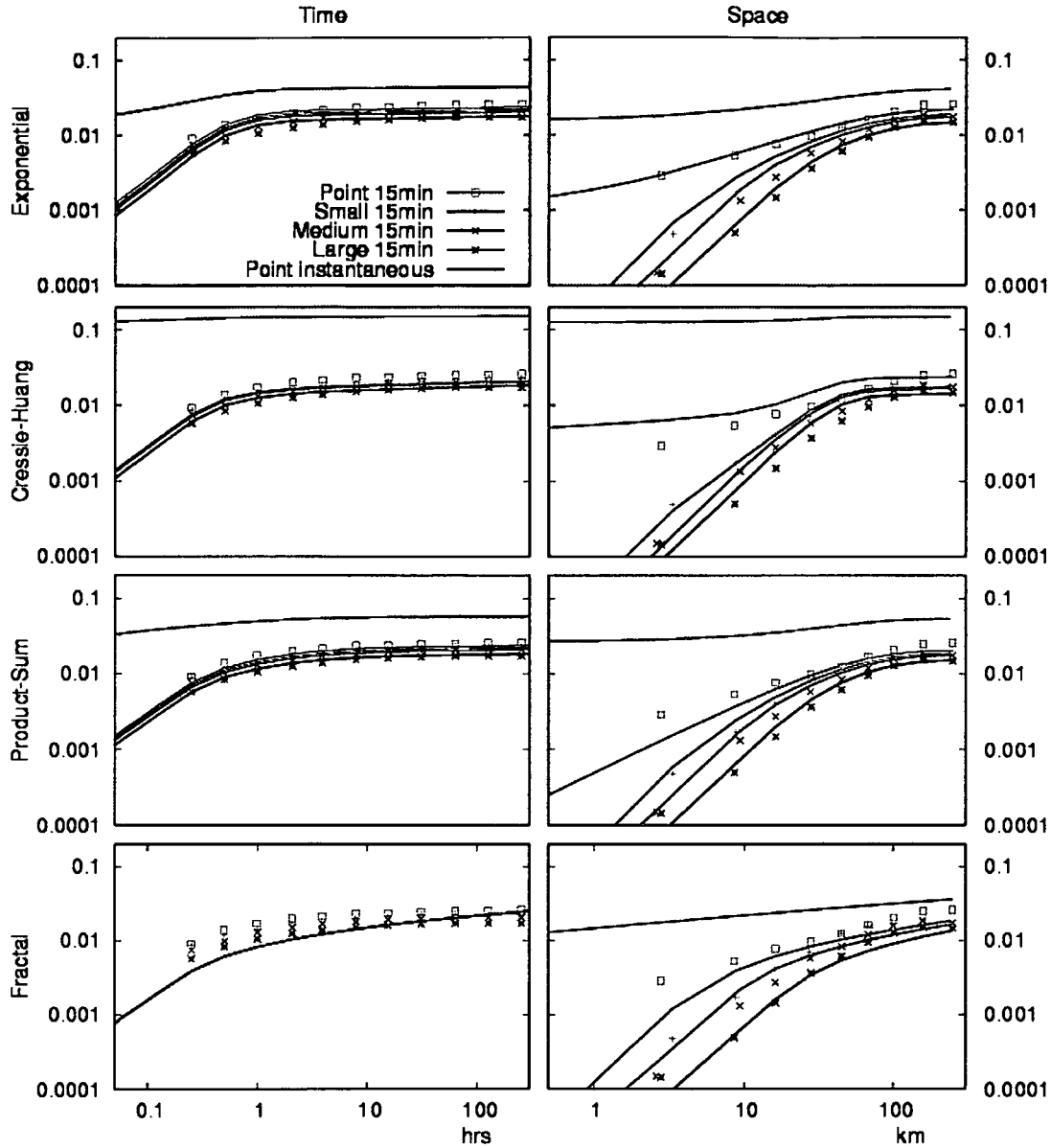


Figure 5.8 Temporal (left) and spatial (right) margins of the spatio-temporal variograms of catchment precipitation as in Fig. 5.7. Sample variograms are shown as points, jointly fitted variogram models (small, medium, large catchment size classes) as well as estimated variogram models (point 15 min, point instantaneous) as lines.

Table 5.4 Objective function (Eq. 5.21) for regularised variograms of precipitation fitted jointly to the spatio-temporal sample variograms for small, medium and large catchment size classes, denoted as "total". Point refers to the objective function for point precipitation with 15 min temporal support which is the verification case.

Variogram model	Total (fitted)	Point (verification)
Exponential model	0.0113	0.0098
Cressie-Huang model	0.0145	0.0191
Product-sum model	0.0094	0.0304
Fractal model	0.1210	0.2549

5.4.4 Jointly fitted variograms of runoff

Regularised spatio-temporal variogram models were fitted to the sample variograms of runoff jointly for all catchment size classes and are shown in rows two to four of Fig. 5.9. There are only minor differences between the regularised variograms from the different models, and they are all similar to the sample variograms. The exception is the fractal model which cannot be fitted as well. It should be noted that this is the model with the smallest number of parameters, so a poorer fit would be expected. The point variogram models back-calculated by the procedure (Fig.5.9, top row) exhibit significantly shorter spatial correlation lengths than any of the catchment scale variogram models. The point variogram models differ in terms of their sills (i.e. the overall levels). Similar to precipitation, the Cressie-Huang and fractal models have the largest sills. It is clear that there is substantial uncertainty associated with these variograms. However, for practical applications this may not be important if the spatio-temporal estimation of runoff is applied to catchments of a size range similar to that used here, as the regularised variograms based on these point variograms are all very similar.

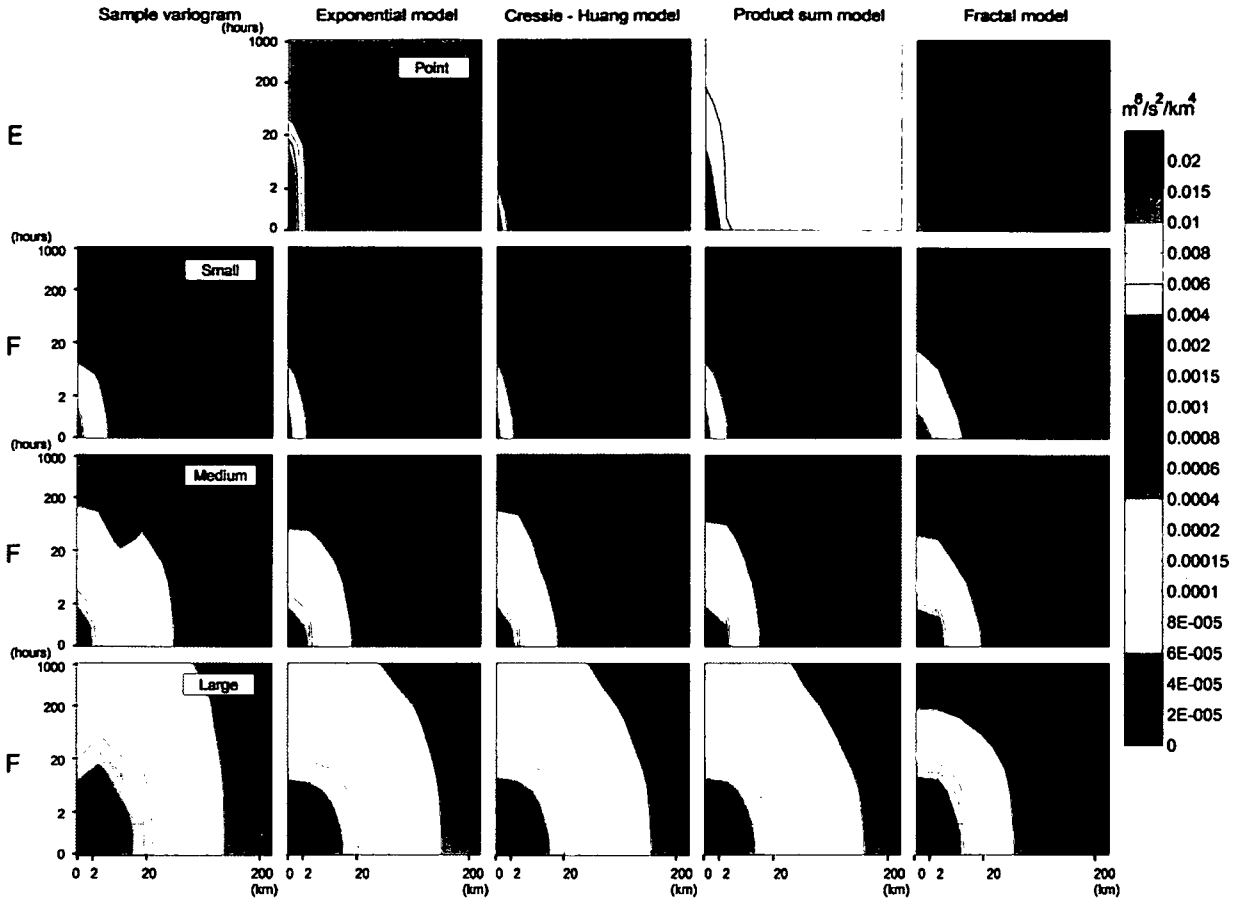


Figure 5.9 Spatio-temporal variograms of runoff. Sample variograms (left column) and jointly fitted variogram models (columns 2-5, rows 2-4). Top row shows the back-calculated variograms for zero spatial and temporal supports (instantaneous point runoff). Letters "E" and "F" stand for estimation and fitting, respectively. The horizontal axes are space lag, the vertical axes are time lag.

Fig. 5.10 shows the margins of the sample variograms and the fitted regularised models for runoff. There are only small differences between the exponential, Cressie-Huang and the product sum models. The temporal margins are almost perfectly modelled, while there are minor deviations between the spatial sample variograms and the estimated variograms. The point variograms are shown in light blue. For the fractal model, the point variogram is larger than the range shown. Table 5.5 indicates that the variogram models give almost equally good fits with the exception of the fractal model. The objective functions of the exponential, Cressie-Huang and product sum models range between 0.02 and 0.03. This is larger than those of the separately fitted variograms (around 0.01 in Table 5.3) which is likely related to the simplifications of the analysis including the assumptions on the unit hydrograph and the

general assumption of linearity. However, the absolute values of the objective functions for the three models are still very small indicating overall excellent consistency.

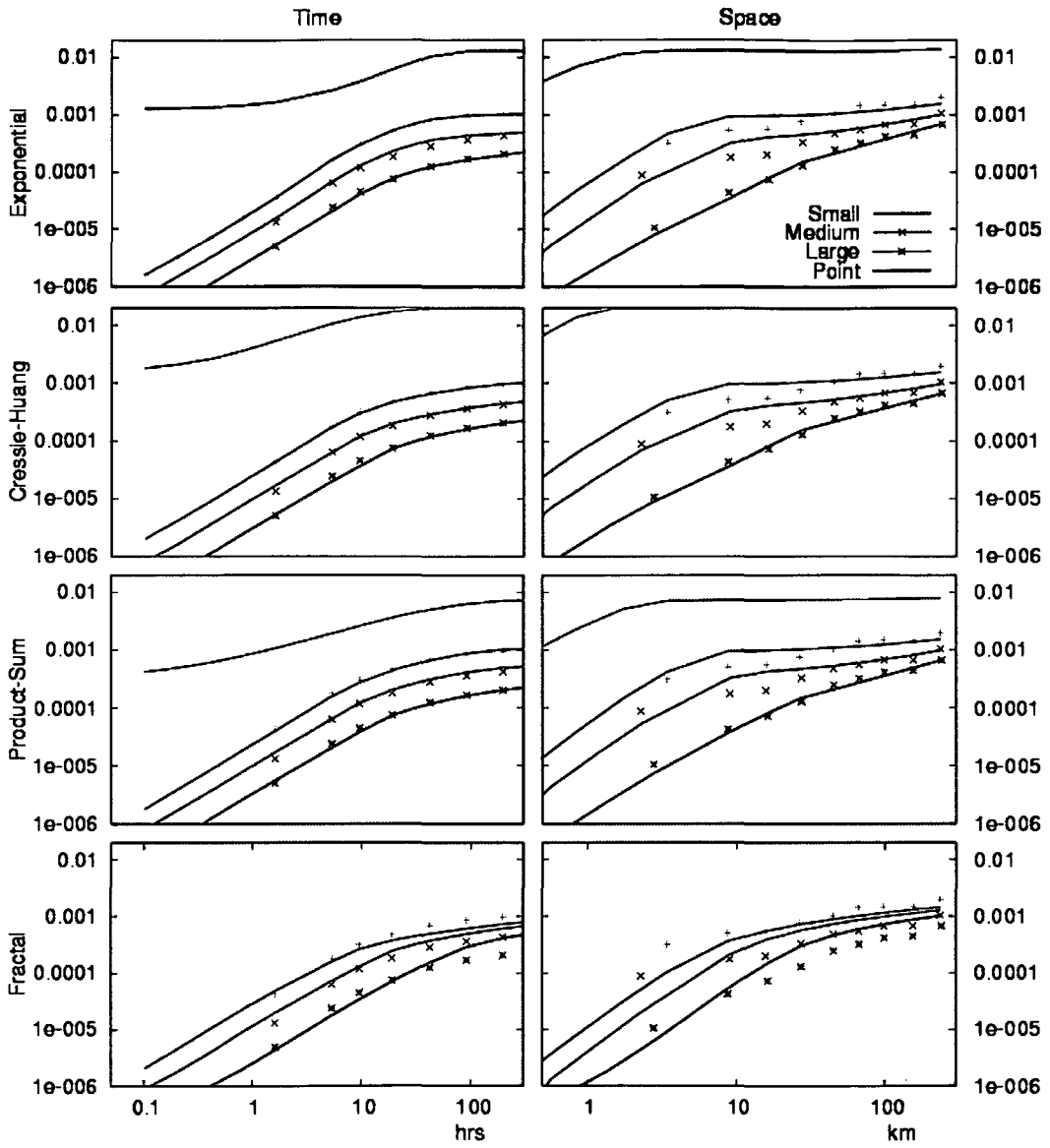


Figure 5.10 Temporal (left) and spatial (right) margins of the spatio-temporal variograms of runoff as in Fig. 5.9. Sample variograms are shown as points, jointly fitted variogram models (small, medium, large catchment size classes) as well as estimated variogram model (point instantaneous) as lines.

Table 5.5 Objective function (Eq. 5.21) for regularised variograms of runoff fitted jointly to the spatio-temporal sample variograms for small, medium and large catchment size classes.

Variogram model	Total
Exponential model	0.0269
Cressie-Huang model	0.0227
Product-sum model	0.0257
Fractal model	0.1544

All variogram models have been fitted ten times with different starting values, which produced somewhat different parameter sets. This is because of local minima in the objective function. The figures have all been based on the parameter sets with the smallest objective functions. To illustrate the uncertainty around these best fits, for each model, the five best parameter sets were selected. The average and the coefficient of variation (CV) was then computed for each parameter and for each model. The CV is a measure of the uncertainty of the parameters. These statistics are shown in Table 5.6, together with the values of the corresponding objective functions. Overall, the uncertainty depends on the parameter estimated. For some parameters, the uncertainty is very small (e.g. d_1) but for other parameters the uncertainty is substantial. These differences are related to the sensitivity of the shape of the variogram to individual parameters. The parameters with the smallest sensitivity have the largest uncertainty but this may not be important for spatio-temporal estimation. It was more difficult to find suitable parameter sets for the product-sum model than for the other models. This is because it is the model with the largest number of parameters. In order to obtain suitable parameters, the parameter search was initiated with parameter sets found in previous optimisation runs. Because of the presence of local minima this tends to reduce the variability of the estimated parameters. The CV values of the parameters of the product-sum model in Table 5.6 hence tend to be smaller than those of the other models. A parameter that is of particular interest is the exponent in the relationship between space and time supports, κ. The uncertainty of this parameter ranges between 1 and 38% depending on the model. The order of magnitude of κ is hence a meaningful estimate. With the exception of the fractal model, for which the fitting was not very good, the κ values of the different models are similar and range between 0.3-0.4. The parameters of the non-stationary parts (a_s, a_t, α and β) are not well constrained as they are controlled by the large time scale and space scale

variability present in the data. For the exponential and Cressie-Huang models, the levels (or sills) of the point variograms are defined by parameters a_1 and a_2, respectively. The a_2 value is significantly larger than a_1 reflecting the larger sills of the Cressie-Huang model as illustrated in Figs. 5.9 and 5.10.

Table 5.6 Statistics of the objective function (ϕ_k) of the joint fitting of the runoff variograms; estimated parameters of the space-time relationship of the supports (μ,κ); and estimated parameters of the point runoff variogram (remaining lines). CV is a measure of the uncertainty of the estimates.

	Exponential model		Cressie-Huang model		Product-sum model		Fractal model	
	Average	CV	Average	CV	Average	CV	Average	CV
ϕ_k	0.0293	0.0786	0.0281	0.1176	0.0258	0.0158	0.1548	0.0020
μ	1.8670	1.0422	2.5991	0.6058	2.0108	0.3027	0.3582	0.0710
κ	0.4193	0.3841	0.3440	0.2777	0.3065	0.1298	0.7936	0.0123
a_s	0.0000	1.2773	0.0007	0.9102	0.0000	0.1613	0.0840	0.5811
a_t	0.0024	1.4488	0.0005	1.0493	0.0001	0.4936	0.0277	0.6813
α	0.5909	0.2939	0.2287	1.0096	0.6718	0.0385	0.0050	0.5056
β	0.1245	0.9893	0.0972	0.4798	0.1847	0.3471	0.0076	0.5847
a_1	0.0131	0.1046						
c_1	0.0295	0.3589						
d_1	1.0298	0.0198						
e_1	1.6427	0.0548						
a_2			0.0256	0.2456				
c_2			0.1755	0.3966				
d_2			1.2517	0.0976				
a_{3s}					0.0070	0.0773		
a_{3t}					0.0070	0.0798		
d_{3s}					1.6841	0.0805		
d_{3t}					31.6109	0.1490		
e_{3s}					1.6814	0.0090		
e_{3t}					0.5550	0.1742		
k					142.8483	0.0780		

5.5 Discussion and conclusions

5.5.1 Sample variograms

A comparison of the spatio-temporal variograms of runoff and precipitation indicates that, for a given catchment size class, the variograms are fundamentally different. The left column of Fig. 5.5, as compared to the left column of Fig. 5.3, suggests that the plots of the runoff variograms are much more elongated in the time direction indicating that the time correlations of runoff are much more persistent than those of precipitation. Obviously, this is because of the time delays as rainfall passes through the catchment system. This is an effect of the catchment operating as a filter to the atmospheric forcing, with the time scale of the filter being directly related to the concentration time of the catchment. The contour lines of the variogram values give an indication of the characteristic velocities, of the type found in Chapter 4. For precipitation of all catchment classes, a typical pair of length and times scales is 70 km and 2 hrs which suggests a typical characteristic velocity of 10 m/s. This is similar to the characteristic velocities found in Chapter 4 and consistent with the schematic of space time scales of Blöschl and Sivapalan (1995). For runoff, again for all catchment size classes, typical pairs of length and times scales are 2 km and 2 hrs, 20 km and 20 hrs, and 50 km and 100 hrs. This translates into typical characteristic velocities of 0.27, 0.27 and 0.14 m/s, respectively. These characteristic velocities are somewhat faster than those found in Chapter 4, which may be related to the higher temporal resolution of the data. The data resolve the event scale in more detail, hence one would expect the estimated scales to be associated with events. The slower characteristic velocities with increasing catchment size are likely related to the larger groundwater contribution in larger catchments.

Independently fitting variogram models to each catchment size class gave excellent to good fits for all the variogram models considered here with the exception of the fractal model. The product-sum model was generally better than the other models for both runoff and precipitation. The differences in the goodness of fit may be partly related to the degrees of freedom; the fractal model has the smallest number of parameters, the product-sum model the largest number of parameters. The objective function is dimensionless, so a comparison of precipitation and runoff is meaningful. The objective functions for runoff and precipitation are similar (both around 0.01 in Tables 5.2 and 5.3) indicating that the variogram models can be fitted equally well to runoff and precipitation.

The variograms change as one moves from small to medium sized and large catchments. The catchment scale effects are significantly larger for the case of runoff than for precipitation, i.e., in the case of runoff the variance reduction with catchment area is much larger (Figs. 5.3 and 5.5). Also, the temporal correlations increase more strongly with catchment area which, again, is related to the travel time of water in the catchments. The stronger time aggregation effects of runoff, as compared to precipitation, may explain the stronger variance reduction with spatial scale than that predicted by spatial aggregation, found by Woods et al. (1995).

5.5.2 Space-time regularisation

The regularisation of precipitation is used here for two purposes; to separate the spatial aggregation effects (moving from point rainfall to catchment rainfall) from spatio-temporal effects that involve runoff routing; and to test the spatial aggregation procedure, in particular the assumption of approximating catchments by squares and the use of a constant catchment size equal to the median in each size class. The comparison of back-calculated point precipitation (zero spatial support, 15 min temporal support) in Fig. 5.7, second row suggests that the assumptions are indeed appropriate for the data set used here, so the spatial regularisation of runoff, that uses the same procedure, is also valid. The objective functions of the joint fitting (Table 5.4) are close to the average objective functions obtained by the separate (direct) fitting (Table 5.2), i.e., around 0.01 in both cases (exponential, Cressie-Huang and product-sum models). This further corroborates the validity of the regularisation procedure.

For the case of runoff, however, the objective functions of the joint fitting (Table 5.5) are larger than those of the separate fitting (Table 5.3). For the joint fitting, the errors of the exponential, Cressie-Huang and product-sum models range between 0.02 and 0.03, depending on the model, while they are around 0.01 for the separate fitting. This means that, for runoff, the space-time aggregation effects of catchments are not fully consistent with the assumptions made here. Specifically, the simplifications include the assumptions of a block unit hydrograph, the general assumption of linearity and, perhaps most importantly, a single relationship between catchment size and catchment response time. However, the overall magnitude of the objective functions are still very small (i.e., errors of 0.02-0.03) indicating that the first order effects of the spatio-temporal variability of runoff is indeed captured well by conceptualising catchments as linear space-time filters. The stronger time aggregation

effects of runoff, as compared to precipitation, have been represented by a relationship between spatial and temporal supports (Eq. 5.15) which seems essential in representing the change of spatio-temporal runoff variograms with changing catchment size.

For precipitation, the results indicate that the point scale product-sum model provided slightly better variograms than the other models in terms of the goodness of fit to the small, medium and large catchment size classes but there was no advantage over the other models in the verification case of 15 min point precipitation. The fit of the fractal model was poorest but it was the model with the smallest number of parameters. There are two reasons for the poor fit of the fractal model. The obvious one is the smallest number of parameters among all variogram models, so the fractal model has the least flexibility. The other, probably equally important, reason is the lack of space-time interaction of the spatio-temporal fractal variogram, i.e., the fact that the partial derivatives of the variogram $\frac{\partial \gamma}{\partial h_s}$ and $\frac{\partial \gamma}{\partial h_t}$ only depend on h_t and h_s, respectively. This lack of space-time interaction also concerns the fractal part of the other variograms, but to a lesser degree, as it only relates to a component of the entire variogram. For runoff, the goodness of fit of the exponential, Cressie-Huang and product-sum models was good (0.02-0.03) suggesting that all three models are suitable for the spatio-temporal estimation of runoff in the study area. Because of the small differences between the models, the choice of model could be based on computational convenience. It is interesting that the product-sum model reduces to a separable model with the fitted parameters, i.e. $a_{3s} = a_{3t} = 0.0070$ and $k = 142$, which is very close to the condition for the product-sum model to reduce to the separable model of Rodríguez-Iturbe and Mejía (1974). Separable models are computationally more convenient for some applications (Fuentes, 2006). The spatial variogram fits in this chapter (Fig. 5.10 right column) are as good as or better than those in Chapter 4 (Fig. 4.6 b) where only spatial aggregation was applied. In addition, these variograms describe the temporal aggregation effects well (Fig. 5.10 left column).

5.5.3 Interpretation of point variograms of runoff

The point variograms of runoff, i.e. the variograms for a local runoff generation process with zero spatial and temporal supports, differ between the models. The fractal model gives the highest point variogram. This model, however, should be treated with caution as the model fits are not very good. For the Cressie-Huang model, the overall level or sill is higher than for the exponential and product sum models. The point model has been estimated from catchment size classes of 42, 119 and 605 km². These are the spatial supports. The associated temporal supports are, depending on the model (Eq. 5.15, Table 5.6), approximately 7, 11 and 20 hours. This means that the back-calculation procedure involves substantial extrapolation to smaller scales, so the differences between the variogram models are not surprising. The shapes of the three point models are, however, not too different (Fig. 5.10 and Fig. 5.9 top row). It should also be noted that for estimation purposes one is usually interested in catchment sizes that are not much smaller than the smallest catchments considered here, e.g., 1 km². For these catchment sizes, the variograms are much more similar. For the practical application of spatio-temporal estimation methods in catchment hydrology the differences in the point variograms may hence not be important.

The correlation lengths of the back-calculated point variograms of runoff are on the order of a kilometre or less, while the small catchments showed correlation lengths of around 10-20 kilometres (Fig. 5.10 right column). A similar value of 0.7 km for point variograms of runoff was found in Chapter 4. The short correlation lengths are plausible as local runoff will likely vary much over short spatial scales because of the variability of local infiltration and soil moisture characteristics (Western et al., 2002, Western et al., 2004). It is also of interest to compare the sills or overall levels of the point precipitation and point runoff variograms. For the exponential, Cressie-Huang and product sum models of point/instantaneous precipitation the sills in space are 0.04, 0.15 and 0.06 $mm^2 \cdot h^{-2}$, respectively (Fig. 5.8 right column). The corresponding values for runoff (Fig. 5.10 right column, with units adjusted) are 0.14 0.26 and 0.12 $mm^2 \cdot h^{-2}$, respectively. This means that the local variability of runoff is between twice and three times the variability of local rainfall. This is plausible as temporal and spatial soil moisture variability contributes to making local runoff more variable than rainfall. In time, local runoff is more coherent than rainfall (Fig. 5.10 left column as compared to Fig. 5.8 left column). This, again, is plausible because of the memory induced by soil moisture and local ponding. The non-stationary (fractal) parts of the variograms are more difficult to interpret. The parameters differ between the variogram models which is likely a result of the

interdependence of the parameters of the fractal part and the other parameters of the variogram models. As the levels of the stationary parts of the point variograms differ, so will the non-stationary parts in the different models.

5.5.4 Catchments as space–time filters

The high temporal resolution of the data used here (15 minutes) made it possible to analyse the connections of space-time variability in more detail than what was possible in Chapter 4, where daily data was used. A time step of 15 minutes resolves individual events even in the small catchment class. The kernel or space-time filter characteristics shed light on the space-time scaling behaviour of the rainfall-runoff transformation (Eq. 5.15). The parameter that is of particular interest is the exponent of the relationship between space and time supports, κ. With the exception of the fractal model, for which the fitting was not very good, the κ values of the different models are similar and range between 0.3-0.4. With a μ value of around 2, this range of κ gives response times of about 5 hours and one day for catchments of 10 and 1000 km² area, respectively. These are plausible event response times for the catchments considered here and are shorter than the low flow recession time scales to be expected in the catchments. The estimated response times seem to reflect large events with relatively short response times, as these events are associated with large absolute differences in the data set and hence contribute most to the second statistical moment. The estimated range of κ represents the average scaling characteristics of catchments within the study region. Obviously, for an individual catchment, the response time may deviate significantly from the general relationship of Eq. 5.15. Fig. 3 of Merz and Blöschl (2003), for example, shows a map of the deviations from such a general relationship based on an exponent of κ = 0.35. However, interpreted as an ensemble average, the range found here is fully consistent with analyses of observed runoff response in numerous catchments (e.g. Fig. 11.4 of Pilgrim, 1987, Anderson and Burt, 1990, Corradini et al., 1995). Similarly, typical values used in regional modelling studies are within this range. For example, Blöschl and Sivapalan (1997) used an exponent of κ = 0.35 in analysing the spatial scaling behaviour of flood frequency. The consistency suggests that the concept of catchments as space-time filters is indeed meaningful. It should be noted that, if known, the deviations of catchment response times from a general relationship could be included in the space-time filter framework proposed here.

Spatio-temporal point variograms of runoff as derived in this chapter can be used in spatio-temporal estimation models, as will be showed in Chapter 7. This method would be particularly useful for filling in missing data of streamflow records based on the records in neighbouring catchments. Clearly, the approach would be expected to work best if the density of the stream network is high as is the case in Austria. If fewer stream gauges are available the variograms will not be as well defined (see chapter 2 and 3 for sampling issues) while knowledge about the physical characteristics of the catchment and climate systems will become relatively more important. The precipitation variograms were treated separately from the runoff variograms in this chapter. Another extension of the work reported here would be to combine these two analyses. This could be based on similar concepts as those proposed by Woods and Sivapalan (1999).

Chapter 6. Top-kriging – geostatistics on stream networks

6.1 Introduction

The Problem of Ungauged Basins (PUB) (Sivapalan et al., 2003) is one of the central problems in hydrology and related sciences. The problem consists of estimating streamflow-related variables at locations where no measurements are available. Estimates can be obtained by a range of methods (e.g. Blöschl, 2005). A particularly appealing set of methods are geostatistics, which allow estimation of a variable including its uncertainty at locations where no measurements are available (Journel and Huijbregts, 1978). The main advantage of geostatistical methods is that they are best linear unbiased estimators (BLUE); best meaning that the mean squared error is a minimum, linear meaning that the estimate is a weighted mean of the data in the area, and unbiased meaning that the mean expected error is zero (Journel and Huijbregts, 1978, p. 304). Geostatistical methods have evolved in the mining industry. The main problem consisted of estimating the expected ore grade (and its uncertainty) of a block using point samples of the ore grade in the area. To this end the spatial correlations of pairs of points are plotted versus their Euclidian distance. From this, the variogram is estimated which is then used to estimate the variable at the location of interest for a given block size from the point samples (Journel and Huijbregts, 1978).

The problem in catchment hydrology is quite different. The main difference is that catchments are organised into subcatchments. Unlike mining blocks, catchments are nested. Water follows a stream network. It is therefore clear that upstream and downstream catchments would have to be treated differently from neighbouring catchments that do not share a subcatchment. Therefore Euclidian distances between catchments are not the natural way of measuring the spatial distance of catchments. Estimation of variables on stream networks needs to use a topology that is different from the usual Euclidian topology.

Most applications of geostatistics to catchment hydrology, so far, have indeed used Euclidian distance between catchments, usually measured as the Euclidian distance between the gauges or the catchment centroids (e.g. Daviau et al., 2000, Adamowski and Bocci, 2001, Eaton et al., 2002, Merz and Blöschl, 2005). Given the obvious nested structure of catchments it is surprising that very little research has been done on extending geostatistical concepts to

catchments. There is one notable exception. Gottschalk and co-workers (Gottschalk, 1993a, b, Sauquet et al., 2000) have addressed this very issue. Gottschalk (1993a) first developed a method for calculating covariance along a river network and used this for interpolation along the network (Gottschalk, 1993b). Sauquet et. al. (2000) further developed this method for mapping annual runoff along the river network using water balance constraints in the estimation procedure.

In this chapter, it is proposed a method of geostatistical estimation on stream networks that builds on the work of Sauquet et. al. (2000). It extends the original work in a number of ways. Firstly, the interpolation method can be used, in an approximate way, for a range of streamflow-related variables including variables that are not fully mass conserving. Sauquet et. al. (2000) interpolated mean annual runoff which is a mass conserving variable. Second, here it will be used variograms while Sauquet et. al. (2000) used covariances. This makes it possible to deal with variables that are non-stationary. Third, the method accounts for local uncertainties of the measurements that may differ between locations. This makes it possible to exploit short records. Last, the method can be used to estimate the uncertainty of the variable of interest in ungauged catchments, which also will be illustrated.

Section 6.2 contains a review the basic concepts, before the methodology is presented in detail. Section 6.3 illustrates the approach for the case of estimating the 100 year specific flood in ungauged catchments in Austria. This includes a comparison of the estimates with Ordinary Kriging as well as an analysis of the estimation uncertainties in ungauged catchments. Section 6.4 summarises the main implications for hydrological regionalisation.

6.2 Method of Top-kriging

6.2.1 Concepts of Top-kriging

There are two main groups of variables that control streamflow (Fig. 6.1). The first group consists of variables that are continuous in space. These variables include rainfall, evapotranspiration and soil characteristics. They are related to local runoff generation. In this context, runoff generation is conceptualised as a point process, i.e. runoff generation is assumed to exist at any point in the landscape. This concept is discussed in Woods and Sivapalan (1999). In a similar way, other streamflow-related variables can be conceptualised as continuous point processes on the local scale. For characterising these variables, Euclidian

distances are appropriate. The spatial statistical characteristics of the point variables can be represented by the variogram (Chapter 4):

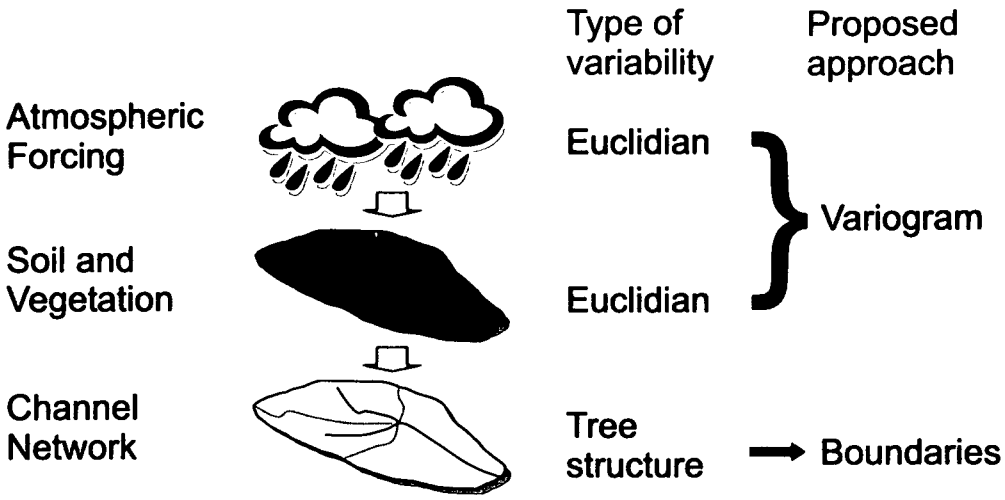


Figure 6.1 Atmospheric forcing and soil and vegetation contribute to the runoff generation process locally and can be represented by point variograms. The channel network organises runoff into streams, which can be represented by the catchment boundaries.

The second group of variables is related to routing in the stream network. These variables are affected by the catchment organisation of nested catchments where runoff accumulates along the stream network. Variables of this type include mean annual discharge, flood characteristics, low flow characteristics, concentrations, turbidity and stream temperature. These variables are only defined for points on the stream network. They cannot be represented by Euclidian distances. Rather they need to be represented by methods that reflect the tree structure of the stream network.

The proposed method combines these two groups of variables in a geostatistical framework. The method is referred to as topological kriging or Top-kriging, as it takes into account the topology of stream networks and nested catchments. The continuous process in space defined for point variables is represented by a variogram. The channel network structure and the similarity between upstream and downstream neighbours are represented by the catchment area that drains to a particular location on the stream network. The catchment areas are defined by their boundaries in space.

6.2.2 Kriging basics

In Euclidian kriging methods (such as Ordinary Kriging), the variable of interest is represented as a random field of values $z(\mathbf{x})$. Spatial similarity is represented by the variance between pairs of points as a function of their Euclidian distance. Kriging is then the best linear unbiased estimator, i.e. an interpolation method where the expected bias is zero and the expected kriging error is minimised. An unknown value $\hat{z}(\mathbf{x}_0)$ of the variable at position $\mathbf{x}_0$ (i.e. the target position) can be estimated as a weighted average of the variable measured in the neighbourhood:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i) \tag{6.1}$$

λ_i is the interpolation weight of the measurement at position $\mathbf{x}_i$ and n is the number of neighbouring measurements used for interpolation. The weights λ_i can be found by solving the kriging system:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \gamma_{ij} - \lambda_i \sigma_i^2 + \mu &= \gamma_{0i} \qquad i = 1,\ldots,n \\ \sum_{j=1}^{n} \lambda_j &= 1 \end{aligned} \tag{6.2}$$

The γ_{ij} refers to the gamma value or the expected semivariance between two measurements i and j, as found from a theoretical semivariogram model. μ is the Lagrange parameter. σ_i^2 represents the measurement error or uncertainty of measurement i. The use of measurement errors in the kriging equations is termed kriging with uncertain data (KUD) (de Marsily, 1986 p. 300, Merz and Blöschl, 2005).

6.2.3 Interpolation and regularisation over catchment boundaries

In Top-kriging, the measurements are not point values but are defined over a non-zero catchment area A. In geostatistical terminology, A is the support. A point variable $z(\mathbf{x})$ can be averaged over an area A as:

$$\bar{z}(A) = \frac{1}{A} \int_A w(\mathbf{x}) z(\mathbf{x})\, d\mathbf{x} \tag{6.3}$$

where $\bar{z}$ is the spatially averaged variable and $w(\mathbf{x})$ is a weighting function. If there is reason to assume the variable is conservative, or approximately conservative, as is done by the

suggested approach, the aggregation is linear, and $w(\mathbf{x}) = 1$. If a non-zero support A is accounted for, the kriging system remains the same, but the gamma values between the measurements need to be obtained by regularisation (Cressie, 1991, p. 66). Assuming the existence of a point variogram γ_p, the gamma value or the semivariance between two measurements with catchment areas A_1 and A_2, respectively, is:

$$\gamma_{12} = 0.5 * Var(z(A_1) - z(A_2)) = \frac{1}{A_1 A_2} \int_{A_1} \int_{A_2} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 - \\ 0.5 * \left[\frac{1}{A_1^2} \int_{A_1} \int_{A_1} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 + \frac{1}{A_2^2} \int_{A_2} \int_{A_2} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 \right] \tag{6.4}$$

$\mathbf{x}_1$ and $\mathbf{x}_2$ are position vectors within each catchment used for the integration. The first part of this expression integrates all the variance between the two catchments, while the second part subtracts the averaged variance within the catchments. The second part is the smoothing effect of the support, which indicates that the variance of the averaged variable decreases as the support area increases. Eq. (6.4) can be used to estimate the variogram of the averaged variable from the point variogram. This procedure is termed regularisation although most textbooks use the term only for the case when $A_1 = A_2$. The gamma values are inserted into the kriging matrix Eq. (6.2) and the kriging system can be solved in the normal way to calculate the weights λ_i for the interpolation scheme. The important thing in Top-kriging is that the integration is performed over the catchment area that drains to a particular location on the stream network. The location on the stream network is the outlet of the target catchment. It is hence possible to perform geo-statistics on stream networks.

The integration in Eq (6.4) will, in most cases, be either very complex or impossible to carry out analytically. The integrals have to be replaced by sums and the catchment area is discretised by a grid. It is important to note that the grid has to be the same for each catchment every time it is discretised (Isaaks and Srivastava, 1989, p. 326). In fact, even slight differences of the grids or randomly chosen points will cause numerical problems that are likely to flaw the results as test simulations with the examples shown later have indicated. Figure 6.2 shows a schematic of two nested catchments, their discretisation by a square grid, and the distances between the discretised points within the catchments.

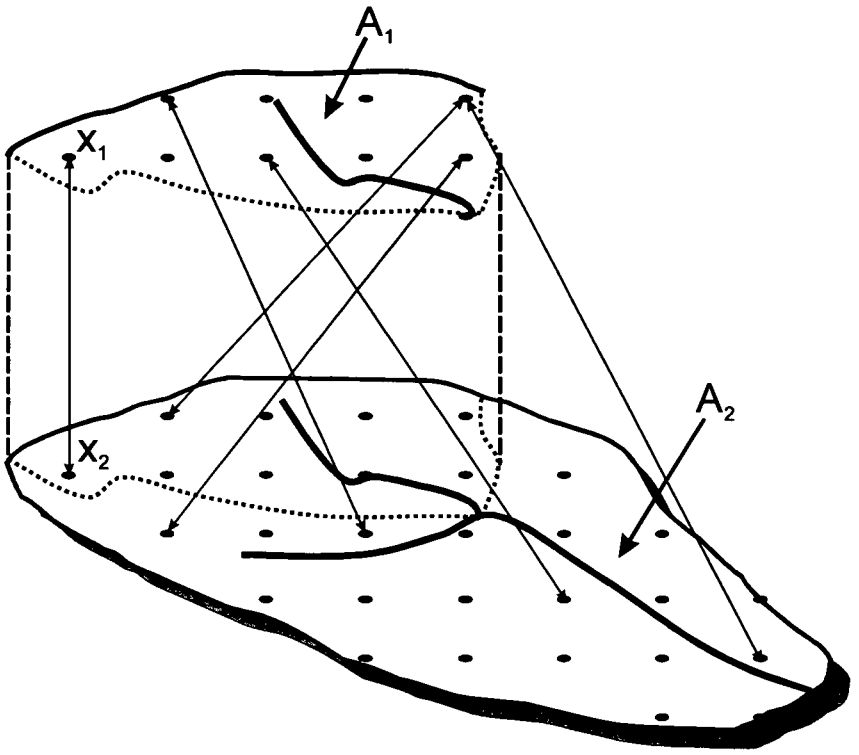


Figure 6.2 Schematic stream network and catchment boundaries with point pairs shown.

Many variables are likely to have a nugget effect that represents variability at scales much smaller than the distance between measurements. In the variogram the nugget appears as a discontinuity close to the origin. A point variogram with nugget variance C_{0p} can be expressed as:

$$\gamma_p(h) = \begin{cases} 0 & if \quad h = 0 \\ f(h) + C_{0p} & \text{otherwise} \end{cases} \tag{6.5}$$

where h is spatial distance and $f(h)$ is the variogram without nugget. If Eq. (6.5) is regularised with Eq. (6.4) directly, the nugget will vanish even for small catchments. The suggestion is therefore to regularise the nugget separately. The nugget variance can be seen as the variance of a spatially independent random variable. Following Journel and Huijbregts (1978, p. 154-156), the regularised nugget variance for two catchments of different size $C_0(A_1, A_2)$, overlapping or not, can be generalised as:

$$C_0(A_1, A_2) = 0.5\left(\frac{C_{0p}}{A_1} + \frac{C_{0p}}{A_2} - \frac{2C_{0p} \cdot \text{Meas}(A_1 \cap A_2)}{A_1 A_2}\right) \tag{6.6}$$

where Meas($A_1 \cap A_2$) represents the area shared by the two catchments with areas A_1 and A_2. If the catchments are nested this will be $\min(A_1, A_2)$, if they are not this will be zero. The regularised nugget effect is then added to the regularised variogram of Eq. (6.4).

An advantage of kriging over some other interpolation methods is that it provides an estimate of the kriging variance of the estimate at any location. The kriging variance σ_R^2 represents the uncertainty of the estimates and is given by:

$$\sigma_R^2 = \sum_{i=1}^{n} \lambda_i \gamma_{i0} + \mu \qquad (6.7)$$

γ_{i0} is the gamma value between the target catchment and the neighbouring catchments.

Figure 6.3 illustrates the merits of Top-kriging over Euclidian kriging methods such as Ordinary Kriging. In all three examples, the neighbouring catchments have the same centre-to-centre distance to the target catchment (20 km in the left panel, 10 km in the centre and right panels). In Ordinary Kriging this would imply that the same weights λ_i are assigned to all the neighbouring catchments (0.25 in the case of four neighbours, 0.5 in the case of two neighbours). In Top-Kriging the weights are different. The example on the left shows the catchment size effect. The largest catchment has the largest weight, because this is regarded as the most certain, or having the least biased measurement in comparison to the mean. Although the 49 km^2 catchment on the right is larger than the 25 km^2 catchment at the bottom, their weights are similar which is because the 49 km^2 catchment is closer to the 400 km^2 catchment. This is because kriging reduces the weights of clustered samples in contrast to, e.g., inverse distance methods (Isaaks and Srivastava, 1989, pp. 318-321).

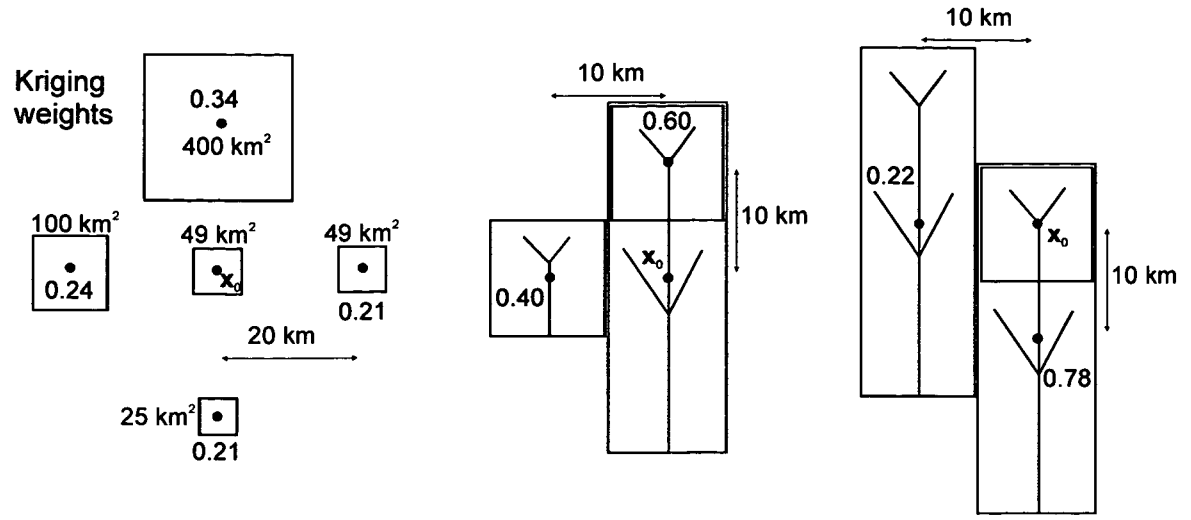


Figure 6.3 Effect of catchment size (left) and nesting (centre and right) on the kriging weights λ_i (red numbers) as estimated by Top-kriging. $\mathbf{x}_0$ indicates the centre of the target catchment.

The centre and right panels in Fig. 6.3 show the effect of nesting. The sizes of the two neighbouring catchments are the same in each case (100 km² in the centre panel, 300 km² in the right panel). The centre panel indicates that the catchment that forms a subcatchment of the target catchment gets a larger weight even though size and distance are the same as those of the other catchment. The right panel shows the reverse case. Although the neighbours have the same areas and the same centre-to-centre distances to the target catchment, more weight is attached to the catchment into which the target catchment drains.

The weights have been obtained by the variogram shown later in this paper with the measurement errors set to zero. The relative effects of size and nesting would be similar for other variograms.

6.3 Example application

6.3.1 Data

The concept of Top-kriging is applied to an Austrian data set of 7000 catchments in this paper. For all of these catchments the stream network and the catchment boundaries were available. About 600 of these catchments were gauged. For these flood data were used with record lengths ranging from 5 to 110 years, most of them from 10 to 50 years. From these data, the specific 100 year flood, Q_{100}, was estimated using a Gumbel distribution. To be able

to compare smaller and larger catchments, the Q_{100} values were normalised to a catchment size of 100 km^2, following Merz and Blöschl (2005):

$$Q_{100N} = (A \cdot \alpha^{-1})^{\beta} Q_{100} \tag{6.8}$$

where A is the catchment area, α is the catchment area used for normalisation (100 km^2) and β was set to –0.33. Although the kriging theory does not require the data to be normally distributed it is an advantage to reduce skewness. Because of this, the Q_{100N} values were logarithmically transformed before using them for interpolation:

$$z = \ln(Q_{100N}) \tag{6.9}$$

The expected variance of the estimate of a Gumbel distribution is (e.g. Plate, 1993, p. 418):

$$\sigma_{QT}^2 = \frac{s_Q^2}{n}\left[1 + 1.1396K_T + 1.100K_T^2\right] \tag{6.10}$$

where T refers to the return period of the flood, s_Q^2 is the variance of the annual flood series, n is the number of annual floods in the series and K_T is a constant dependent on the return period of the estimate:

$$K_T = -\frac{\sqrt{6}}{\pi}(\ln(-\ln(1-1/T)) + 0.5772) \tag{6.11}$$

T=100 in the present case. The variance was also logarithmically transformed by

$$\sigma_i^2 = \sigma_z^2 = \ln(1 + CV_{Q100N}^2) \tag{6.12}$$

where the coefficient of variation is expressed as $CV_{Q100N} = \sigma_{Q100N} / \mu_{Q100N}$. μ_{Q100N} is the mean of the data series. Before presentation, all values were back transformed.

The Q_{100N} values of the gauged catchments are referred to as measurements in this paper to illustrate the characteristics of Top-kriging. Their variances σ_{Q100}^2 are referred to as measurement errors to illustrate how Top-kriging estimates uncertainties in ungauged catchments.

6.3.2 Estimation of point variogram

In order to apply Top-Kriging a point variogram is needed which was back-calculated from the $\ln(Q_{100N})$ values at the gauged catchments. Kyriakidis (2004), Mockus (1998) and

Chapters 2-5 provide methods for back-calculation. As there are too many stations for using the variogram cloud for fitting, a sample variogram is in this chapter inferred with the pairs grouped into bins, similar to Matheron's (1965) traditional estimator, but with two more dimensions, the areas of the two catchments of a pair:

$$\gamma_{obs}(A_1, A_2, h) = \frac{1}{2n(A_1, A_2, h)} \sum_{i=1}^{n(A_1,A_2,h)} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \qquad (6.13)$$

where $h = |\mathbf{h}|$ is the distance between the centroids of the catchments, $n(A_1,A_2,h)$ is the number of catchment pairs with areas A_1 and A_2, and distance h between the centroids. The bins were logarithmically distributed in all three dimensions. To reduce the number of bins and to increase the number of pairs in each bin, A1 was always chosen as the smaller area of the two catchments in a pair.

Similar to Chapter 4, a point variogram with a nugget effect of the following shape was assumed:

$$\gamma_p(h) = ah^b(1 - e^{-(h/c)^d}) + C_{0p} \qquad (6.14)$$

a, b, c and d are parameters. The parameters can be interpreted as following: a is related to the sill of the variogram, c is a correlation length, while b and d define the long and short distance slope of the variogram in a log-log plot, respectively. For a given bin, the regularised gamma values and nugget variance was calculated according to Eqs. (6.4) and (6.6) using average areas and distances within each bin and a square catchment area shape as an approximation. For each bin, a weighted relative difference was calculated between the observed gamma value γ_{obs} and the regularised gamma value including nugget effect γ_{mod}, according to the Weighted Least Squares method of Cressie (1985):

$$Err(A_1, A_2, h) = n(A_1, A_2, h)\left(\frac{\gamma_{obs}(A_1, A_2, h)}{\gamma_{\text{mod}}(A_1, A_2, h)} - 1\right)^2 \qquad (6.15)$$

By minimising the sum of Err the parameters of the point variogram Eq. (6.14) was found to be a=2.99, b=0.0812, c=9690, d=0.2568, C_{0p}=1.9668.

Figure 6.4 shows the back-calculated point variogram together with a number of regularised variograms for different catchment areas, as examples. In all cases a square catchment shape was assumed. As the catchment area increases, the gamma values decrease because of the smoothing effect of regularisation. Catchments of different size will always have a variance between them, also when the centre-to-centre distance is zero. This is the reason why all

variograms between catchments of different size start with an apparent nugget effect. The effect of the point nugget effect C_{0p} is dependent on the catchment size and degree of overlapping. There may be some uncertainty with estimating the point variogram as different point variograms can give similar regularised variograms. However, Top-Kriging is not very sensitive to this as regularised variograms are used for catchment sizes that are of the same order of magnitude as those on which the observed variogram is based (Chapter 5):

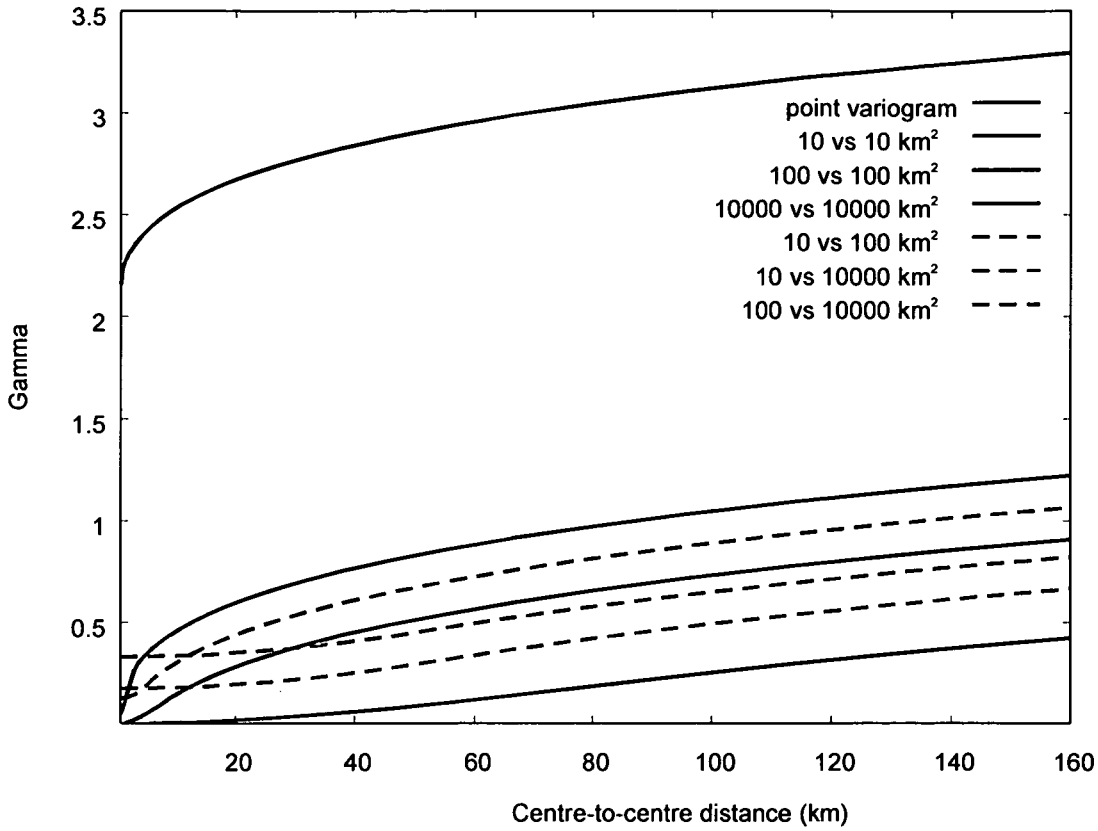


Figure 6.4 Point variogram and regularised variograms of different catchment sizes.

Figure 6.5 shows a comparison of the observed gamma values γ_{obs} and the gamma values γ_{mod} regularised from the point variogram. For the presentation, the observed gamma values were grouped and the mean and standard deviations of the modelled gamma values for each group are shown. The model has a tendency of overestimating the gamma values for small observed gamma values and the standard deviations of the modelled gamma values are relatively large. This is partly because the observed variogram Eq. (6.13) is more complex than the regularised version of Eq. (6.14) and partly because in some of the bins there are only

few pairs. For the purpose of demonstrating the characteristics of Top-kriging, however, the fit was considered acceptable.

To assess the merit of Top-kriging relative to Ordinary Kriging the Q_{100N} values were also estimated by Ordinary Kriging using a variogram of the type of Eq. (6.14) with the parameters a=0.58, b=0.0010, c=29.8, d=0.667, C_{0p}=0.0236

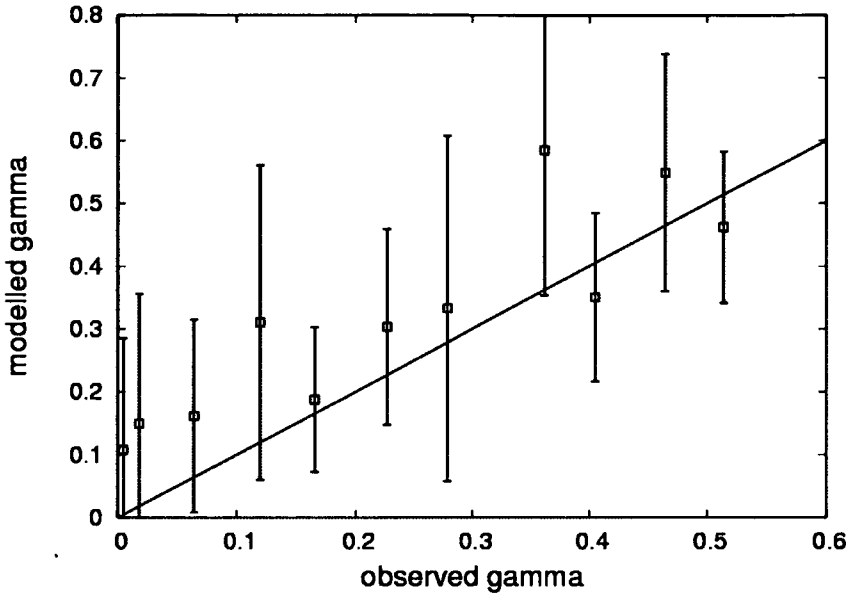


Figure 6.5 Comparison between observed semivariances and mean and standard deviation (error bars) of modelled semivariances. Green line shows $\gamma_{obs} = \gamma_{\mathrm{mod}}$.

6.3.3 Results

To understand and illustrate the characteristics of Top-kriging this paper focuses on two sub-regions of the entire data set. The estimates for individual catchments will be discussed in a regional context for both regions. The first region is the river Mur and tributaries in Styria, southern Austria. In this region the focus is on the estimates on the tributaries. The second region is the river Inn and tributaries in Tyrol, western Austria, where the focus is on the main stream. In both examples, the Top-kriging estimates are compared with Ordinary Kriging estimates that use the Euclidian distance between the catchment centres to estimate the gamma values in the Kriging system.

Figure 6.6 shows the catchment areas of the Mur and tributaries as well as the stream gauges. Stream gauges 1-3 are situated directly on the river Mur, the other gauges on the tributaries.

The catchment area of the Mur almost doubles between stations 1 and 3, from 2300 km^2 to 4400 km^2. There are four tributaries entering the river on this reach with catchment areas ranging from 200-500 km^2.

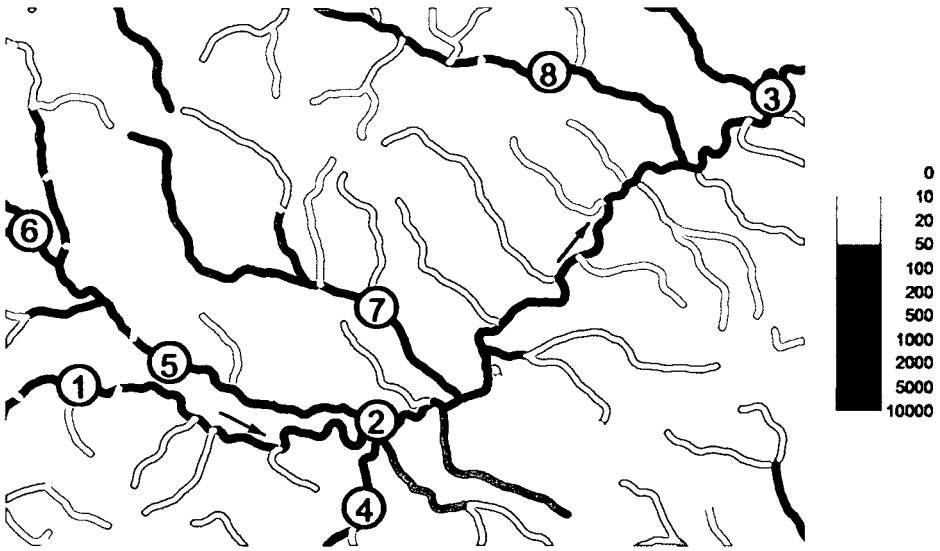


Figure 6.6 Catchment area of the Mur and tributaries (km²). Arrows show flow direction. Numbers refer to the stream gauges. Gauges 1-3 are situated directly on the Mur, the other gauges on the tributaries. The region represents a rectangle of 52 by 35 km.

Figure 6.7 presents the estimates of the normalised specific 100 year flood Q_{100N} from Top-kriging (top) and Ordinary Kriging (bottom) colour coded on the stream network. The measurements are shown as circles in both figures with the same colour coding. For both methods, the estimates next to the stream gauges are almost equal to the measurements of the stream gauge itself.

The Top-kriging estimates on the main river are similar to the measurements on the main river (gauges 1-3) and they do not change much along the reach. The estimates on the northern tributaries are much smaller than those on the main stream which is consistent with the measurements on the same tributaries (gauges 5 and 7). This is also reflected in the estimates for the other northern tributaries. On the southern side, the measurements are larger, so the Top-kriging estimates are generally much larger than those on the northern side of the Mur.

The Ordinary Kriging estimates differ substantially from the Top-kriging estimates. The main difference is that the estimates are not similar along the stream network as is the case of Top-kriging but similar along Euclidian distance in space. Although gauge 7 has measurements of

0.4 (red colour), most of the Ordinary Kriging estimates along this tributary are around 0.6 (yellow to green colours). This is because the estimates along this tributary are too much influenced by the measurements along the main river while they should be mainly influenced by the downstream gauge as is the case in Top-kriging. On the other hand, the estimates on the main stream are somewhat underestimated by Ordinary Kriging as they are too much affected by the measurements on the tributaries.

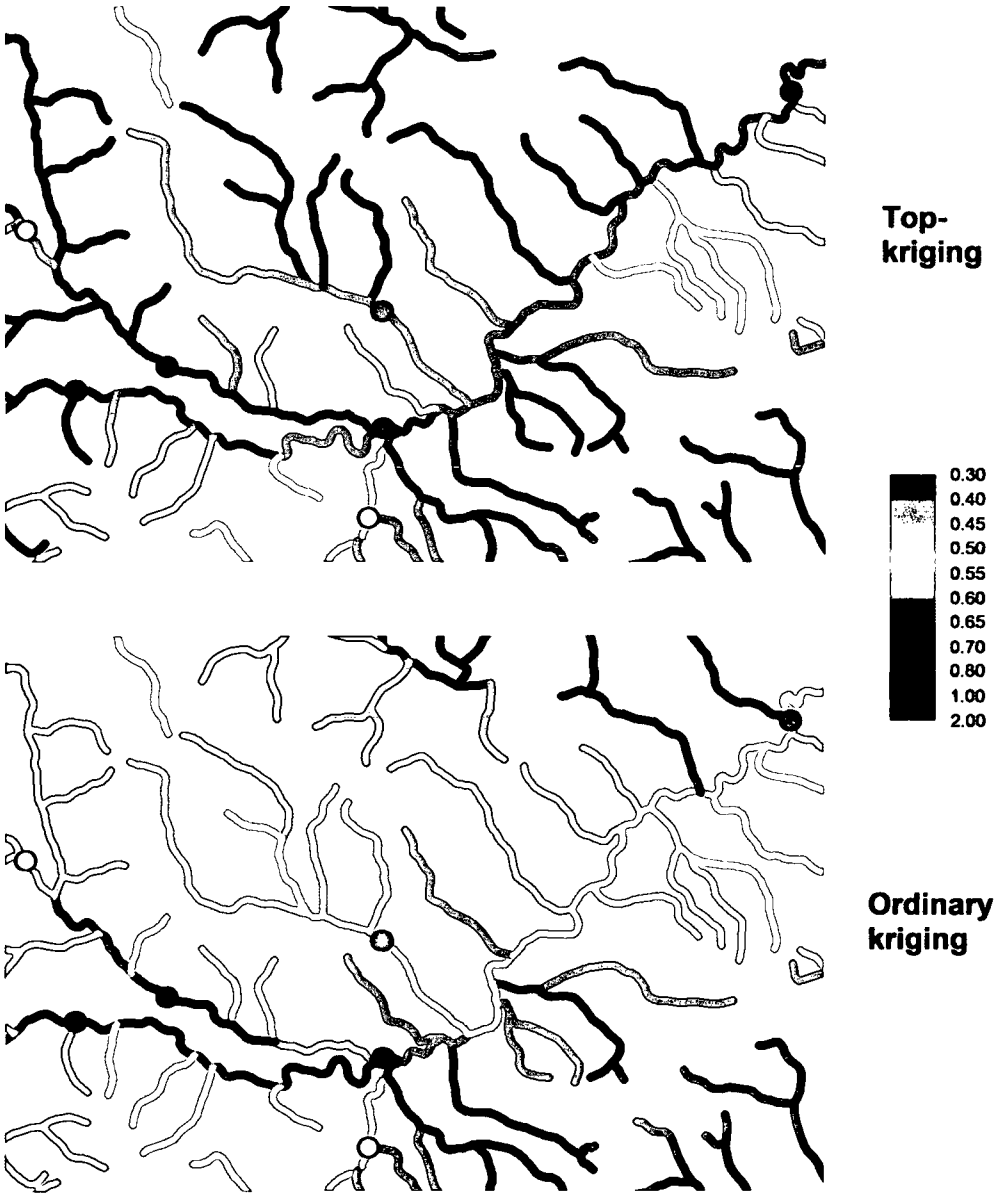


Figure 6.7 Estimates of the normalised specific 100 year flood Q_{100N} from Top-kriging (top) and Ordinary Kriging (bottom) colour coded on the stream network of the Mur region. The measurements (i.e. values at the stream gauges) are shown as circles. Units are in $m^3/s/km^2$.

In order to examine the merits of Top-kriging more quantitatively, a cross validation procedure was carried out for both methods. A measurement was temporarily discarded from the sample data set and the Q_{100N} was then estimated by both Top-kriging and Ordinary Kriging for the same location from the remaining samples (Journel and Huijbregts, 1978, p. 352). The difference between the estimate so obtained and the measurement is a measure of the interpolation error. Figure 6.8 shows these differences for the two methods. The figure indicates that the estimates from Top-kriging are similar or better than the estimates from Ordinary Kriging in all cases but one (gauge 6). The Ordinary Kriging error for gauge 6 is smaller because it assigns more weights to gauges left of the region shown that possess larger Q_{100N} values. It is also worth noting that the errors are generally small for estimates on the main river (gauges 1-3), while they can be larger for the tributaries.

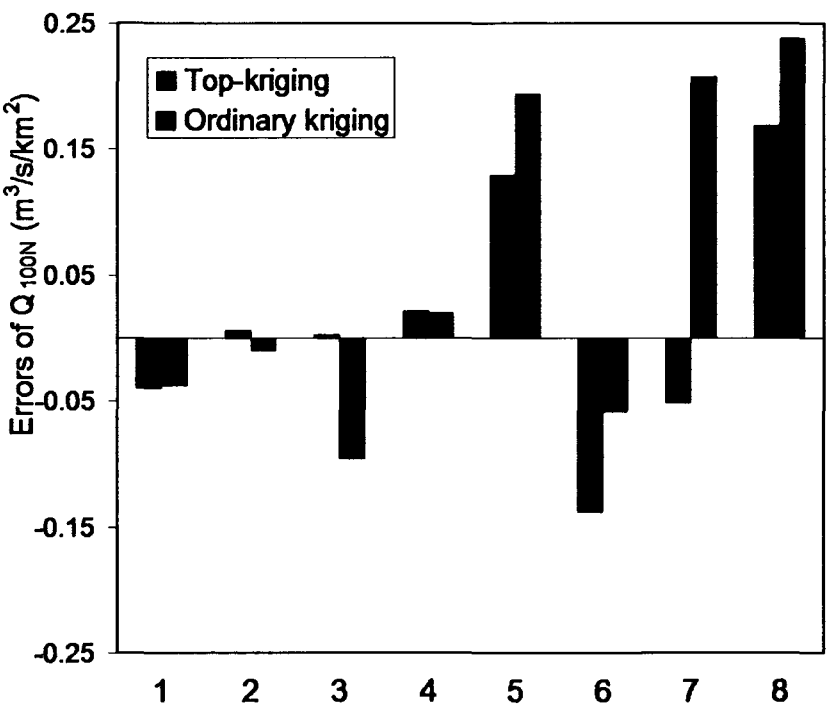


Figure 6.8 Cross validation errors of Q_{100N} for the eight gauges in the Mur region. Station numbers as shown in Fig. 6.7.

The kriging variances of the estimates in the Mur region, Eq. (6.7), are shown in Fig. 6.9 colour coded on the stream network. The kriging variances are expressed as the coefficient of variation (CV) of the estimate:

$$CV = \frac{\sigma_R}{Q_{100N}} \tag{6.16}$$

The points represent the measurement error σ_i^2 at the stream gauges, again expressed as a coefficient of variation. The measurements have CVs in the range 0.05-0.3. This is rather large, and is related to relatively short record lengths for some of the gauges. Both procedures estimate the lowest uncertainties close to the measurements, equal to or larger than the CV of the measurements. Note that the uncertainties on the stream network have been plotted for stream reaches of finite lengths as for these the catchment boundaries were available. The small uncertainties, strictly speaking, apply to the immediate neighbourhood of the gauges.

The uncertainties estimated by Top-kriging and Ordinary Kriging are very different for most of the stream network. Top-kriging (Fig. 6.9 top) gives relatively small uncertainties on the main river with CVs of around 0.2. This is only slightly larger than the CV of the measurements. On the other hand, the uncertainties of some of the tributaries are considerably larger. The uncertainties are small for those tributaries where measurements are available, but rather large for tributaries without any measurements. It is interesting that the uncertainty increases substantially with decreasing catchment area. For some of the smallest catchments, i.e. headwater catchments, CVs of more than 1 are estimated. These point to very uncertain estimates, which is not surprising as no measurements are near.

The uncertainties estimated by Ordinary Kriging (Fig. 6.9 bottom) contradict what one would intuitively expect. Most disturbing is that some of the smallest catchments have uncertainties equal to or smaller than the uncertainty of the main river. This is of course a result of the uncertainty being a function of Euclidian distance between catchment centre and measurements only, and not a function of the size and nesting of the catchments.

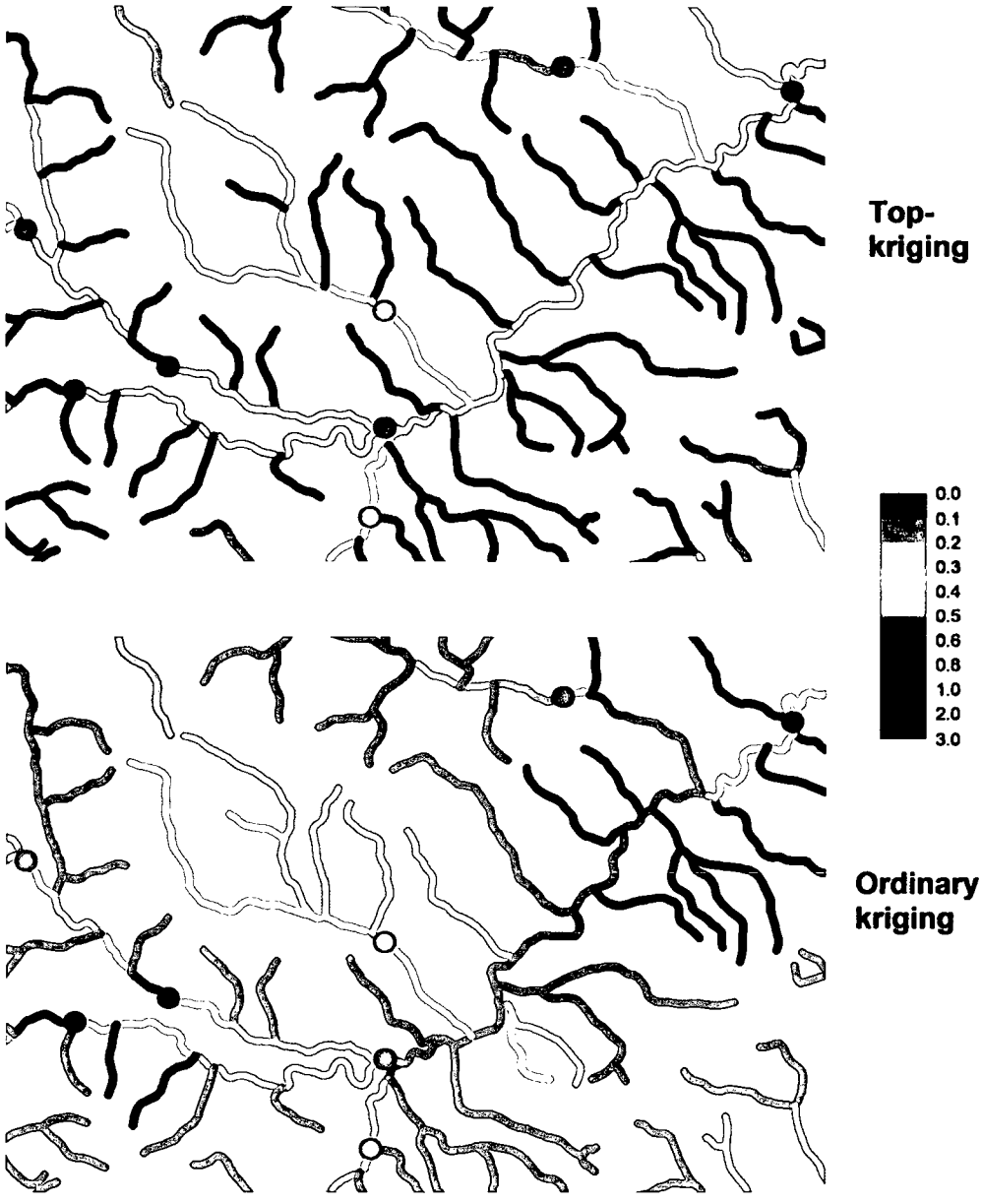


Figure 6.9 Uncertainties of the normalised specific 100 year flood Q_{100N} from Top-kriging (top) and Ordinary Kriging (bottom), expressed as the coefficient of variation, colour coded on the stream network of the Mur region. Uncertainties of the measurements (i.e. values at the stream gauges) are shown as circles.

The second example presented in this paper is the Inn region for which the catchment areas are shown in Fig. 6.10. Stream gauges 1-3 are situated directly on the river Inn, the other gauges on the tributaries. Between gauges 1 and 3, the catchment area increases from 5800 km^2 to 8500 km^2. There are only two tributaries with an area of more than 100 km^2 entering the Inn within this reach, but they are larger than the tributaries of the Mur. The Sill is the

tributary gauged by gauge 4 just before the junction with the Inn, and has an area of 850 km^2. The Ziller is the tributary gauged by gauges 7-10, and has an area of 1135 km^2.

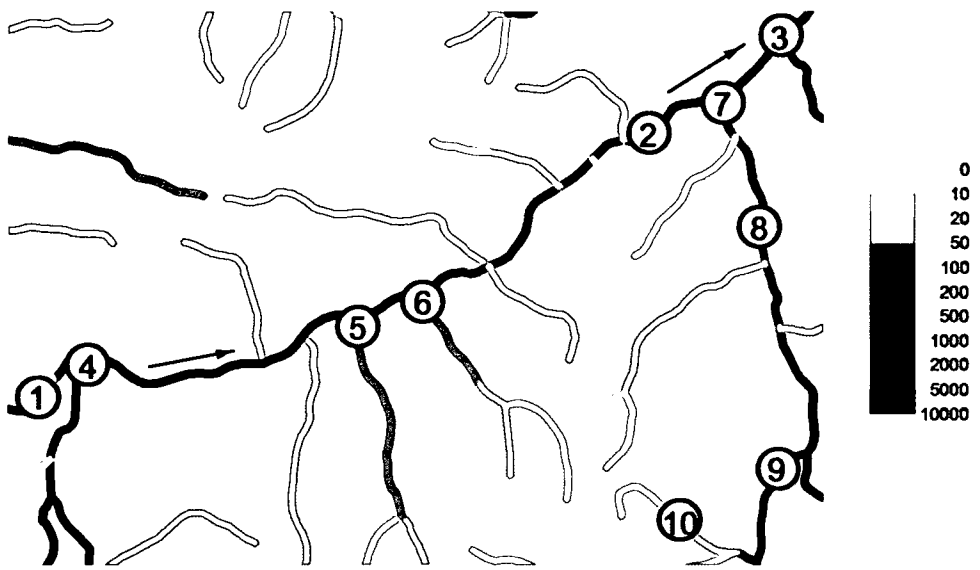


Figure 6.10 Catchment area of the Inn and tributaries (km²). Arrows show flow direction. Numbers refer to the stream gauges. Gauges 1-3 are situated directly on the Inn, the other gauges on the tributaries. The region represents a rectangle of 43 by 29 km.

The Top-kriging and Ordinary Kriging estimates of the normalised specific 100 year flood Q_{100N} are shown in Fig. 6.11 top and bottom, respectively. The Q_{100N} measured at the Sill (gauge 4) is similar to that measured at the Inn at gauge 1. The Q_{100N} measured at the Ziller (gauge 7) is larger than that measured at the Inn at gauge 1 but the Q_{100N} at the smaller tributaries (gauges 5 and 6) are much smaller than those of the Inn. Although the measurements on the tributaries show large variations, the Top-kriging estimates on the Inn change very little, which is consistent with the measurements.

Similar to the Mur case, the Ordinary Kriging estimates deviate considerably from Top-kriging but they deviate in a different way because of different gauge locations. For the Mur case, the main differences were the estimates of the tributaries, which were too much affected by the main river. In the Inn case (Fig. 6.11), the estimates on the tributaries are similar to those of Top-kriging, but the estimates on the main river are obviously too much influenced by measurements on tributaries. The tributaries for which the centres are close to the centre of

the Inn lie outside the region shown in Fig. 6.11 and have small measured Q_{100N} similar to the southern tributaries in Fig. 6.11.

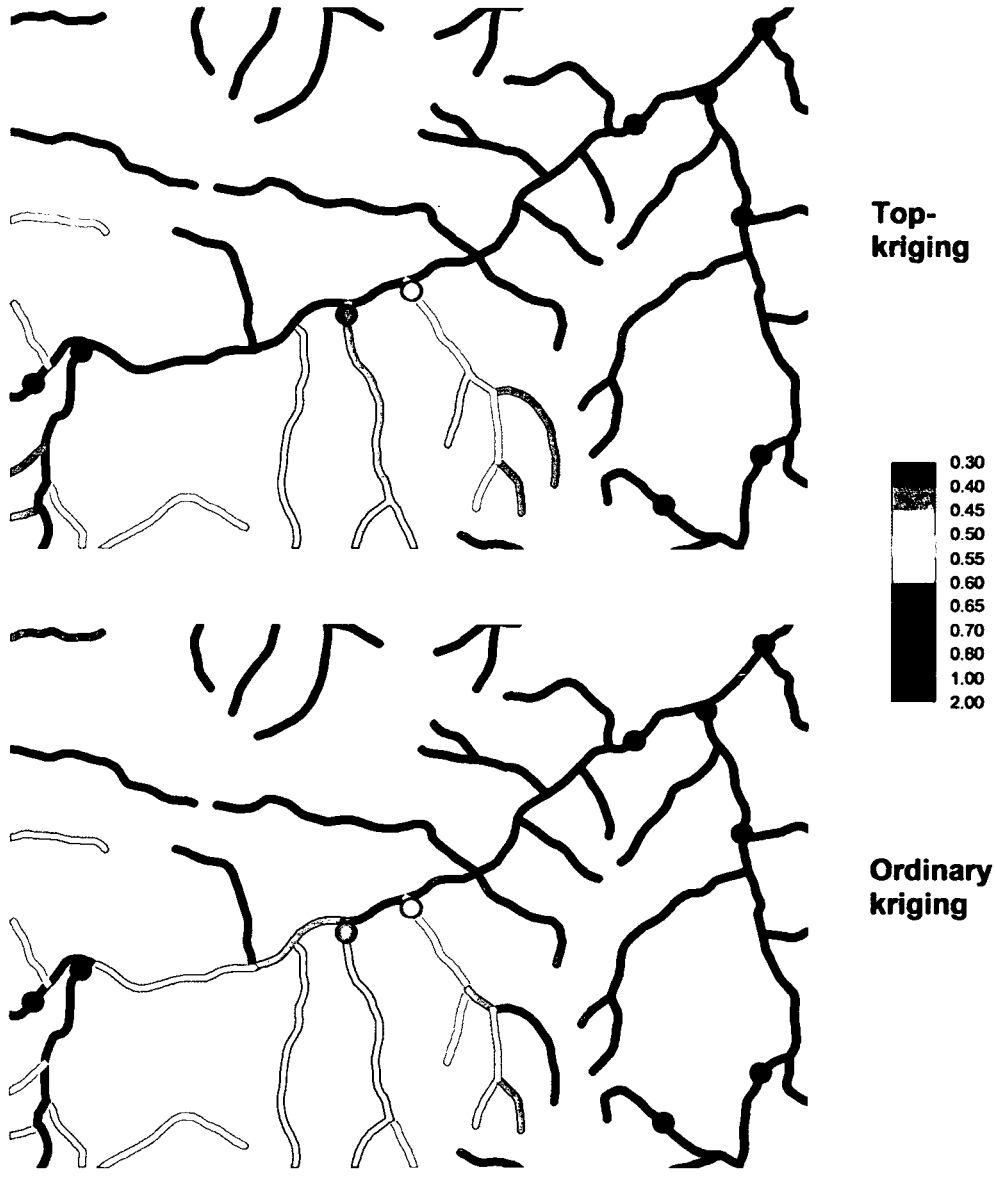


Figure 6.11 Estimates of the normalised specific 100 year flood Q_{100N} from Top-kriging (top) and Ordinary Kriging (bottom) colour coded on the stream network of the Inn region. The measurements (i.e. values at the stream gauges) are shown as circles. Units are in $m^3/s/km^2$.

The cross validation procedure confirms more quantitatively that the interpolation errors of Top-kriging are much smaller than those of Ordinary Kriging (Fig. 6.12). The difference between the two methods is largest along the main river (gauges 1-3). This is because estimates from Ordinary Kriging have their largest errors along the main river, while Top-

kriging only gives small errors on the main river. The difference is smaller along the tributaries, but Top-kriging does give smaller errors than Ordinary Kriging for the majority of the gauges.

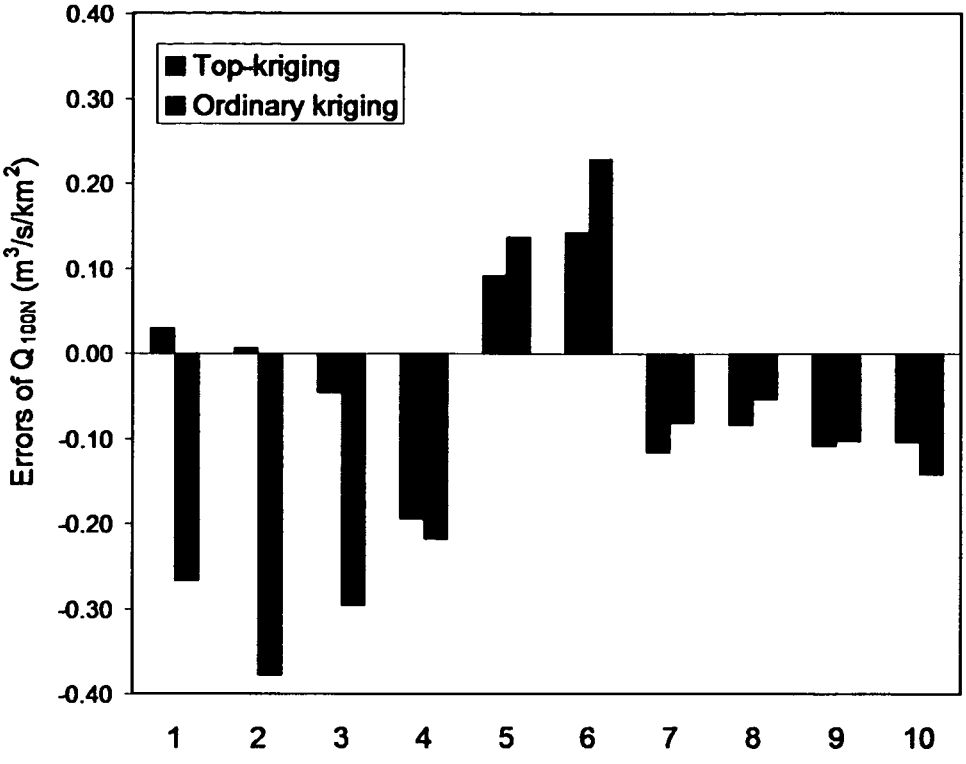


Figure 6.12 Cross validation errors of Q_{100N} for the eight gauges in the Inn region. Station numbers as shown in Fig. 6.7.

The estimated uncertainties of the estimates in the Inn region are shown in Fig. 6.13. Similar to Fig. 6.9, Fig. 6.13 shows that the uncertainties next to the stations are equal to the measurement uncertainties. The Top-kriging uncertainties along the main river and along tributaries with measurements are lower than the uncertainties of tributaries without measurements as would be expected. The uncertainties generally increase with decreasing catchment size. For the tributaries that are gauged close to the confluence with the main river (gauges 5 and 6) the uncertainty gradually increases as one moves away from the gauge towards the headwaters. However, if the tributary is gauged (e.g. by gauges 5 and 6) the estimates of the headwater catchments are less uncertain than the headwaters of ungauged tributaries (e.g. the tributaries close to stream gauge 2). Overall, the Ordinary Kriging uncertainties indicate that the Euclidian distances do not reflect the intuitive distribution of estimation errors. The uncertainties are far too uniform within the region as they do not take

into account the amount of information shared by gauged and ungauged catchments. In contrast, Top-kriging captures exactly this information, as expected.

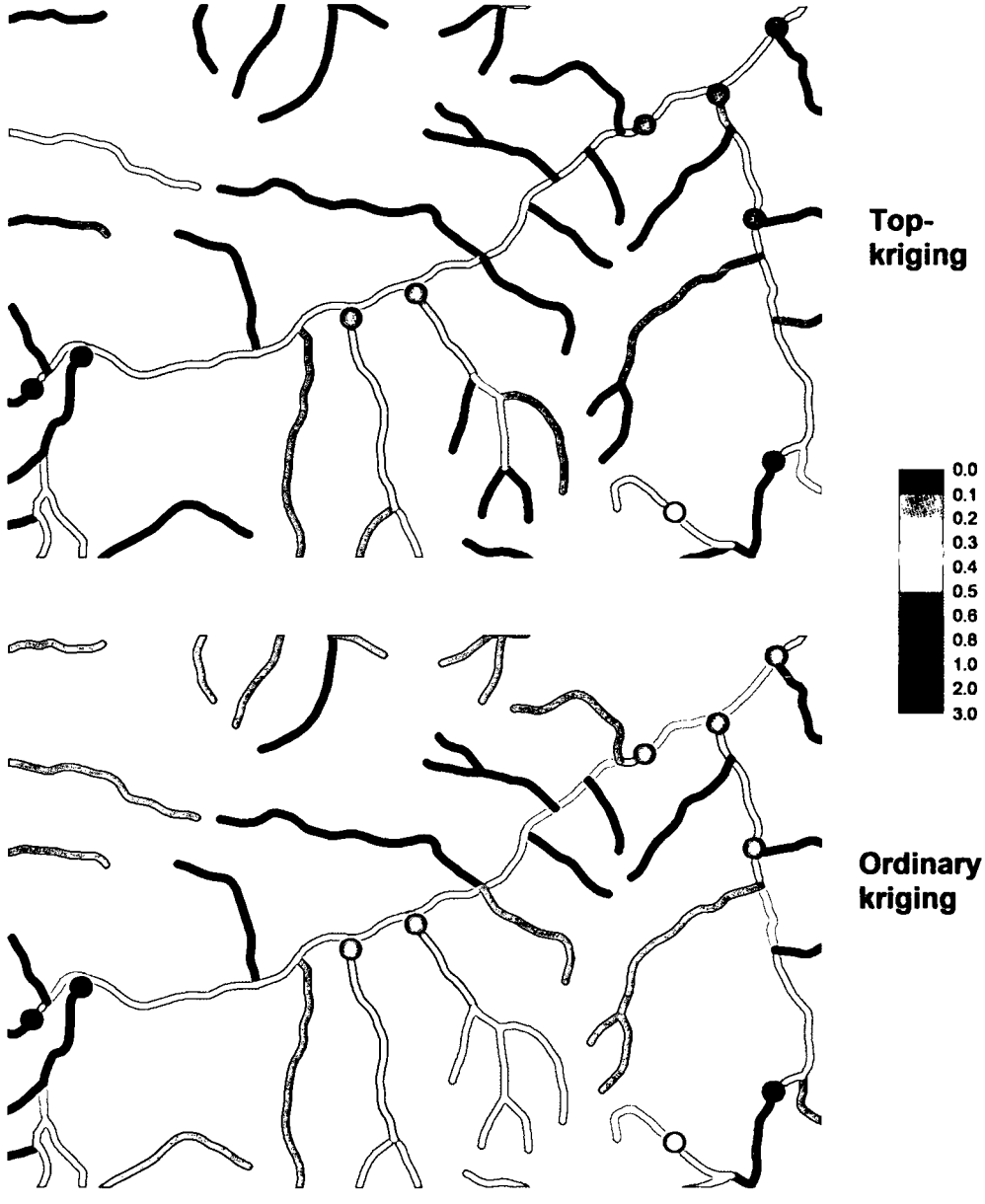


Figure 6.13 Uncertainties of the normalised specific 100 year flood Q_{100N} from Top-kriging (top) and Ordinary Kriging (bottom), expressed as the coefficient of variation, colour coded on the stream network of the Inn region. Uncertainties of the measurements (i.e. values at the stream gauges) are shown as circles.

6.4 Conclusions

This chapter presented Top-kriging as a spatial estimation method for streamflow-related variables. It takes both the area and the nested nature of catchments into account. The main appeal of the method is that it is a best linear unbiased estimator (BLUE) adapted for the case of stream networks without any additional assumptions. Because of the minimum number of assumptions, it is here suggested that the Top-kriging approach is the most natural method of estimating streamflow-related variables on stream networks.

The method provides more plausible and, indeed, more accurate estimates of the specific 100 year flood than Ordinary Kriging in the regions examined here. In the example of the Mur region it was showed how the estimates of the tributaries are improved over Ordinary Kriging. In the example of the Inn region it was showed how the estimates on the main stream are improved. In general, both the estimates on the main stream and the tributaries will be superior to Ordinary Kriging and the relative magnitude of the improvement will depend on the location of the gauges. In essence, Top-kriging treats upstream and downstream catchments differently from neighbouring catchments that do not share a subcatchment. This is also what one would do in manual interpolation. Typically, the effect of nesting will differ for upstream and downstream catchments. Larger catchments are generally given larger weights than smaller catchments (Fig. 6.3) and they are also likely to be well correlated with other large neighbours. It is therefore more likely that the nesting will have more impact on the weights of an upstream catchment than on those of a downstream catchment.

In addition to accounting for nested catchments, the method as presented here is able to exploit the information contained in short records (Eqs. 6.2 and 6.10). Variables estimated from short records are less certain than those from longer records. By allowing the measurement error to vary between gauges it was possible to use both short and long records and attribute more confidence to the longer records.

Top-kriging not only provides estimates of the variable of interest in ungauged catchments but also provides estimates of their uncertainty. The uncertainty patterns estimated by Top-kriging are very different from those of Ordinary Kriging. While the Ordinary Kriging uncertainty only depends on the centroid distances of gauged and ungauged catchments, the Top-kriging uncertainties fully take into account the nested nature of catchments. Figures 6.9 and 6.13 are typical of the way Top-kriging estimates interpolation uncertainty. On the main stream, where most of the gauges are, the estimated uncertainties are smallest. On tributaries that are gauged close to their confluence with the main river, the uncertainty gradually

increases as one moves away from the gauge towards the headwaters. If a tributary is gauged, the estimates of the headwater catchments are less uncertain than the estimates of headwater catchments of similar size without a downstream gauge on the tributary. Locations between gauges at large rivers can also be considered as ungauged. However, it is in the headwater catchments where most of the uncertainty resides. The IAHS Decade on Predictions in Ungauged Basins (PUB) (Sivapalan et al., 2003) has predictive uncertainty of hydrological variables as its main focus. This is exactly what Top-kriging provides for the most natural case of best linear unbiased estimators.

Top-kriging assumes linear aggregation as it is a linear estimator. This means that, strictly speaking, the method only applies to variables that are mass conserving over nested catchments. It is still suggested that the method can also be profitably used, as an approximation, for variables that do not aggregate linearly but show a degree of averaging. The example of the specific 100 year flood shown here is not mass conserving and it was demonstrated that the Top-kriging estimates are much better than the Ordinary Kriging estimates. In fact, although Top-kriging is based on linear aggregation it does not necessarily reproduce the mass-balance of the variable of interest (Sauquet et al. 2000). This is consistent with the here suggested approximate use of Top-kriging for a range of streamflow-related variables. Such variables of interest in hydrology include mean annual discharge, flood characteristics, low flow characteristics, concentrations, turbidity and stream temperature.

Top-kriging as presented in this paper is based on a simple linear aggregation scheme. There are numerous opportunities for extending this simple scheme and still retaining the merits of a regionalisation procedure that naturally takes into account the area and nested structure of catchments. There are more complex ways of representing the logarithmic back-transformation of the variable of interest (e.g. Clark, 1998, Krige, 2004) and these could be used in a Top-kriging context. In a different context Top-kriging was extended to a more complex scheme that takes auxiliary variables into account (such as mean annual precipitation and a lake index) to improve the estimates beyond simple interpolation. This is being used to estimate T-year floods for 26000 km of Austrian streams, which will be used for hazard zone planning in a project known as HORA. In Chapter 7, the Top-kriging approach will be extended to space-time aggregation.

Chapter 7. Spatio-temporal Top-kriging of runoff time series

7.1 Introduction

The estimation of runoff related variables at locations where no measurements are available is a key problem in hydrology which is generally termed the Problem of Ungauged Basins (PUB) (Sivapalan et al., 2003). A range of methods addressing this problem has been proposed in the literature (Blöschl, 2005). The traditional approach is to use deterministic rainfall-runoff model with parameters inferred from neighbouring, gauged catchments. The advantage of this method is that it explicitly represents causal processes such as precipitation and runoff generation. However, as Blöschl (2005) noted, there is significant uncertainty in the estimates, mainly due to lack of representativeness of catchment data, rainfall data as well as identifiability problems of the runoff model parameters (Montanari, 2005b). For some applications it may not be necessary to invoke causal relationships directly. For example, when assessing the hydropower potential one is interested in the flow duration curves of ungauged sites (e.g. Castellarin et al., 2004) and there is usually no need to vary rainfall characteristics. Another application are environmental flows where one is interested in the runoff dynamics and seasonality at ungauged sites without the need to examine scenarios (Gippel, 2005). For these applications, an alternative to the traditional rainfall-runoff modelling approach may be to directly estimate the runoff time series from observed time series of neighbouring catchments without recourse to rainfall data. This would be of particular interest in those countries where a rather dense stream gauge network exists.

An obvious choice of estimating runoff time series directly from observed runoff of neighboring catchments are geostatistical methods which exploit spatial correlations. Although geostatistical methods such as kriging are best linear unbiased estimators (BLUE) (Journel and Huijbregts, 1978, p. 304) they have not been used much in catchment hydrology. This is because geostatistical methods have evolved in the mining industry where a typical objective is to estimate the expected ore grade of a cuboid block from point samples. The estimation procedure makes use of the variogram, which is the spatial correlation of pairs of points of the variable of interest plotted against their Euclidian distance. The problem in catchment hydrology is quite different in that, unlike mining blocks, catchments are organized

into subcatchment and the organization is defined by the stream network. Upstream and downstream catchments would have to be treated differently from neighboring catchments that do not share a subcatchment. The Euclidian distance is hence not the natural way of measuring the spatial distance of catchments. Rather a topology needs to be used that is based on the stream network.

Gottschalk (1993) was probably the first to develop a method for calculating covariance along a stream network. This method was extended by Sauquet et al. (2000) to map annual runoff along the stream network using water balance constraints in the estimation procedure and by Gottschalk et al. (2006) to map the coefficient of variation of runoff along the stream network. This method was extended in Chapter 6 where it was used to the 100 year floods at ungauged locations as well as their uncertainty. The concept was based on the assumption by Woods and Sivapalan (1999) that assumes that local runoff or rainfall excess can be defined at each point in space and can be integrated over a catchment to form catchment runoff. The method in Chapter 6 was termed topological kriging or Top-kriging as it exploits the topology of nested catchments in addition to the spatial correlation of runoff. This chapter is concerned with estimating time series of runoff for ungauged locations rather than with estimating a single quantity as was the case in Sauquet et al. (2000) and in Chapter 6. It is hence necessary to extend the spatial estimation procedure of the earlier work to a space-time estimation procedure, taking into account both spatial and temporal correlations of runoff.

Spatio-temporal geostatistical models have been used in a number of disciplines (e.g. Snepvangers et al., 2003, Jost et al., 2005). In a review of spatio-temporal methods, Kyriakidis and Journel (1999) note that there are three options of representing the random variable –full space-time models, simplified representations as vectors of temporally correlated spatial random fields, and simplified representations as vectors of spatially correlated time series. The latter reduces to a spatial estimation problem and is of interest for variables that are rich in time but poor in space as is the case of runoff (Rouhani and Wackernagel, 1990). Full spatio-temporal kriging is more complicated than the two simplifications as the kriging system needs to be solved simultaneously for both spatial and temporal kriging weights (Kyriakidis and Journel, 1999). As still another alternative, spatio-temporal cokriging has been suggested (Rouhani and Wackernagel, 1990), where information from different time steps are treated as covariates. Spatio-temporal cokriging includes more unbiasedness conditions, while spatio-temporal kriging is strictly only valid when the mean does not change with time (Bogaert, 1996)

The objective of this chapter is to propose a method of spatio-temporal Top-kriging that is able to estimate runoff time series at ungauged locations. The chapter goes beyond the work of Sauquet et al. (2000) and what was presented in Chapter 6 by accounting for space-time correlations rather than spatial correlations and by including routing effects. The new method is tested by jack-knifing to assess its accuracy for the case of ungauged locations based on an Austria data set. The characteristics of the method in terms of representing the dynamics of the hydrograph are examined and its predictive power is compared to that of traditional regionalisation based on deterministic rainfall runoff models.

7.2 Data

The data used in this chapter stem from a comprehensive hydrographic data set of Austria. Austria has a varied climate with mean annual precipitation ranging from 500 mm in the eastern lowland region up to about 3000 mm in the western alpine region. Runoff depths range from less than 50 mm per year in the eastern part of the country to about 2000 mm per year in the Alps. Potential evapotranspiration is on the order of 600-900 mm per year. Austria has a dense stream gauge network. Hourly runoff data over the period Aug 1, 1990 to July 31, 2000 are used in this chapter. The raw runoff data were screened to exclude catchments with significant anthropogenic effects, karst and strong lake effects. The remaining data set consisted of 376 stream gauges with catchment areas ranging from 10 to 10000 km² (Fig. 7.1).

To analyze the dynamic characteristics of the estimation method, the Innviertel region in northwestern Austria is examined in more detail (Fig. 7.2). The region covers an area of approximately 1500 km² and contains 19 stream gauges. Mean annual runoff is rather uniform in the region, ranging from 350 to 510 mm/year. However, there are differences in the runoff dynamics (Table 7.1). This is because of a rather complex hydrogeology consisting of a mixture of moraine, clay and marl with local gravel fillings of the valleys (BMLFUW, 2003, Map 6.2). The temporal variance of runoff in the most dynamic catchment (Osternach) is more than ten times that of the slowest responding catchment (Weghof). The latter contains significant gravel deposits in the valleys while the former does not.

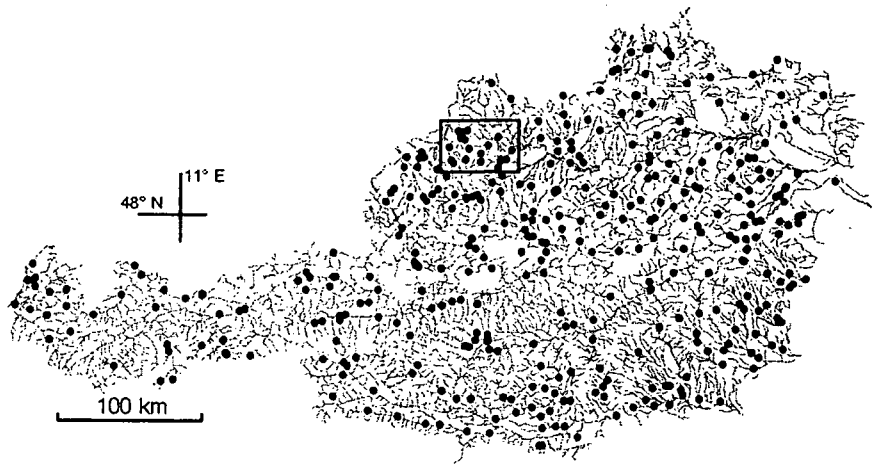


Figure 7.1 Stream gauges (circles) in Austria used in this chapter. The square represents the Innviertel region.

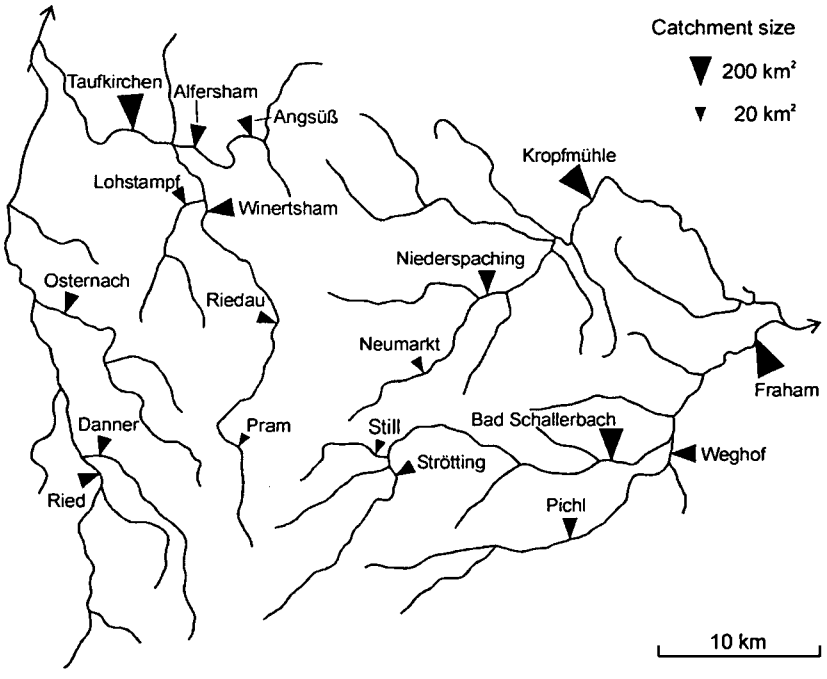


Figure 7.2 Innviertel region with stream gauges (triangles). The triangles are scaled according to catchment area. The catchments are listed in Table 7.1.

Table 7.1 Stream gauges in the Innviertel region with catchment area, mean annual runoff and temporal runoff variance.

Stream gauge	River	Area (km^2)	Average runoff ($*10^{-2} m^3 s^{-1} km^{-2}$)	Temporal variance ($*10^{-4} m^6 s^{-2} km^{-4}$)
Ried	Rieder Bach	69.3	1.53	7.79
Danner	Antiesen	55.6	1.45	4.14
Osternach	Osternach	68.6	1.42	10.97
Pram	Pram	14.2	1.60	9.36
Riedau	Pram	59.5	1.44	7.52
Winertsham	Pram	128.1	1.33	6.48
Taufkirchen	Pram	303.3	1.38	4.65
Angsüß	Pfudabach	64.1	1.59	3.80
Alfersham	Pfudabach	81.3	1.50	2.38
Lohstampf	Messenbach	39.3	1.31	7.04
Still	Stillbach	19.4	1.33	6.32
Strötting	Trattnach	52.0	1.44	3.06
Bad Schallerbach	Trattnach	183.8	1.23	3.19
Pichl	Innbach	66.2	1.23	1.32
Weghof	Innbach	116.5	1.09	0.79
Fraham	Innbach	361.8	1.14	1.62
Neumarkt	Dürre Aschach	29.7	1.22	6.78
Niederspaching	Aschach	104.0	1.30	7.17
Kropfmühle	Aschach	312.5	1.37	4.98
Average		112.1	1.36	5.23
Median		68.6	1.37	4.98

7.3 Method

7.3.1 Spatio-temporal estimation of time series

Specific runoff is assumed to be $q(\mathbf{x}_i, t_\omega)$ at location $\mathbf{x}_i$ on the stream network and time t_ω can be represented as a spatio-temporal random field that is related to runoff Q by

$$q(\mathbf{x}_i, t_\omega) = Q(\mathbf{x}_i, t_\omega) / A_i \tag{7.1}$$

where A_i is the catchment area. In this chapter there are three variants of Top-kriging suggested, which are termed, for simplicity in terminology, spatial kriging, spatio-temporal kriging and spatio-temporal cokriging.

7.3.1.1 Spatial kriging

The first variant takes into account that the runoff time series represent spatio-temporal runoff with a much higher resolution in time than in space. As a simplification runoff is hence considered to consist of a set of spatially correlated time series (Kyriakidis and Journel, 1999). For each time step t_ω, specific runoff $\hat{q}(\mathbf{x}_i, t_\omega)$ of an (ungauged) target catchment defined by location $\mathbf{x}_i$ is estimated from observed specific runoff $q(\mathbf{x}_j, t_\alpha)$ at the same point in time $t_\alpha = t_\omega$ of neighbouring (gauged) catchments located at $\mathbf{x}_j$ as

$$\hat{q}(\mathbf{x}_i, t_\omega) = \sum_{j=1}^{n} \lambda_j q(\mathbf{x}_j, t_\alpha) \tag{7.2}$$

where λ_j is the weight given to the runoff from each gauged catchment and n is the total number of stream gauges used. This means that each time step is treated independently in the estimation procedure. The weights λ_j are found by solving the kriging system:

$$\begin{aligned} \sum_{k=1}^{n} \lambda_k \gamma_{jk} - \lambda_j \sigma_j^2 + \mu &= \gamma_{ij} \qquad j = 1, \ldots, n \\ \sum_{j=1}^{n} \lambda_j &= 1 \end{aligned} \tag{7.3}$$

where the gamma values γ_{ij} and γ_{jk} are the expected semivariances (or variogram values) between the target catchment i and the neighbours j used for estimation, and between two neighbours j and k, respectively. The gamma values are obtained by regularising a theoretical point variogram (see section 7.3.2). σ_j^2 represents the local uncertainty of runoff which may be due to measurement error and small scale variability. The use of local uncertainty in the kriging equations is termed kriging with uncertain data (KUD) (de Marsily, 1986 p. 300, Merz and Blöschl, 2005). μ is the Lagrange parameter. The number of neighbours n is limited to five to minimise potential numerical problems with the regularisation procedure. Also, in some cases the kriging weights were adjusted as described in section 7.3.1.4 before they were used in Eq. 7.2.

It is assumed that the same variogram is applicable to all time steps and that the same stream gauges j can be used as neighbours for all time steps. The kriging equation has hence only to be solved once for each target catchment i, and the same weights are used for all time steps. As noted above, only runoff data for the time step of interest are used in the estimation.

7.3.1.2 Spatio-temporal kriging

Spatio-temporal kriging on the other hand does take into account information from different time steps (Kyriakidis and Journel, 1999):

$$\hat{q}(\mathbf{x}_i, t_\omega) = \sum_{j=1}^{n} \sum_{\alpha=1}^{p} \lambda_{j\alpha} q(\mathbf{x}_j, t_\alpha) \tag{7.4}$$

where p is the number of time steps used for the estimation. Extending the kriging system of Eq. 7.3 gives

$$\begin{aligned} &\sum_{k=1}^{n} \sum_{\beta=1}^{p} \lambda_{j\beta} \gamma_{jk\alpha\beta} - \lambda_{j\alpha} \sigma_j^2 + \mu = \gamma_{ij\alpha\omega} \qquad \text{for } j = 1, \cdots, n \qquad \alpha = 1, \cdots, p \\ &\sum_{j=1}^{n} \sum_{\beta=1}^{p} \lambda_{j\beta} = 1 \end{aligned} \tag{7.5}$$

which consists of $n \cdot p + 1$ linear equations. $\gamma_{jk\alpha\beta}$ and $\gamma_{ij\alpha\omega}$ are, again, found by regularising a theoretical point variogram (see section 7.3.2). Again, for numerical robustness, $n = 5$. Also the number of time steps p was limited, and two cases were considered. In the first case, five time steps ($t_\alpha - t_\omega$= -5, -2, 0, 2, 5 hrs) were used and in the second case nine time steps ($t_\alpha - t_\omega$= -20, -10, -5, -2, 0, 2, 5, 10, 20 hrs) were used.

7.3.1.3 Spatio-temporal cokriging

An alternative method, termed spatio-temporal cokriging, uses measurements from other time steps as covariates. Eq. 7.4 is used here as well but the difference from spatio-temporal kriging is that the sum of weights for each time step is set equal to zero, except for the weights of the time step of interest, t_ω, which sum up to one. Cokriging hence involves more constraints than spatio-temporal kriging. The kriging weights $\lambda_{j\alpha}$ can be found by solving the following linear system (Rouhani and Wackernagel, 1990):

$$\sum_{k=1}^{n}\sum_{\beta=1}^{p}\lambda_{j\beta}\gamma_{jk\alpha\beta}-\lambda_{j\alpha}\sigma_j^2+\mu_j=\gamma_{ij\alpha\omega} \qquad \text{for } j=1,\cdots,n \qquad \alpha=1,\cdots,p$$

$$\sum_{j=1}^{n}\lambda_{j\omega}=1 \tag{7.6}$$

$$\sum_{j=1}^{n}\lambda_{j\beta}=0 \qquad \forall\,\beta\neq\omega$$

$\gamma_{jk\alpha\beta}$ is the cross-variogram between $q(\mathbf{x}_j,t_\alpha)$ and $q(\mathbf{x}_k,t_\beta)$. There exist different definitions of the cross-variogram in the literature, the definition of Clark et al. (1987), which has been recommended by Cressie (1991), is used in this chapter. With information from different time steps as the covariates, the cross-variogram is defined as:

$$\gamma_{jk\alpha\beta}=\frac{1}{2}\text{var}\left[q(\mathbf{x}_j,t_\alpha)-q(\mathbf{x}_k,t_\beta)\right] \tag{7.7}$$

which is equal to a spatio-temporal variogram (Chapter 5). As for spatio-temporal kriging, n = 5, and p = 5 and 9.

7.3.1.4 Adjustment of kriging weights

If measurements are close to each other, i.e. clustered in space, they tend to be correlated. One advantage of kriging is to account for this effect by reducing the kriging weights of clustered measurements. However, in the limit, if two measurements are located at the same point in space, the kriging system will only give the sum of the weights and it is impossible to identify the weights individually. Measurements that are close may hence lead to ill-conditioned kriging systems and numerical problems may arise. Some of the kriging weights obtained may then be much smaller than zero. This is not a desirable result as they can produce artefacts in the estimates. As catchments are, occasionally, very close, and runoff measurements are highly correlated in time, this needs to be dealt with in Top-kriging. The issue was addressed in two ways in this chapter. First, non-zero local uncertainty of runoff, σ_j^2, was used. The larger σ_j^2 the fewer numerical problems will arise as it effectively decreases the correlation between neighbouring catchments. σ_j^2 has been set to a physically realistic value of 1% of the average temporal variance of runoff. However, occasional large negative weights still remained. These were adjusted by rescaling them.

The method of rescaling weights usually recommended in the literature (see Yamamoto, 2000) is to remove all negative weights and to rescale the remaining weights to a sum of

unity. However, as a large negative weight is usually the counterpart of a large positive weight, this method gives a loss of information. Instead, it is in this chapter proposed a method that keeps some of the negative weights and, in most instances, maintains the internal ranking of the weights. The starting point is that ill-conditioned kriging systems can be identified by large absolute sum of weights:

$$\Lambda = \sum_{j=1}^{n}\sum_{\beta=1}^{p}\left|\lambda_{j\beta}\right| >> 1 \tag{7.8}$$

where $\lambda_{j\beta}$ are the kriging weights. An upper limit had to be defined for this sum and, after testing different limits, $\Lambda_{max} = 1.5$ was found to be a suitable choice. All weights were then scaled according to:

$$\lambda'_{j\beta} = \lambda_{j\beta}\Lambda_{max} / \Lambda \tag{7.9}$$

As the sum of the weights is not longer unity the difference to unity is added to all weights.

$$\lambda''_{j\beta} = \lambda'_{j\beta} + \frac{1 - \sum_{j=1}^{n}\sum_{\beta=1}^{p}\lambda'_{j\beta}}{np} \tag{7.10}$$

For spatial and spatio-temporal kriging all weights were adjusted in this way while for spatio-temporal cokriging only the weights for time step t_ω are adjusted by Eqs. 7.9 and 7.10. Eq. 7.8 was used to recalculate Λ using the weights of Eq. 7.10, typically, giving $\Lambda > \Lambda_{max}$. The procedure (7.9, 7.10, 7.8) was then repeated until $abs(\Lambda - \Lambda_{max}) < \delta_\Lambda$ with $\delta_\Lambda = 0.05$. It is here suggested that the weights adjusted by this procedure respect the kriging system to a larger extent than the methods presented by Yamamoto (2000).

7.3.2 Concept of catchments as space-time filters

Runoff at the catchment outlet is a result of a complicated non-linear averaging process within the catchment over a period of time. This averaging process was in Chapter 5 described as a space-time filter that operates on local runoff. The same concept is used in this chapter to find the gamma values to be used in Eqs. 7.3, 7.5 and 7.6.

Following Woods and Sivapalan (1999) and Chapter 5, local runoff or rainfall excess $R(\mathbf{r},t)$ is conceptualised as a continuous point process in space $\mathbf{r}$ and time t. To account for routing on the hillslopes and in the channels within the catchment, a weighting function $u(\mathbf{r},t)$ is

introduced which allows to combine local instantaneous runoff into runoff at the catchment outlet, Q_i :

$$Q_i(t) = \int_{A_i} \int_{t-T_i}^{t} R(\mathbf{r},\tau) \cdot u(\mathbf{r},\tau) d\tau d\mathbf{r} \tag{7.11}$$

where A_i is the catchment area and T_i is the time interval that influences the output. As an approximation, it is assumed that for a given catchment, the weighting function is constant within the integration limits both in space and time, i.e. $u(\mathbf{r},t) = u_i = 1/T_i$. For a constant weighting function, Eq. 7.11 becomes a linear filter or a convolution integral. In time, the weighting function is equivalent to a unit hydrograph that is constant between 0 and T_i and zero elsewhere. In space, the weighting function is constant within the catchment area and zero elsewhere. The specific runoff at the catchment outlet then becomes:

$$q_i(t) = \frac{1}{A_i T_i} \int_{A_i} \int_{t-T_i}^{t} R(\mathbf{r}, t-\tau) d\tau d\mathbf{r} \tag{7.12}$$

In geostatistical terminology, A_i and T_i are the spatial and temporal supports, respectively. Again following Chapter 5, it is assumed that the temporal support is related to catchment area by:

$$T_i = \mu \cdot A_i^{\kappa} \tag{7.13}$$

where μ and κ are parameters to be estimated from the data. For $\kappa > 0$ the temporal support (or time base of the unit hydrograph) increases with catchment area.

In a geostatistical framework, the linear aggregation of Eq. 7.12 is performed on the second moments. A point variogram of runoff represents the second moment of local, instantaneous runoff. From the point variogram with zero support in space and zero support in time (i.e. instantaneous) one can estimate variograms that are valid for finite support areas and finite support times by a procedure that is usually referred to as regularisation (Journel and Huijbregts, 1978). Following Cressie (1991, p. 66) and Chapter 5, the gamma value $\gamma_{ij\alpha\beta}$ between two catchments i and j, and two different time steps, t_α and t_β is found by regularising a spatio-temporal point variogram γ_{st} :

$$\gamma_{ij\alpha\beta}=\frac{1}{A_i A_j T_i T_j}\int_{A_i}\int_{A_j}\int_{0}^{T_i}\int_{0}^{T_j}\gamma_{st}(|\mathbf{r}_1-\mathbf{r}_2|,|\tau_1+h_t-\tau_2|)d\tau_1 d\tau_2 d\mathbf{r}_1 d\mathbf{r}_2 -$$

$$0.5*\left[\frac{1}{A_i^2 T_i^2}\int_{A_i}\int_{A_i}\int_{0}^{T_i}\int_{0}^{T_i}\gamma_{st}(|\mathbf{r}_1-\mathbf{r}_2|,|\tau_1-\tau_2|)d\tau_1 d\tau_2 d\mathbf{r}_1 d\mathbf{r}_2+\frac{1}{A_j^2 T_j^2}\int_{A_j}\int_{A_j}\int_{0}^{T_j}\int_{0}^{T_j}\gamma_{st}(|\mathbf{r}_1-\mathbf{r}_2|,|\tau_1-\tau_2|)d\tau_1 d\tau_2 d\mathbf{r}_1 d\mathbf{r}_2\right]$$

(7.14)

where $h_t=|t_\alpha-t_\beta|$, $\mathbf{r}_1$ and $\mathbf{r}_2$ are spatial integration vectors within the two catchments, and τ_1 and τ_2 are the temporal integration variables. For the integration in Eq. 7.14 over A_i and A_j the catchment boundaries from the digital data base were used for each catchment. For spatial kriging, h_t=0 in Eq. 7.14.

Based on the comparison of different variogram models in Chapter 5, the exponential model is chosen, as the number of parameters is limited and some physical interpretation of the parameters is possible.

$$\gamma_{st}(h_s,h_t)=a(1-\exp(-((ch_t+h_s)/d)^b))+a_s h_s^{b_s}+a_t h_t^{b_t} \qquad (7.15)$$

The first term of this variogram represents the stationary part. a gives the variance of the (stationary) process, c relates space and time, d is the combined correlation length, and b gives the slope of the variogram. The second and the third terms give the non-stationary parts of the variogram in space and time, respectively.

7.3.3 A simple routing model

The space-time filter accounts for the effects of routing within the catchment on the spatial and temporal variance. However, it does not explicitly represent time lags between catchments, as the direction of time is independent of the direction in space. To account for time lags as a result of in-stream routing in the estimation procedure a routing model is proposed that consists of applying a time lag that is constant with time. To estimate runoff at time step t_ω of catchment i, $q(\mathbf{x}_j,t_\alpha^*)$ at time step $t_\alpha^*=t_\alpha+t_\alpha'$ is used instead of t_α in Eqs. 7.2 and 7.4. Depending on the relative position on the stream network of catchments i and j, the time lag t_α' can be positive or negative.

Three variants of determining the time lags were examined. In the first variant (termed "nested routing") travel times were inferred from cross variograms for catchment pairs estimated from the runoff data (Eq. 7.20 below). The minimum variance of the cross variograms typically occurred for a temporal separation larger than 0. Dividing the spatial

distance between the stream gauges by that temporal separation gave a flow velocity for each pair of catchments. For the Innviertel region this method indicated an average velocity of v=0.67 m/s. For non-nested catchments the time lag was set to zero.

$$\begin{array}{ll} t_\alpha' = d_{ij} / v & \text{if } i \text{ and } j \text{ are nested} \\ t_\alpha' = 0 & \text{if } i \text{ and } j \text{ are non-nested} \end{array} \tag{7.16}$$

where d_{ij} is the distance between the two stream gauges. d_{ij} is positive when j is a downstream neighbour and negative when j is an upstream neighbour of i. If d_{ij} / v does not correspond to an integer time step, a weighted average of q of the two time steps is used.

The second variant (termed "routing all") uses the lags for nested catchments as above but makes different assumptions about the lags for non-nested catchments. Typically, the characteristic velocity of precipitation is much larger than that of flow routing processes in catchments (Chapter 4). This means that the difference in the timing of a flood event of two catchments of different size that are close to each other will be mainly due to routing differences as the time difference due to the motion of the rainfall system will be much smaller. As an approximation, the time lag is related to the size of the catchments. For a single catchment i the time lag T_{Li} of runoff relative to rainfall has been estimated as:

$$T_{Li} = \xi \cdot A_i^{\psi} \tag{7.17}$$

following Melone et al. (2002). For the Innviertel region $\xi = 1.5$ and $\psi = 0.35$ (Merz and Blöschl, 2003). The time lag of two non-nested catchments is then assumed to be the difference of the individual time lags:

$$\begin{array}{ll} t_\alpha' = d_{ij} / v & \text{if } i \text{ and } j \text{ are nested} \\ t_\alpha' = T_{Lj} - T_{Li} & \text{if } i \text{ and } j \text{ are non-nested} \end{array} \tag{7.18}$$

To analyze the effect of the routing model on runoff estimation a third variant termed "no routing" was also examined where all time lags were set to zero:

$$t_\alpha' = 0 \tag{7.19}$$

7.3.4 Hydrologic interpretation of Top-kriging

There are two main groups of processes that control runoff. The first group consists of variables that are continuous in space and include rainfall, evapotranspiration and soil characteristics. In Top-kriging their variability is represented by the point variogram that is

based on Euclidian distances. The second group of processes is related to routing on the hillslope and in the stream network. Their effect cannot be represented by Euclidian distances. Top-kriging represents these processes in three ways:

(a) The channel network structure and the similarity between upstream and downstream neighbours are represented by the catchment area that drains to a particular location on the stream network. The catchment areas are defined by their boundaries in space.

(b) Advective runoff routing (Fig. 7.3) is represented by a simple routing model (Eqs. 7.16-7.19). This model takes into account the travel time between upstream and downstream neighbours and involves direction in both space and time. The routing is incorporated directly in the estimation procedure (Eqs. 7.2 and 7.4) by a time lag in estimating runoff as a weighted average of the runoff of neighbouring catchments.

(c) Dispersive routing is represented by the space-time filter (Eq. 7.14). Dispersive effects include hillslope routing and what Rinaldo et al. (1991) refer to as hydrodynamic and geomorphologic dispersion (Fig. 7.3). Hydrodynamic dispersion is caused by different travel times in the stream within individual reaches and is related to the pressure term in the St. Venant equation. Geomorphologic dispersion is related to the different lengths and junctions of the stream network and results in a superposition of runoff from the tributaries. The representation of dispersion in Top-kriging has an analogy in the unit hydrograph concept. Application of the unit hydrograph concept involves two steps – estimation of catchment rainfall (e.g. by areal reduction factors, Sivapalan and Blöschl, 1998) and convolution of catchment rainfall with the unit hydrograph. The former step is a spatial filter, the latter step a temporal filter. Both are represented in the space-time filter of Top-kriging that is used to estimate the gamma values and hence the kriging weights.

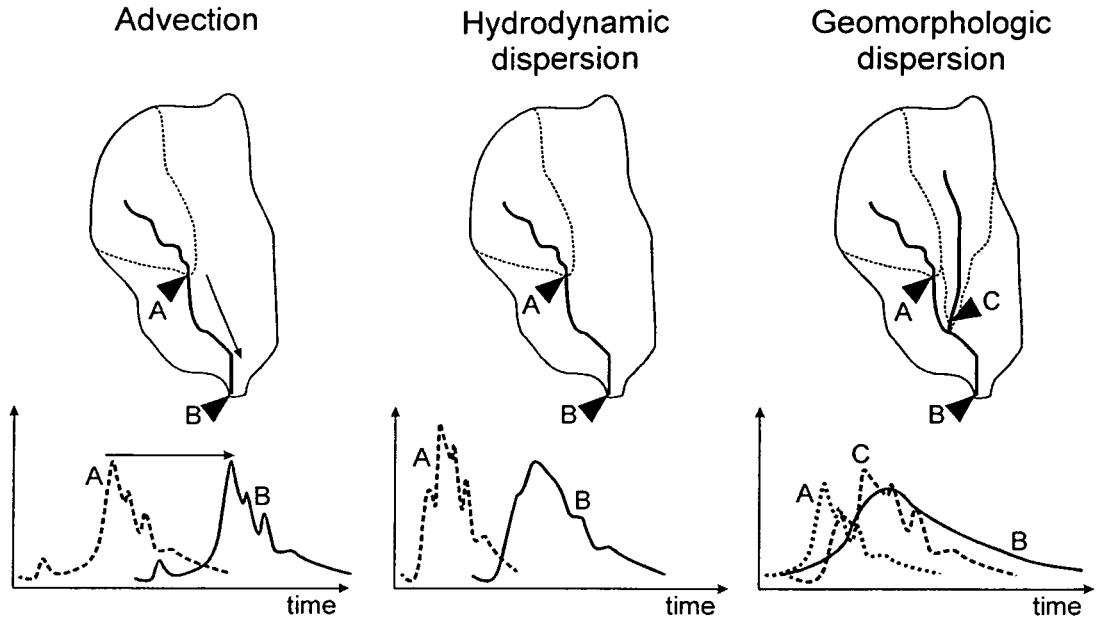


Figure 7.3 Schematic of runoff routing processes represented by Top-kriging.Hydrographs for locations A, B, C are shown.

7.3.5 Estimation of the spatio-temporal point variogram

For applying Top-kriging, a spatio-temporal point variogram is needed in the region of interest. The point variogram was in this Chapter inferred from the runoff data in the Innviertel region in the following way. In a first step, cross variograms $\hat{\gamma}_{ij}(h_t)$ were estimated for all pairs of catchments i and j, which resulted in a family of temporal variograms:

$$\hat{\gamma}_{ij}(h_t) = \frac{1}{2n(h_t)} \sum_{t=1}^{n(h_t)} (q(\mathbf{x}_i, t_t + h_t) - q(\mathbf{x}_j, t_t))^2 \qquad (7.20)$$

where $q(\mathbf{x}_i, t_t)$ is runoff at time t_t of stream gauge i with spatial location $\mathbf{x}_i$, h_t is the temporal lag, and $n(h_t)$ is the number of pairs of runoff measurements in time for the temporal lag bin associated with h_t.

In a second step, theoretical gamma values $\gamma_{ij}(h_t)$ were estimated for pairs of catchments from a theoretical point variogram γ_{st} using the same regularisation method as in Eq. 7.14. The parameters of the point variogram Eq. 7.15 (a, b, c, d, a_s, a_t, b_s, b_t) are initially unknown as are the parameters of Eq. 7.13 (μ, κ). To obtain these parameters the theoretical gamma values $\gamma_{ij}(h_t)$ were fitted to the sample variograms $\hat{\gamma}_{ij}(h_t)$ by minimising the objective function Φ:

$$\Phi = \frac{2}{N(N+1)M} \sum_{i=1}^{N} \sum_{j=i}^{N} \sum_{m=1}^{M} \min\left\{ \left[\frac{\hat{\gamma}_{ij}(h_t)}{\gamma_{ij}(h_t)} - 1 \right]^2, \left[\frac{\gamma_{ij}(h_t)}{\hat{\gamma}_{ij}(h_t)} - 1 \right]^2 \right\} \tag{7.21}$$

N is the total number of catchments and M is the total number of temporal bins. Eq. 7.21 is a modified version of the Weighted Least Squares (WLS) method. The WLS-method as introduced by Cressie (1985) only contains the first term of min{...} in Eq. 7.21. This gives asymmetrical errors, as errors of overestimation are limited to 1, while errors of underestimation can be very large and can mask the errors from other catchment pairs and temporal lags. The modified version used here limits all errors to the range from 0 to 1. Eq. 7.21 is normalised to also give Φ in the range from 0 to 1.

Initial tests indicated that the type of routing method did not change the parameter values of the point variogram much. The parameters are hence estimated from cross-variograms inferred without routing.

As a guidance for the reader, Fig. 7.4 presents a flow scheme of the estimation process.

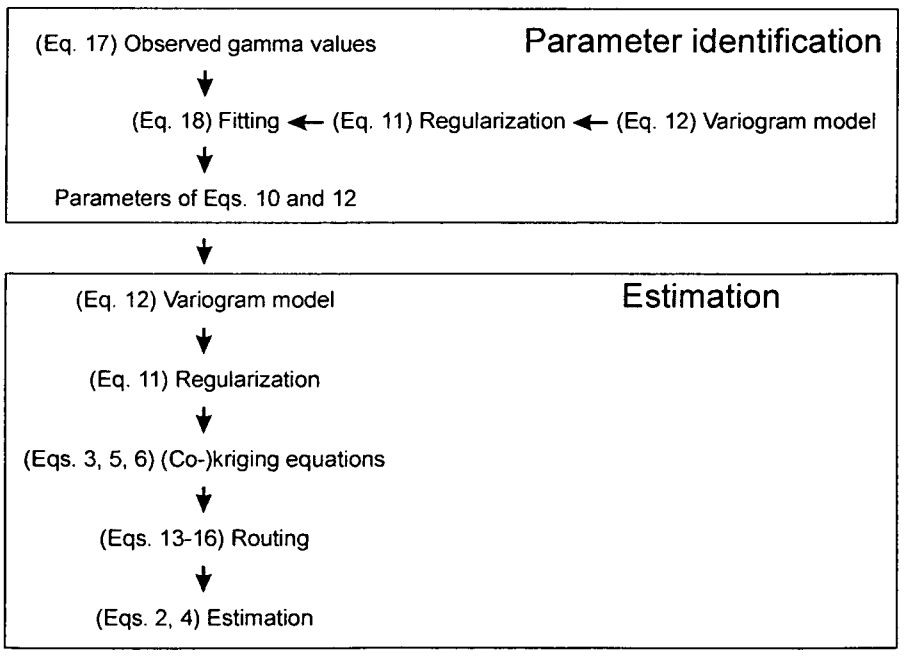


Figure 7.4 Flow scheme of the estimation process presented in this chapter.

7.3.6 Evaluation of methods

To examine the predictive performance of the proposed method for ungauged catchments cross-validation analysis was carried out. One runoff record was withheld from the data set,

the runoff time series estimated for that catchment from measurements of the neighbouring stream gauges and finally the estimates were compared with the runoff data of that catchment. The model efficiency (ME_i) according to Nash and Sutcliffe (1970) was used as performance statistic for each target catchment i:

$$ME_i = 1 - \frac{\sum_{\omega=1}^{\Omega}(q_i(\omega) - \hat{q}_i(\omega))^2}{\sum_{\omega=1}^{\Omega}(q_i(\omega) - \overline{q_i(\omega)})^2} \tag{7.22}$$

where Ω is the number of time steps estimated (Ω=87672 for the ten years analyzed in this chapter) and $\overline{q_i(\omega)}$ is the mean of the runoff data for the same time period. The efficiency $ME \leq 1$, where ME=1 indicates perfect estimation and ME=0 means that the estimation method performs no better than the mean of the runoff data. In addition to the Nash-Sutcliffe efficiency, the estimated hydrographs were examined visually to understand the dynamics of the estimated runoff time series.

A total of 15 estimation variants were tested; three routing models (no routing, routing model for nested catchments, and routing model for all catchments), five methods of Top-kriging (spatial kriging, and spatio-temporal kriging and spatio-temporal cokriging with the five and nine time steps) and combinations thereof.

The proposed method is an alternative to deterministic runoff models that use regionalised model parameters. To illustrate the relative merits of the two genres of methods, the Top-kriging estimates were compared with simulations of a deterministic rainfall-runoff model taken from the study of Parajka et al. (2005). The runoff model is a conceptual soil moisture accounting scheme that uses precipitation and air temperature data as inputs and runs on a daily time step. It consists of a snow routine, a soil moisture routine and a flow routing routine and involves 14 model parameters. Three of the parameters were preset in their study, leaving 11 parameters to be found by model calibration. Parajka et al. (2005) first calibrated the model to 320 catchments in Austria. They then regionalised the calibrated model parameters by different methods and examined the model performance for the ungauged catchment case by cross-validation. Their results (both locally calibrated and regionalised) were contrasted with the results from Top-kriging. As their analysis is based on a daily time step, the hourly Top-kriging estimates were averaged to daily values and compared the daily runoff time series. A first comparison focuses on the Innviertel region and involves 17 stream gauges that are common to both studies. To provide context, a second comparison examines 208 stream

gauges in Austria that are common to both studies. The same procedure and parameters as in the Innviertel region are used for all of Austria including, for consistency, the same point variogram. The results from their calibration period (1987-1997), which has 74% overlap with the period used in this study (1990-2000), were used for the comparison.

7.4 Results

7.4.1 Estimation of point variogram

The estimated parameters of the point variogram are shown in Table 7.2. The parameters are similar to those obtained in Chapter 5 for 488 catchments in Austria although differences exist. For example, the spatial correlation length of $d = 2.3$ km found here is somewhat larger than those found in Chapter 5 (1.0 km) which may be due to the somewhat more homogeneous runoff in the Innviertel region as compared to the rest of Austria. The exponent κ relating temporal and spatial supports in (Eq. 7.13) is smaller (κ=0.17 instead of 0.4) and the scale μ is similar (2.9 as compared to 1.9) indicating that the catchments in the Innviertel respond somewhat faster than the average of the catchments in Austria. For example, the temporal support of a 100 km² catchment found in this study is $T_i = 6$ hrs, as compared to 12 hrs found Chapter 5. For consistency across the two scales (Innviertel and Austria) the parameters of Table 7.2 were used for all analyses in this chapter.

Figure 7.5 shows a comparison of the sample gamma values and the gamma values obtained by regularising the point variogram (Eq. 7.15, Table 7.2). The regularised gamma values exhibit some scatter as indicated by the error bars, but they are very close to unbiased. This suggests that the point variogram parameters of Table 7.2 are a realistic representation of the space time variability of runoff.

As estimates of the local uncertainty in Eqs. 7.3, 7.5 and 7.6 were unavailable in this study, a value of $\sigma_j^2 = 0.000005$ $m^6s^{-2}km^{-4}$ was used based on test simulations. This value is about 1% of the average temporal variance of runoff in the Innviertel region (Table 7.1).

Table 7.2 Parameters of the point variogram (Eq. 7.15) and Eq. 7.13 estimated for the Innviertel region.

Parameter	Value	Units
a	0.00139	$m^6s^{-2}km^4$
b	0.445	
c	0.300	km hr^{-1}
d	2.31	km
a_s	0.00003	$m^6s^{-2}km^4$
a_t	0.00009	$m^6s^{-2}km^4$
b_s	0.0247	-
b_t	0.186	-
μ	2.90	hrs
κ	0.167	-

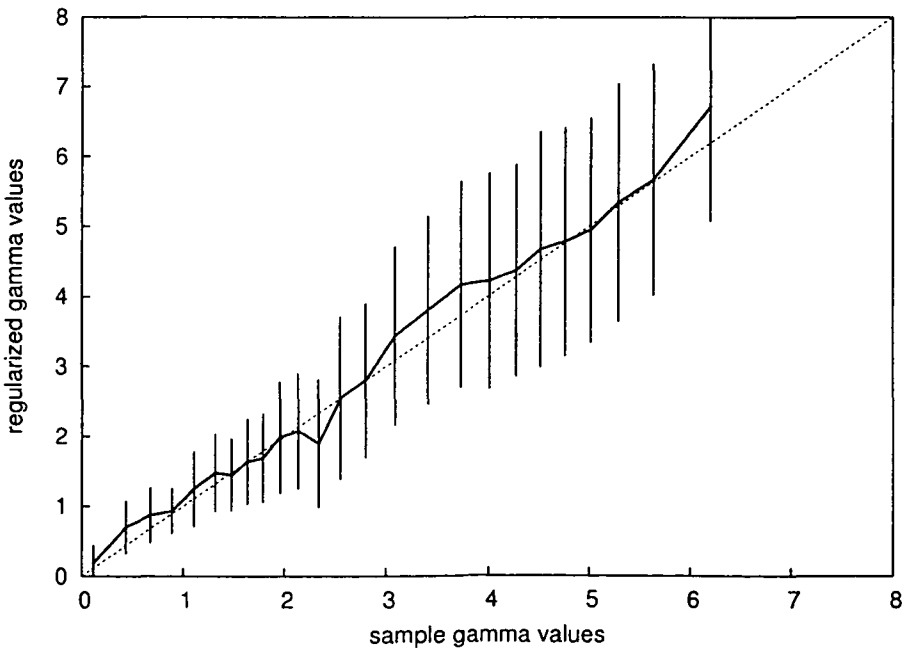


Figure 7.5 Comparison of sample gamma values (Eq. 7.20) with gamma values obtained by regularising the point variogram (Eq. 7.15, Table 7.2) in terms of the mean (thick solid line) and standard deviation (error bars). Dashed line shows 1:1 line. Units are $m^6s^{-2}km^{-4}$ $*10^{-4}$.

7.4.2 Estimation performance

Table 7.3 shows model efficiencies (ME) of the hourly runoff time series estimated by Top-kriging in the Innviertel region. The largest model efficiency (ME) for each catchment is printed in bold, the largest ME for each catchment and each routing model is printed in italics.

Considering spatial kriging first, ME ranges between 0.72-0.96 with an average ranging between 0.84 and 0.86 depending on the routing model. The routing all model gives the highest ME for eleven out of the 19 catchments and the nested routing model gives the highest ME for seven catchments indicating that routing does improve the runoff estimates over the variant where no routing is used. For spatio-temporal kriging and spatio-temporal cokriging the advantage of the routing model is similar, with the routing all model and the nested routing model giving the highest ME for eleven and seven catchments, respectively.

A comparison of the efficiency of spatial kriging with that of spatio-temporal kriging and cokriging for the routing all model indicates that spatial kriging outperforms the other variants of kriging in twelve catchments. In the remaining seven catchments the efficiency of spatio-temporal kriging or cokriging is in most cases only marginally higher than that of spatial kriging, whereas one or more models of spatio-temporal kriging and cokriging perform considerably poorer than spatial kriging for almost all catchments. The median and the average show a similar tendency for spatial kriging to outperform spatio-temporal kriging and spatio-temporal cokriging.

The best runoff estimates are obtained for those catchments that have one or more close upstream neighbours. The highest ME is found for Taufkirchen (ME around 0.96, depending on the choice of routing model). Upstream gauges cover 248 km² of the 303 km² of the catchment. It is therefore not surprising that a weighted average of the upstream neighbours gives an estimated hydrograph that is similar to the observed hydrograph. The more of the catchment area that is shared with neighbouring catchments the larger the efficiency. ME is around 0.9 for most of the catchments with several upstream neighbours, or which are a large part of a downstream catchment (e.g. Winertsham, Kropfmühle, Fraham, Bad Schallerbach). The catchments without such neighbours generally have ME on the order of 0.8 or lower (e.g. Ried, Osternach, Danner, Pram, Pichl).

Table 7.3 Model efficiencies of hourly Top-kriging estimates of runoff in the Innviertel region for the period Aug. 1, 1990 to July 31, 2000. S refers to spatial kriging, STK1 and STK2 refer to spatio-temporal kriging with 5 and 9 time steps, respectively, and STC1 and STC2 refer to spatio-temporal cokriging with 5 and 9 time steps, respectively. The largest model efficiency (ME) for each catchment is printed in bold, the largest ME for each catchment and each routing model is printed in italics.

Catchment	No routing					Nested routing					Routing all model				
	S	STK1	STK2	STC1	STC2	S	STK1	STK2	STC1	STC2	S	STK1	STK2	STC1	STC2
Ried	0.76	**0.77**	0.74	0.76	0.76	0.76	**0.77**	0.74	0.76	0.76	0.75	*0.77*	0.74	0.76	0.76
Danner	0.76	*0.78*	0.76	0.76	0.76	0.76	*0.78*	0.76	0.76	0.76	0.79	**0.80**	0.78	0.78	0.78
Osternach	*0.82*	0.79	0.70	0.81	0.81	*0.82*	0.79	0.70	0.81	0.81	**0.82**	0.79	0.70	0.82	0.82
Pram	*0.80*	0.74	0.67	0.76	0.76	*0.80*	0.74	0.67	0.76	0.76	**0.80**	0.75	0.67	0.77	0.77
Riedau	0.84	*0.88*	0.85	0.87	0.87	0.87	**0.90**	0.87	0.89	0.89	0.87	*0.90*	0.87	0.89	0.89
Winertsham	0.88	*0.90*	0.83	0.89	0.89	*0.91*	0.90	0.83	0.91	0.91	0.91	0.92	0.84	**0.93**	0.93
Taufkirchen	*0.94*	0.94	0.89	0.92	0.92	0.96	**0.96**	0.89	0.96	0.96	0.96	**0.96**	0.89	0.96	0.96
Angsüß	0.77	*0.79*	0.77	0.74	0.75	0.82	*0.83*	0.79	0.77	0.78	0.82	**0.84**	0.80	0.79	0.80
Alfersham	0.73	0.70	*0.76*	0.56	0.60	*0.78*	0.72	0.78	0.57	0.61	**0.78**	0.70	0.77	0.53	0.57
Lohstampf	0.84	0.83	0.73	*0.86*	0.85	*0.87*	0.85	0.76	0.87	0.87	**0.88**	0.83	0.73	0.86	0.86
Still	0.89	0.90	0.81	0.91	**0.91**	0.86	0.89	0.82	0.90	*0.90*	0.86	0.87	0.82	0.87	*0.87*
Strötting	*0.84*	0.83	0.76	0.82	0.81	*0.87*	0.81	0.77	0.83	0.83	**0.87**	0.82	0.77	0.83	0.83
Bad Schallerb.	*0.92*	0.90	0.83	0.91	0.91	*0.93*	0.90	0.83	0.91	0.91	**0.93**	0.91	0.84	0.91	0.91
Pichl	*0.80*	0.70	0.59	0.66	0.62	**0.83**	0.73	0.58	0.68	0.64	*0.82*	0.72	0.58	0.67	0.63
Weghof	*0.83*	0.43	0.44	0.27	0.22	**0.86**	0.46	0.44	0.30	0.25	*0.85*	0.45	0.44	0.29	0.24
Fraham	*0.89*	0.84	0.85	0.81	0.81	**0.93**	0.90	0.86	0.90	0.89	*0.92*	0.90	0.86	0.90	0.89
Neumarkt	**0.89**	0.85	0.76	0.87	0.87	0.88	0.86	0.77	*0.89*	0.89	0.88	0.86	0.77	*0.89*	0.89
Niederspaching	*0.92*	0.83	0.75	0.86	0.87	**0.93**	0.84	0.75	0.87	0.87	*0.92*	0.86	0.75	0.89	0.89
Kropfmühle	*0.90*	0.88	0.83	0.88	0.88	*0.90*	0.88	0.83	0.88	0.88	**0.91**	0.88	0.83	0.87	0.86
Average	*0.84*	0.80	0.75	0.79	0.78	*0.86*	0.82	0.76	0.80	0.80	**0.86**	0.82	0.76	0.80	0.80
Median	*0.84*	0.83	0.76	0.82	0.81	**0.87**	0.84	0.77	0.87	0.87	*0.87*	0.84	0.77	0.86	0.86

7.4.3 Runoff dynamics

To analyze the ability of Top-kriging to reproduce the runoff dynamics, a large number of event hydrographs were examined visually. A few examples are given below to illustrate the main characteristics of Top-kriging. The illustration begins with estimates of the spatial kriging method. Taufkirchen is the catchment with the highest model efficiency (ME) according to Table 7.3. One event is shown in Fig. 7.6. Although there is a large variety in the runoff from the neighbouring catchments (bottom panel) and the peaks occur at different times, Top-kriging with routing for all catchments is able to estimate both the magnitude and the timing of the peak with high accuracy (top panel).

The bottom panel of Fig. 7.6 gives the kriging weights (first number) and the time lags (second number) of the neighbours used. A positive time lag refers to a downstream or larger catchment (use of future measurements) while a negative number refers to an upstream or smaller catchment (use of earlier measurements). Winertsham (128.1 km²) is the largest tributary (Fig. 7.2), and is associated with the largest weight of 0.43. Alfersham (81.3 km²) is smaller and has a weight of 0.37, while its upstream neighbour Angsüß is associated with a weight of 0.02. This very low weight is a result of the large correlation between Alfersham and Angsüß. The sum of weights from this tributary is then 0.39. The smallest tributary is Lohstampf (39.3 km²) with a weight of 0.17. Riedau has a weight of -0.001. Angsüß and Riedau are upstream neighbours of Alfersham and Winertsham respectively, and each pair of nested catchments can be regarded as a clustered sample, reducing the weight for the upstream neighbour in this case.

The time lags are all negative for Taufkirchen. This is because all catchments are upstream neighbours, and a peak flow will reach these catchments before it reaches Taufkirchen. The time lag is largest for Riedau, which is furthest away (13.7 km), with 5.70 hours. The smallest time lag is for Alfersham (2.9 km away) with 1.19 hours.

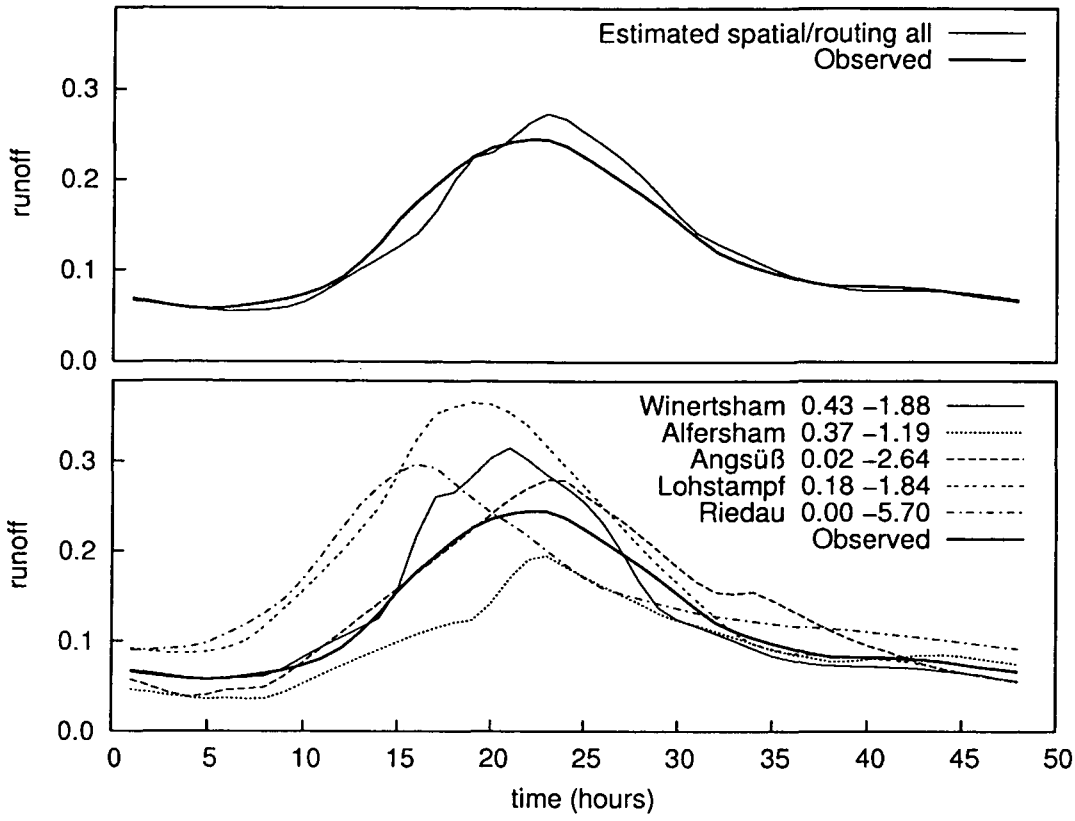


Figure 7.6 Observed and estimated hydrographs for Taufkirchen (303 km²) (top panel) and hydrographs of the neighbours (bottom panel) for the period March 17-18, 1993. The first number after neighbour name represents the kriging weight of the neighbour, the second number the lag of the routing model in hours. Spatial kriging, routing all model. Units of runoff are $m^3s^{-1}km^{-2}$.

Although Top-kriging works best with several upstream neighbours, the estimates are also satisfying for the smaller catchments. Figure 7.7 shows the observed and the estimated hydrographs of an event in August 1991 for the 52 km² Strötting catchment. This is a catchment without upstream neighbours, but with two downstream neighbours, Bad Schallerbach (184 km²) and Fraham (362 km²). As Bad Schallerbach is the first downstream neighbour, it gets the largest weight with 0.61. This is partly compensated by a negative weight of –0.10 for Fraham, which is the downstream neighbour of Bad Schallerbach, so that the total weight of the downstream measurements is 0.51. These catchments are clustered, and Top-kriging gives a negative weight to the one least correlated with Strötting. The non-nested neighbouring catchment Pichl (66 km²) gets a weight of 0.42, partly compensated by the weight of –0.15 of its downstream neighbour, Weghof (117 km²). Danner (56 km²) is slightly smaller than Pichl and gets a weight of 0.22.

The most important stream gauge for the estimation of runoff at Strötting is Bad Schallerbach. The peak of this catchment is quite similar to the peak of Strötting, but slightly delayed. The other catchments east and south of Strötting show considerably lower peaks, while the non-nested Danner catchment has a peak more than twice of that at Strötting. The weighted average gives a reasonably well-estimated peak. The close match is also a result of the time lags applied. A time lag of 5.64 hours was applied to Bad Schallerbach, which is also apparent in the figure. As Pichl, Weghof and Danner are not nested with Strötting, their time lags have been estimated by Eqs. 7.17-7.18. Although the time lags are considerably smaller than those of Bad Schallerbach and Weghof, Fig. 7.7 does confirm that the time lags are reasonable.

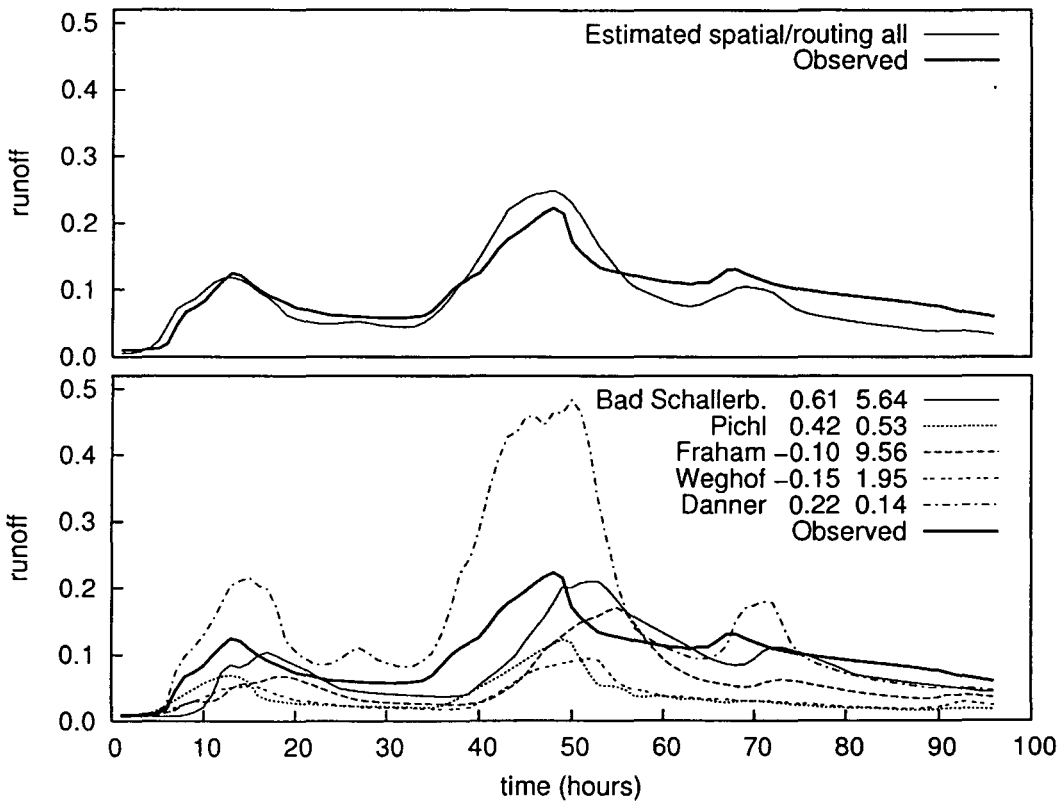


Figure 7.7 Observed and estimated hydrographs for Strötting (52 km²) (top panel) and hydrographs of the neighbours (bottom panel) for the period August 1-4, 1991. The first number after neighbour name represents the kriging weight of the neighbour, the second number the lag of the routing model in hours. Spatial kriging, routing all model.

There are also some events where the Top-kriging estimates differ significantly from the observed hydrographs. This typically happens for local events that do not appear in the hydrographs of the neighbouring catchments. One example for Ried (69 km²) is presented in

Fig. 7.8. This catchment has neither nested upstream nor downstream neighbours. In the three day period shown two events occurred. The first event (shown at time 18 hours) was properly estimated. The second event (shown at time 48 hours) was not. This is because it was apparently the result of a local precipitation event. None of the surrounding catchments experienced more than a small change in runoff for this period. It is interesting that the local event was also missed by the rain gauges in the region, so rainfall-runoff models could not have simulated this event either.

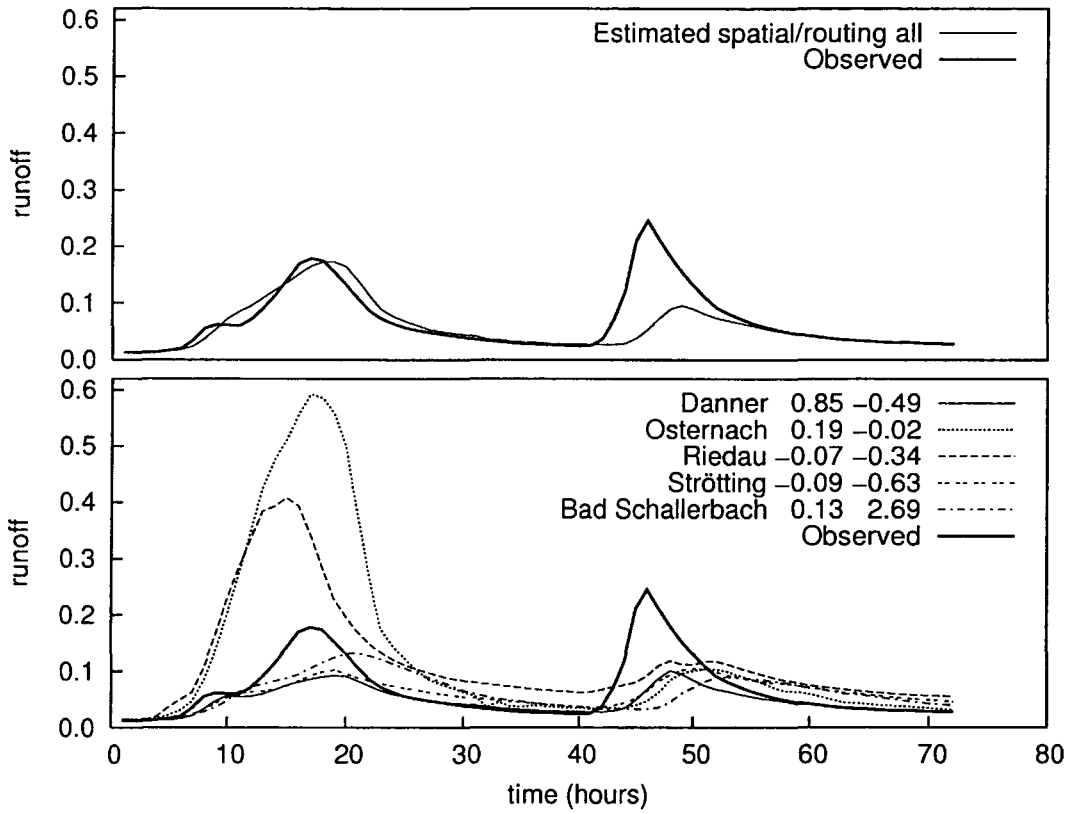


Figure 7.8 Observed and estimated hydrographs for Ried (69 km²) (top panel) and hydrographs of the neighbours (bottom panel) for the period Oct. 29 –Nov. 1, 1998. The first number after neighbour name represents the kriging weight of the neighbour, the second number the lag of the routing model in hours. Spatial kriging, routing all model. Units of runoff are $m^3s^{-1}km^{-2}$.

The merits of the routing model, in many cases, are obvious in the hydrographs. The top panel of Fig. 7.9 shows the estimated hydrograph for Fraham (362 km²) with and without the use of a routing model. The time lags presented in the bottom panel show that all the neighbours have their peaks earlier than Fraham. The largest weights are given to the two large tributaries Bad Schallerbach (184 km²) with 0.49 and Weghof (117 km²) with 0.51. The larger weight for

Weghof is partly compensated by the negative weight of –0.10 of Pichl, giving 0.41 as the total weight for this tributary. The upstream neighbour of Bad Schallerbach, Strötting (52 km²), has a negative weight of -0.02. The non-nested neighbour of equal size, Kropfmühle (313 km²), has a weight of 0.12. Fraham is a large catchment, so the time lags are considerably larger than those of the earlier examples. Strötting is 23 km from Fraham and the time lag is 9.6 hours. The other nested stream gauges are closer, with time lags ranging from 3 to 6 hours.

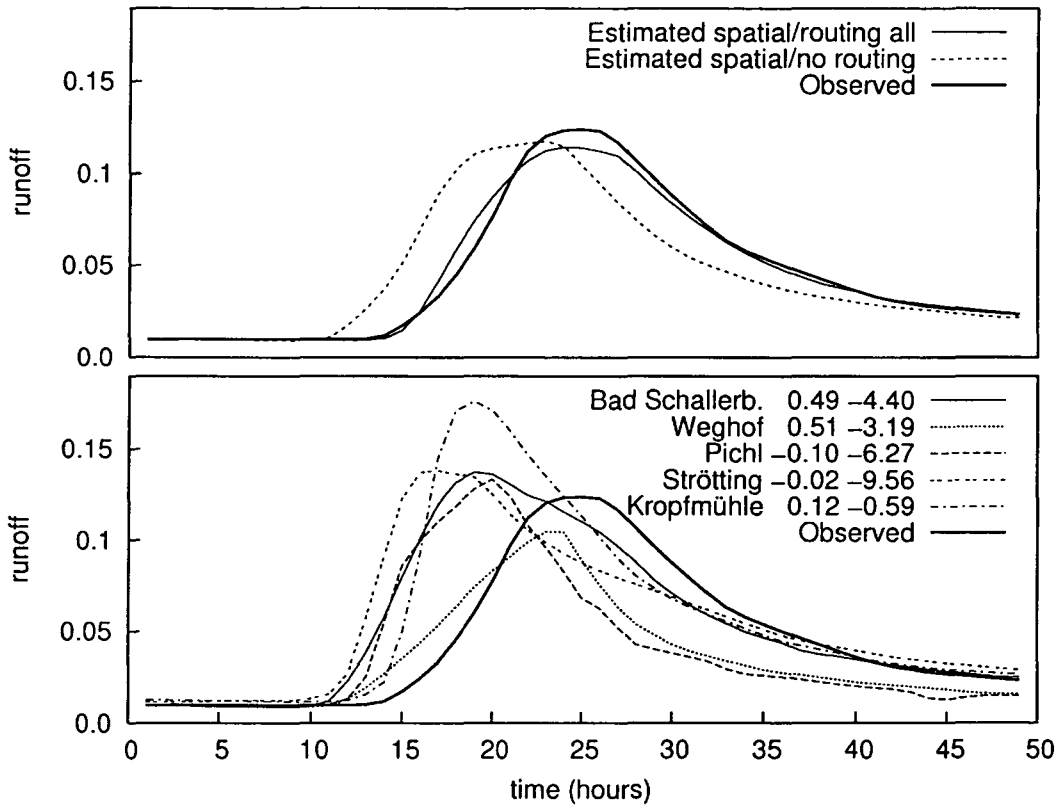


Figure 7.9 Observed and estimated hydrographs for Fraham (362 km²) (top panel) and hydrographs of the neighbours (bottom panel) for the period Oct. 23-25, 1993. The first number after neighbour name represents the kriging weight of the neighbour, the second number the lag of the routing model in hours. Spatial kriging and two routing models. Units of runoff are $m^3s^{-1}km^{-2}$.

The model efficiencies presented in Table 7.3 indicate that spatio-temporal kriging and spatio-temporal cokriging usually do not perform as well as spatial kriging. However, there are events where the spatio-temporal approach performs better than spatial kriging. One example is an event at Taufkirchen in November 1990, presented in Fig. 7.10. The top panel shows that spatial kriging and spatio-temporal cokriging are quite similar and overestimate the peak.

The hydrograph estimated by spatio-temporal kriging on the other hand is closer to the observed hydrograph, although it is slightly too flat at the beginning and the end of the event. There are two reasons for this effect. First, spatio-temporal kriging, in this case, involves weights for nine time steps ($t_\alpha - t_\omega$ ranging from –20 to 20 hrs) as opposed to spatial kriging where all weights are associated with the same time step ($t_\alpha = t_\omega$). As the weights are distributed over more time steps, the estimation method more strongly smoothes peaks observed at the neighbouring catchments. Second, the weights were rescaled (see section 7.3.1.4). This tended to reduce the weights of the closest neighbours and increase the weights of the more remote neighbours. Summed over all time steps, the weights associated with the neighbours of Taufkirchen ranged from 0.15 to 0.26 instead of 0.00 to 0.43 for spatial kriging. These two effects consistently gave smoother hydrographs than spatial kriging. In many cases they were smoother than the observed hydrographs.

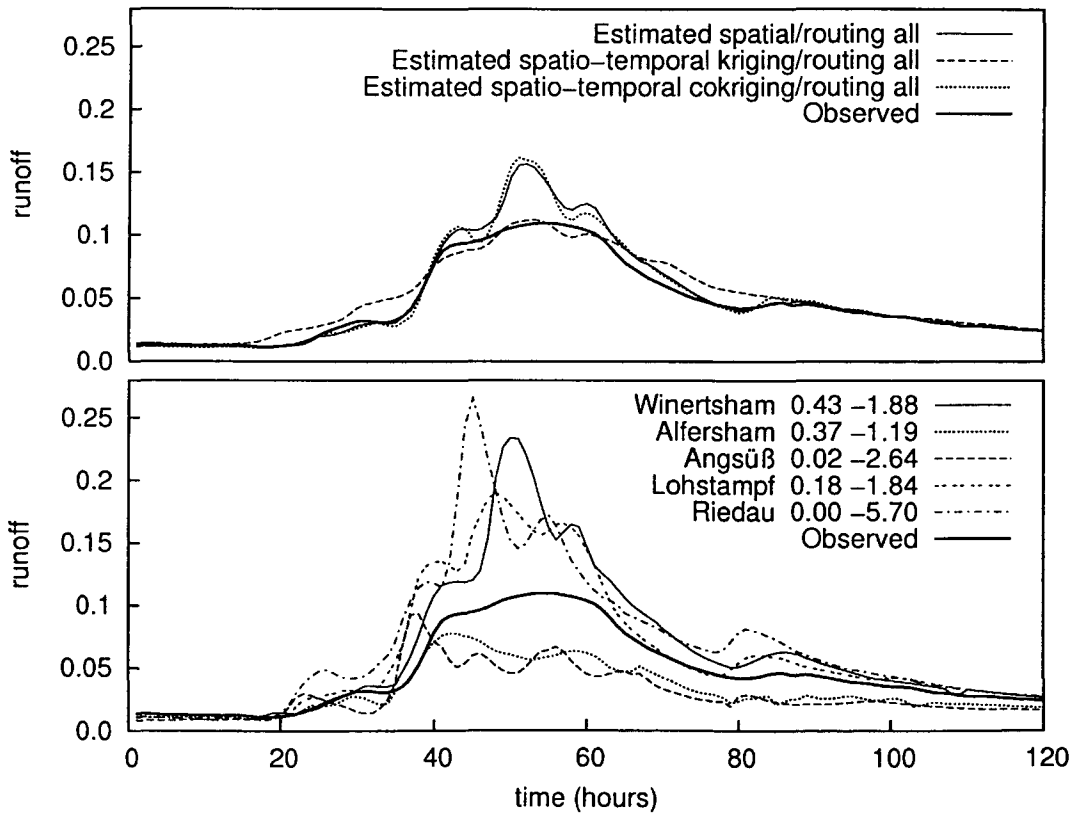


Figure 7.10 Observed and estimated hydrographs for Taufkirchen (303 km²) (top panel) and hydrographs of the neighbours (bottom panel) for the period Nov. 18 –23, 1990. The first number after neighbour name represents the kriging weight of the neighbour, the second number the lag of the routing model in hours. Spatial kriging, spatio-temporal kriging (nine time steps) and spatio-temporal cokriging (nine time steps). Units of runoff are $m^3s^{-1}km^{-2}$.

7.4.4 Temporal average and variance of the time series

When estimating runoff time series of ungauged catchments it is important to represent the statistical moments of the temporal variability well. The left panel of Fig. 7.11 presents the average runoff estimated by spatial kriging plotted against the observed averages for the Innviertel catchments. There is a good correspondence with a slight tendency for the small averages to be overestimated and the larger averages to be underestimated. There is no obvious relationship between this tendency and catchment area.

The right panel of Fig. 7.11 presents an analogous plot for the temporal variance of runoff. Overall, there is a relatively good correspondence between the estimated and observed variances although the largest variances are underestimated. This is particularly the case for the Ried, Osternach and Pram catchments. These are headwaters, and Osternach and Ried do not have nested neighbours at all. Danner is located between Osternach and Ried but has a considerably lower temporal variance which is, conversely, overestimated. Clearly, Top-kriging has difficulties with representing local hydrological effects well that are not reflected in the runoff data of neighbouring catchments. When examining a hydrogeological map (BMLFUW, 2003, Map 6.2) there are apparent hydrological reasons for the differences in the temporal variances of these catchments. Danner has a large number of artesian springs while Osternach has none. The subsurface contribution to runoff in Danner is hence larger than that in Osternach. Some of the Danner catchment area is forested with the remaining land being agricultural while all of the Osternach catchment is agricultural which is likely another factor contributing to the differences in the dynamics. Out of the catchments in the Innviertel region, the Weghof catchment has the smallest temporal variance, which can be estimated well even though the low variance is due to local gravel deposits in the valleys of the catchment.

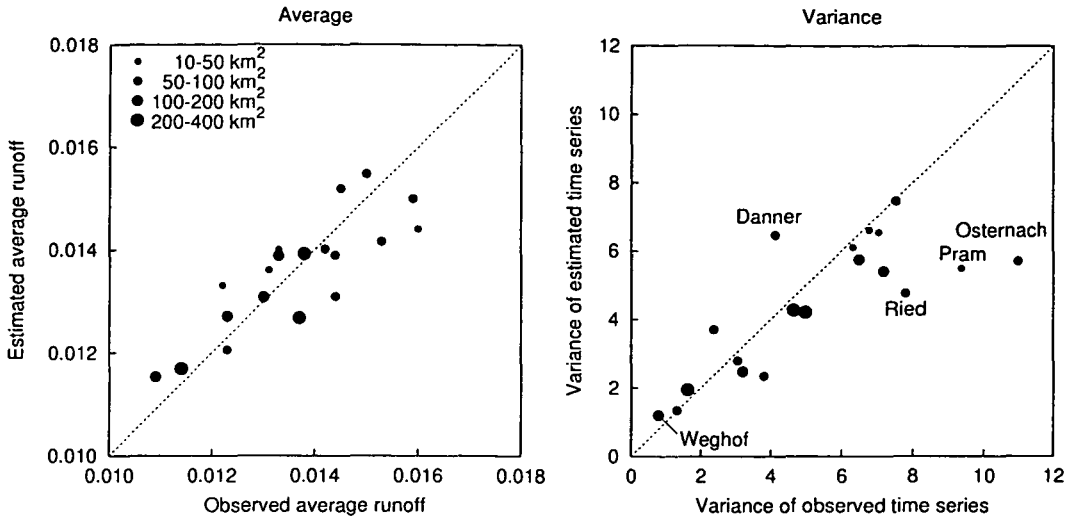


Figure 7.11 Left: Average runoff estimated by Top-kriging (spatial kriging, routing all model) plotted against average runoff of the observed time series for the 19 catchments in the Innviertel region. Right: Temporal variance of the estimated time series plotted against the variances of the observed time series. The size of the points indicates catchment area. Units are $m^3s^{-1}km^{-2}$ for average runoff and $m^6s^{-2}km^{-4}$ $*10^{-4}$ for the variance.

The underestimation of the large temporal variances is more pronounced for spatio-temporal kriging and spatio-temporal cokriging than for spatial kriging as illustrated in Fig. 7.12. In 14 out the 19 catchments, spatio-temporal kriging and cokriging underestimate the variance, which is clearly not a desirable property. The figure indicates that spatio-temporal cokriging retains more of the observed variance than spatio-temporal kriging. It also indicates that the spatio-temporal estimation methods with 5 time steps retain more of the variance than those using 9 time steps. It is particularly the catchments with large observed variances where the estimated time series have less variance than the observed time series. For catchments with small variances, the estimates tend to be slightly too large and there are only minor differences between the different spatio-temporal estimation methods.

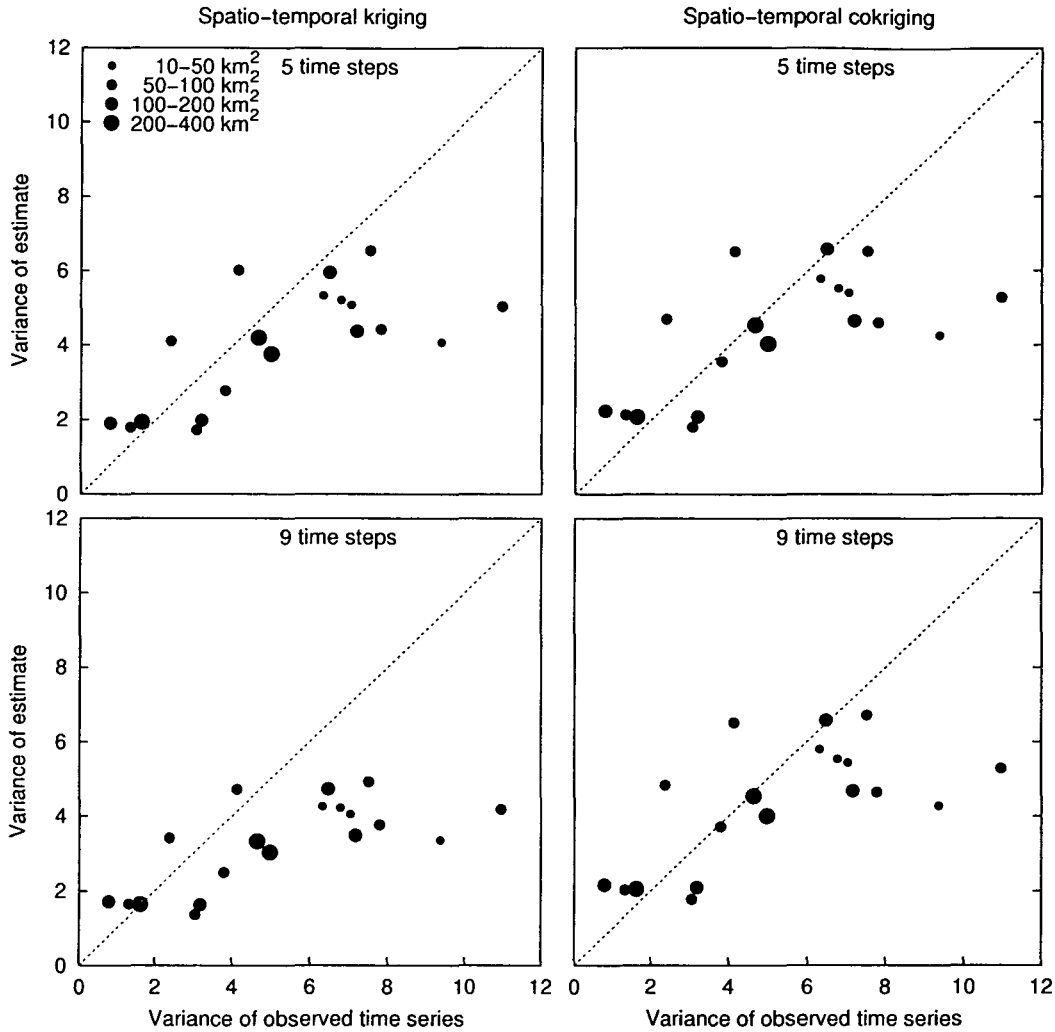


Figure 7.12 Temporal variance of the runoff time series estimated by various Top-kriging variants plotted against the variances of the observed time series. Left panels: Spatio-temporal kriging with 5 (top panel) and 9 (bottom panel) time steps. Right panels: Spatio-temporal cokriging with 5 (top panel) and 9 (bottom panel) time steps. Units are $m^6s^{-2}km^{-4}$ *10^{-4}.

7.4.5 Comparison with a deterministic rainfall-runoff model

The runoff time series estimated by Top-kriging were finally compared with the simulation results of a deterministic rainfall-runoff model obtained by (Parajka et al., 2005). For comparison, the Top-kriging estimates were aggregated to daily values so there is almost no difference between the routing models. The results are hence only presented for spatial kriging with the routing all model. Out of the 19 catchments in the Innviertel, 17 were part of the data set of (Parajka et al., 2005).

The model efficiencies (ME) for Top-kriging and the deterministic model are shown in Table 7.4. The first column gives ME for the deterministic model calibrated *at site*. This is the case

where the parameters have been obtained by fitting the runoff simulations directly to the runoff data, i.e., it is the gauged catchment case. The second column gives the ungauged catchment case of the deterministic model using parameters that have been regionalised from the neighbouring catchments by kriging. This is the regionalisation method of (Parajka et al., 2005) that performed best (see their Table 2, line 6 and Fig. 8 top left green line). The at site calibration efficiency is in the range of 0.58-0.79, with an average of 0.71. When using regionalised parameters, the efficiency decreases to a range of 0.43-0.66, with an average of 0.58. This is quite a significant decrease which is a result of the uncertainty associated with the ungauged catchment problem. The third column of Table 7.4 shows the efficiencies for Top-kriging which, again, is an ungauged catchment case. Note that the Top-kriging efficiencies in Table 7.4 are considerably larger than those in Table 7.3. This is because errors in estimating the time of peak and the dynamics of the runoff hydrograph will be less important for the daily averages of Table 7.4 than for the hourly values of Table 7.3. Top-kriging gives efficiencies in the range of 0.87-0.98 with an average of 0.93. This is remarkably better than the corresponding results of the deterministic model (average of 0.58). In fact, for every single catchment the Top-kriging estimates are much better than the corresponding deterministic estimates of regionalisation. This is not surprising for, say, Taufkirchen where 80% of the catchment area is covered by upstream gauges. However, the much better performance is true for all catchments including those without upstream or downstream neighbours.

Table 7.4 Model efficiencies ME of daily runoff estimated by Top-kriging (spatial kriging, routing all model) and the deterministic runoff model in the Innviertel region. PUB indicates the ungauged catchment case as assessed by cross-validation.

Stream gauge	Deterministic model *at site*	Deterministic model regionalised (PUB)	Top-kriging (PUB)
Ried	0.71	0.56	0.87
Osternach	0.69	0.60	0.88
Pram	0.74	0.50	0.90
Riedau	0.77	0.64	0.94
Winertsham	0.70	0.55	0.96
Taufkirchen	0.73	0.62	0.98
Angsüß	0.73	0.63	0.92
Alfersham	0.71	0.66	0.90
Lohstampf	0.67	0.59	0.92
Still	0.74	0.57	0.94
Strötting	0.75	0.62	0.91
Bad Schallerbach	0.72	0.64	0.96
Pichl	0.58	0.43	0.89
Weghof	0.66	0.52	0.89
Fraham	0.71	0.52	0.97
Niederspaching	0.72	0.59	0.96
Kropfmühle	0.79	0.61	0.94
Average	0.71	0.58	0.93
Median	0.72	0.59	0.92

To illustrate the differences of the two approaches in terms of representing the runoff dynamics Fig. 7.13 shows an example of an event for the Kropfmühle catchment (313 km²). The solid line is the runoff time series estimated by Top-kriging. The dashed and dotted lines are the deterministic model results based on at site and regionalised parameters, respectively. The observed runoff is shown as points. The Top-kriging estimates are very close to the

observed hydrograph for the entire period including the peak of the events and the recession. The deterministic model results tend to underestimate the peak. Interestingly, the first event is better captured by the regionalised model parameters while the second event is better captured by the locally calibrated parameters. It is also of interest that Top-kriging tends to estimate the low flow period better than either of the two deterministic model setups.

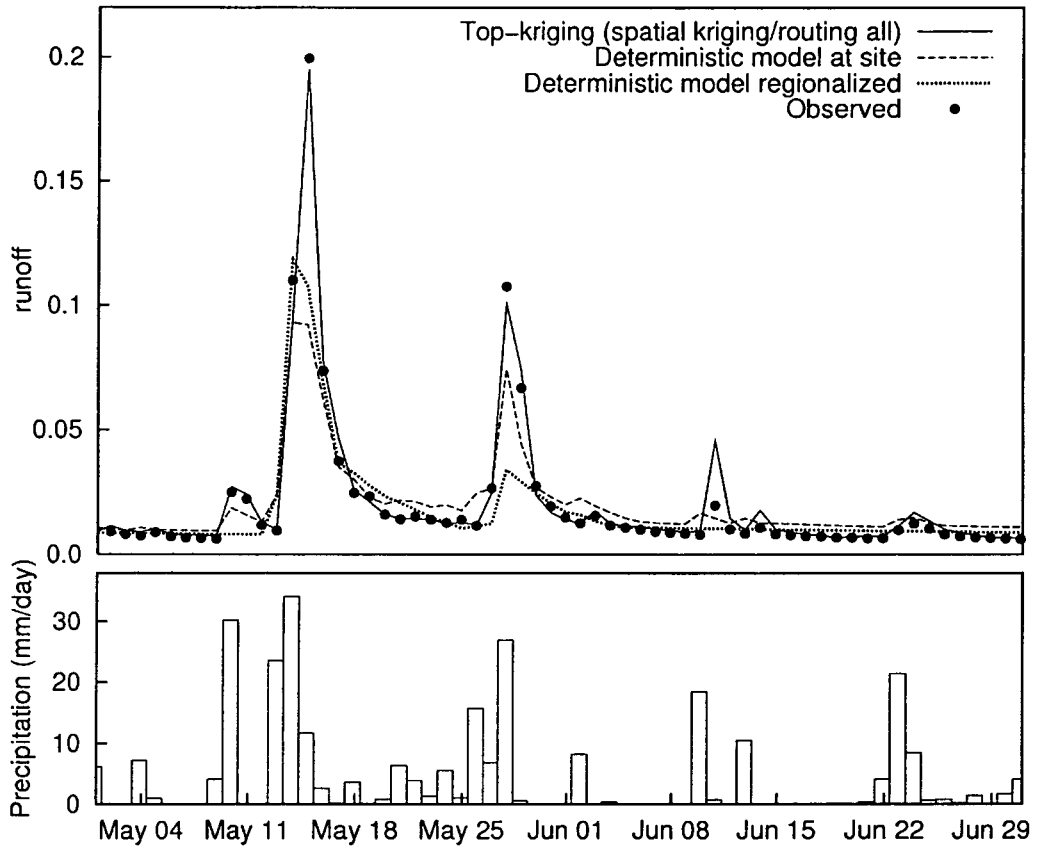


Figure 7.13 Comparison of runoff time series estimated by Top-kriging and the deterministic model for the period of May 1– June 30, 1996 at Kropfmühle (313 km²). The Top-kriging weights of the neighbours are: Niederspaching 0.53, Fraham 0.39, Taufkirchen 0.31, Bad Schallerbach –0.13 and Winertsham –0.09. Units of runoff are $m^3km^{-2}s^{-1}$.

Although the hydrogeology of the Innviertel region is complex, average annual runoff is rather uniform (Table 7.1). It is therefore of interest to put the results of the Innviertel into the context of all of Austria, which is hydrologically much more diverse. For the entire Austrian data set of 367 catchments, the median model efficiency of Top-kriging in terms of estimating hourly runoff time series is 0.75 (Table 7.5). This is lower than the median of 0.87 for the Innviertel region where the stream gauge density is higher. It is also of interest to compare the

results to the deterministic results of Parajka et al. (2005). There were a total of 208 catchments that were common to Parajka et al. (2005) and this chapter. For these catchments, the regionalised deterministic model gives a median efficiency is 0.67 while it is 0.87 for Top-kriging, both based on daily runoff time series for the ungauged catchment case (Table 5). This means that Top-kriging performs much better than the deterministic model. To illustrate the between-catchment variability of model performance, Fig. 7.14 shows the cumulative distribution function (cdf) of the model efficiencies for the deterministic model (both *at site* calibration and regionalised) and Top-kriging. The figure, again, indicates that Top-kriging indeed outperforms the regionalised deterministic model substantially. There are a small number of catchments where Top-kriging gives efficiencies less than zero. Examining these catchments suggested that a number of factors contribute to the poor performance. The majority of these catchments were close to the border of Austria and/or spatially isolated with a large distance to the nearest stream gauge. Some of the catchments, however, did have close neighbours but the hydrogeology differed significantly from that of the neighbouring catchments. A typical example in Austria are gravel deposits in some valleys that increase the groundwater component of runoff and produce much slower response as compared to neighbouring catchments without such gravel deposits.

While the Top-kriging cdf in Fig. 7.14 is always above that of the regionalised deterministic model, the efficiencies do not necessarily relate to the same catchments. In Fig. 7.15, the Top-kriging efficiencies have been plotted against those of the regionalised deterministic model. The comparison indicates that there are indeed catchments were Top-kriging performs poorer than the deterministic model. As indicated above, these are mostly catchments that do not have a stream gauge in their vicinity. However, it is only 15% of the catchments where this is the case. For the remaining catchments Top-kriging performs better.

Table 7.5 Median model efficiencies for Topkriging as compared to the deterministic runoff model for the ungauged catchment case. For 208 catchments, both Top-kriging and deterministic model results were available, 17 of these are in the Innviertel region.

Region	Number of catchments	Top-kriging hourly runoff	Top-kriging daily runoff	Regionalised deterministic model, daily runoff
Innviertel	19	0.87	0.92	
Innviertel	17	0.87	0.92	0.59
Austria	376	0.75	0.82	
Austria	320			0.68
Austria	208	0.82	0.87	0.67

To identify locations with potentially poor estimates, geostatistical analysis generally makes use of the kriging variance. This is also possible for the Top-kriging approach (Chapter 6). For spatial kriging (Eqs. 7.2-7.3), the kriging variance is

$$\sigma_{Ri}^2 = \sum_{j=1}^{n} \lambda_j \gamma_{ij} + \mu \qquad (7.23)$$

In this chapter the kriging variance σ_{Ri}^2 was estimated in a cross-validation mode, i.e., in solving the kriging system to obtain the set of weights λ_j for a given catchment i it was assumed that it is ungauged and only runoff data for the neighbouring catchments j were available. This means that σ_{Ri}^2 should be an indicator of the uncertainty of estimating runoff in ungauged catchments. To examine this idea, Fig. 7.16 shows the Top-kriging efficiencies of 376 catchments plotted against the kriging variance. The model efficiencies are indeed well correlated with the kriging variance indicating that the kriging variance can be used as an indicator to identifying catchments with potentially poor estimates. For example, catchments with kriging variances less than $4*10^{-5}$ $m^6s^{-2}km^{-4}$ are associated with median model efficiencies of hourly time series of 0.9. If the kriging variance is more than $12*10^{-5}$ $m^6s^{-2}km^{-4}$, median model efficiencies of 0.5 can be expected.

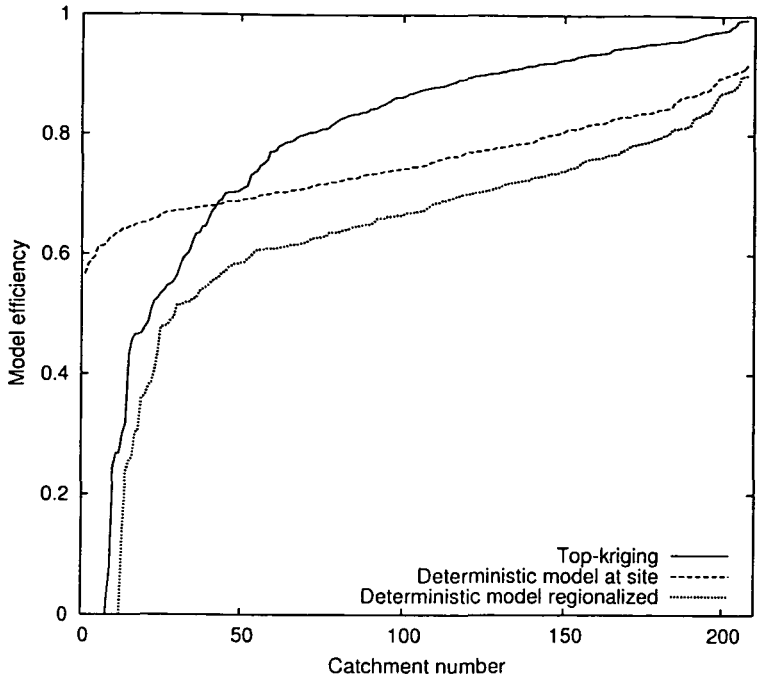


Figure 7.14 Cumulative distribution functions of model efficiencies ME of daily runoff estimated by Top-kriging (spatial kriging, routing all model) and the deterministic runoff model for 208 catchments in Austria. Top-kriging and the regionalised deterministic model relate to the ungauged catchment case as assessed by cross-validation.

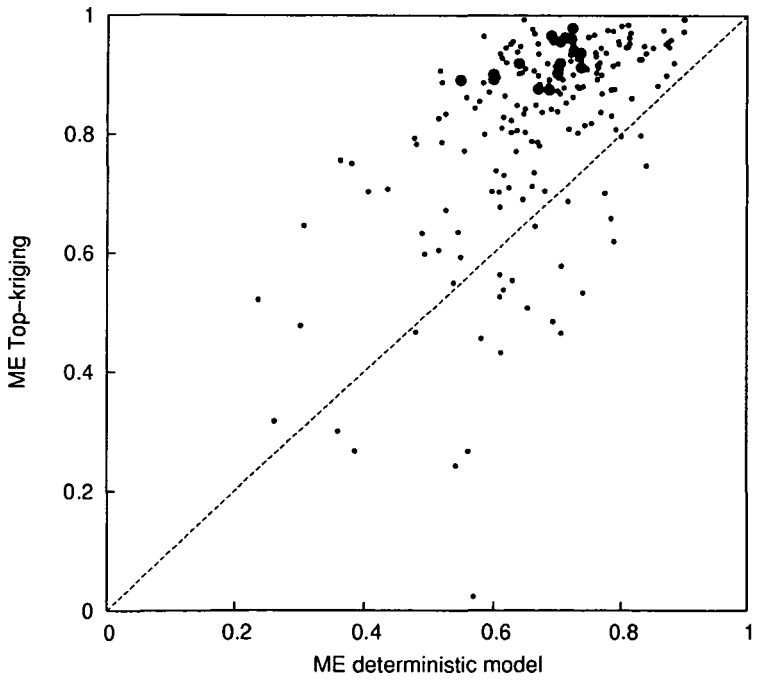


Figure 7.15 Model efficiencies ME of daily runoff estimated by Top-kriging (spatial kriging, routing all model) and the regionalised deterministic runoff model for 208 catchments in Austria. Large circles indicate the Innviertel catchments. All model efficiencies relate to the ungauged catchment case as assessed by cross-validation.

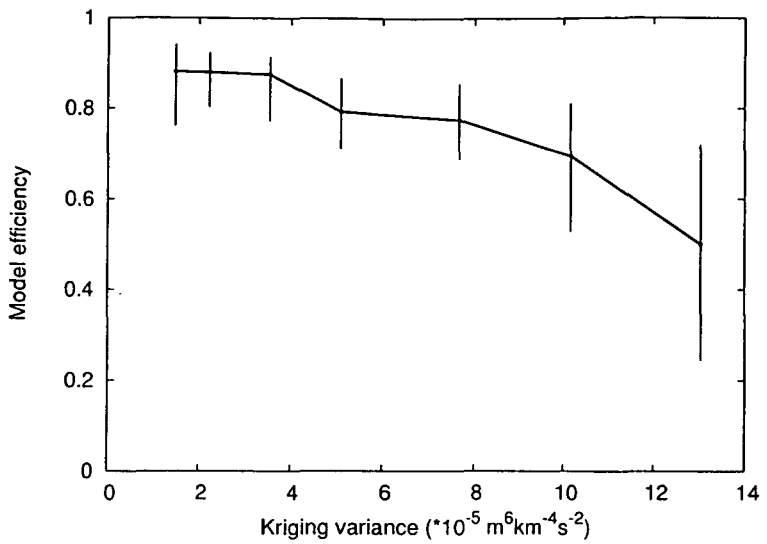


Figure 7.16 Model efficiencies ME of hourly runoff estimated by Top-kriging (spatial kriging, routing all model) plotted as a function of the kriging variance (Eq. 7.23) for 376 catchments in Austria. The thick line represents the median ME and the error bars represent the 25 and 75 percentiles. Both model efficiencies and the kriging variance relate to the ungauged catchment case obtained in a cross-validation mode.

7.5 Discussion and conclusions

The Top-kriging approach from Chapter 6 was in this chapter extended to account for hydrodynamic and geomorphologic dispersion as well as routing. The main appeal of the method is that it is a best linear unbiased estimator adapted for the case of stream networks. It can therefore be claimed that it is the most natural way of statistically estimating runoff time series in ungauged catchments.

Examination of the Top-kriging results in the Innviertel region suggests that the kriging weights are plausible. In particular, they take into account the upstream and downstream similarities of catchments as well as similarities with catchments that do not share a subcatchment. They also account for clustering effects, giving less weight to clusters of catchments that consist of catchments with highly correlated runoff records.

7.5.1 Top-kriging performance

The cross-validation analyses in the Innviertel region indicate that the Nash-Sutcliffe model efficiency of the spatial kriging variant of Top-kriging is on the order of 0.8-0.95 for hourly runoff time series and 0.85-0.98 for daily runoff time series. The efficiency is slightly lower

for the rest of Austria with a median of 0.82 for hourly runoff and 0.87 for daily runoff for a set of 208 catchments, and 0.75 for hourly runoff and 0.82 for daily runoff for a set of 376 catchments. These efficiencies are excellent as compared to what can usually be obtained by deterministic rainfall-runoff models in ungauged catchments. For the same 208 catchments, the deterministic runoff model of Parajka et al. (2005) gave median efficiencies of daily runoff ME=0.67 if the catchments were treated as ungauged. This means that the Top-kriging efficiency is 0.20 higher than the corresponding efficiency of a deterministic runoff model.

These efficiencies also compare well with other studies around the world. A review of Merz et al. (2006) suggests that, typically, the runoff modelling efficiency of daily runoff in ungauged catchments ranges between 0.6 and 0.7, depending on data availability. If the focus is on simulating runoff with hourly temporal resolution, the efficiency is usually lower, mainly because of poorer rainfall information and errors in the timing of runoff dynamics. An example is the DMIP distributed model intercomparison project (Reed et al., 2004). In the DMIP project, 12 deterministic runoff models were used to simulate hourly runoff for 8 catchments (65 km^2-2484 km^2) in the USA for a period of 7 years (1994-2000). The models were run both in an *at site* mode where local runoff data were used to calibrate model parameters and in an uncalibrated mode that represents the ungauged catchment case. In the *at site* (calibrated parameter case) the median of the hourly model efficiencies over the eight catchments ranged between 0.30 and 0.73, depending on the model with a mean of 0.58. The efficiencies for the uncalibrated case was significantly lower with the median over the eight catchments ranging between –0.50 and 0.61, depending on the model, with a mean of 0.22. The median efficiencies obtained by Top-kriging(0.82 and 0.75 for the 208 and 376 catchment data sets, respectively) is much higher thanthat of the best model of that study. There may exist differences in the efficiency that are due to differences in the climate but the relatively low DMIP efficiencies are typical of the general difficulty of estimating runoff in ungauged catchments. What sets Top-kriging apart from deterministic models such as those of Parajka et al. (2005) and Reed et al. (2004) is that Top-kriging does use concurrent runoff data of neighbouring catchments. This means that, for some of the applications envisaged in Parajka et al. (2005) and Reed et al. (2004), such as flood forecasting, Top-kriging is not applicable. However, for those applications where concurrent runoff data are available, the better performance of Top-kriging as compared to the traditional deterministic modelling approach can be a real advantage.

Another indicator of model performance are the statistical moments of temporal runoff variability. Capturing the temporal variability well is of particular importance for assessing

the hydropower potential of streams and for ecological studies. The analyses of the Innviertel region indicate that, overall, Top-kriging does represent the moments well with little bias. However, the temporal variances of the catchments with the largest observed temporal variances were underestimated and the temporal variance of the catchments with the smallest observed temporal variances were overestimated. This means that the spatial variance of the temporal variance was underestimated. This would be expected as kriging, generally, does not retain the spatial variance, being a best estimator. Rather, kriging introduces smoothing as it minimises the error variance. Methods such as stochastic simulation do retain the spatial variance but the errors of such methods are larger (e.g. Deutsch and Journel, 1992). A similar effect of spatial smoothing was found by Gottschalk et al. (2006, Fig. 6, 1 hour) who estimated the first two moments of runoff directly from neighbouring catchments by geostatistical techniques for 17 catchments in France.

7.5.2 Factors controlling performance

For deterministic models, obviously, the uncertainty in rainfall is one of the main factors contributing to simulation uncertainty, particularly for hourly models (Faures et al., 1995). Other uncertainties are problems with parameter identifiability and the representativeness of catchment characteristics such as soils data (Blöschl, 2005). The factors controlling the uncertainties of Top-kriging are quite different. The main limitations are the assumption of spatial homogeneity of runoff, the spatial distribution of stream gauges and the quality of the runoff measurements. Much of the spatial and temporal variability of the atmospheric forcing is filtered by the catchments. Both the runoff of the target catchment and the runoff at the neighbouring stream gauges therefore represent aggregated values that have much larger spatial and temporal correlations than the atmospheric input (Chapters 4 and 5). With this in mind, it is clear that the Innviertel region with a high density of stream gauges is a favourable case for the Top-kriging approach. For the deterministic model, the Innviertel region was on the other hand a region with poorer results than for the average of Austria. This can be attributed to two reasons. First, the response times in this region are relatively small, which makes it more difficult for a daily runoff model to simulate runoff well (see Figs. 7.6-7.10). Second, although the mean runoff is relatively homogenous in the Innviertel region, the dynamics reflect a rather heterogeneous hydrogeology which may contribute to parameter identifiability issues and regionalisation uncertainty. This is also indicated by the decrease in

median model performance from 0.72 to 0.59 in the Innviertel region when moving from gauged to ungauged catchments which is substantial.

In other parts of Austria, stream gauge density is lower and there are, in particular, a number of stream gauges located at the Austrian border with no close neighbours. If one moves from the Innviertel to all of Austria, the Top-kriging performance hence decreases from a median efficiency of 0.87 to 0.75 (hourly data). In addition to the lower stream gauge density, the variogram may contribute to the decrease in performance as it was estimated for the Innviertel and used for all of Austria for consistency. Also, the quality of the Innviertel data set may be above average. The data set of 208 catchments used by Parajka et al. (2005) gives a better Top-kriging performance (ME=0.82 for hourly data) than the entire Austrian data set because Parajka et al. (2005) have been more restrictive in selecting high quality runoff records.

It was shown that the kriging variance (Fig. 7.16) is an indicator of expected model performance. The kriging variance takes into account the catchment shape, stream network organisation, routing characteristics and, most importantly, the location of the stream gauges including the effects of catchment nesting (Chapter 6). However, kriging variance does not take into account local particularities in the hydrogeology. Because of this, for some catchments, the efficiency was considerably poorer than what could be expected from the kriging variance. To capture these effects, maps of the hydrogeology would have to be consulted or, preferably, reconnaissance field trips undertaken to assess the runoff dynamics in ungauged catchments (Blöschl, 2005).

7.5.3 Top-kriging variants

A number of variants of Top-kriging were tested to examine what variant would produce the most realistic runoff hydrographs. The results indicated that inclusion of a routing model almost always improves estimation of the runoff time series. Figure 7.9 shows a typical example. The difference between the two routing models was small, both in Fig. 7.9 and more generally in the Innviertel (Table 7.3). This is related to the fact that the non-nested neighbours usually have smaller weights and, therefore, the contribution to the estimated hydrograph is usually small. This means that use of a routing model, albeit such a simple linear model as used in this chapter, is an advantage but, in general, the assumptions made about the non-nested catchments may not be very important.

Spatio-temporal kriging and cokriging variants of Top-kriging were also tested. From a conceptual point of view spatio-temporal kriging and cokriging are more complete than

spatial kriging as they account for the temporal filtering (i.e. the unit hydrograph) effects of catchments in the estimation equation (Eq. 7.4) while spatial kriging estimates runoff from concurrent time steps only. It should be noted that spatial kriging does account for the temporal filtering effects on the regularised variogram, i.e., the unit hydrograph enters the estimation indirectly through the gamma values. The analyses indicate that spatio-temporal kriging and spatio-temporal cokriging, in some cases, improved the results over spatial kriging, but this was not generally the case. The median model efficiency of spatio-temporal cokriging was similar to that of spatial kriging but there was a clear trend of underestimating the temporal variance of runoff. Spatio-temporal kriging underestimated the temporal variance more and the underestimation increased with the number of time steps used in the estimation. Clearly, this is related to the smoothing effect of kriging. The better performance of spatio-temporal cokriging as compared to spatio-temporal kriging is a result of the additional unbiasedness constraints. Given that spatio-temporal kriging and spatio-temporal cokriging increases the computational burden significantly over spatial kriging, the additional complexity does not seem to be warranted in the light of the results of this chapter. This is in line with the results of Goovaerts et al. (2006) who suggested that, in their case, spatio-temporal kriging did not generally improve the estimation results over spatial kriging.

7.5.4 Top-kriging assumptions

Top-kriging as presented in this chapter involves a number of assumptions. Some of them were made for clarity of presentation and can be easily relaxed if suitable data are available. In this chapter it was assumed that the temporal support (Eq. 7.13) is a function of catchment area only but it can be easily made a function of physiographic catchment characteristics. Similarly, the simple routing model based on a constant velocity used here (Eqs. 7.16 and 7.18) was mainly chosen for clarity. Also, Eq. 7.17 can be relaxed by introducing more complex relationships between lag and catchment characteristics. Specific runoff was assumed to be a random field. It would be possibly to include a temporal trend model (Montanari, 2005a) or a spatial trend model and apply Top-kriging to the residuals of the trend model. A spatial trend model could, for example, be based on mean annual rainfall, to account for differences in runoff in ungauged catchments as a result of climate. In fact, one could also use a deterministic runoff model as a trend model and apply Top-kriging as an error model to improve runoff estimates over those of the deterministic runoff model (Goovaerts et al., 2006).

An assumption that would be more difficult to relax, is the assumption of the variogram not changing with time. It is likely that low flows are correlated over larger distances in both space and time than flood flows, so one would expect temporal differences in the space-time variogram of runoff. It is in this chapter assumed that the variograms of runoff time series will be dominated by the variability of high flows, as the variogram is the second moment of runoff. Changing the variogram with time would have to involve additional assumptions about temporal aggregation. Catchments were conceptualised to operate as a linear space-time filter, which is at the heart of Top-kriging. Introducing non-linearity would in effect modify Top-kriging to become a spatially distributed runoff model.

Relaxing assumptions will generally require additional data and add complexity, which may not be warranted as suggested by the comparison of the spatial and spatio-temporal kriging variants of Top-kriging. Also, the high model efficiencies obtained in this chapter indicate that the assumptions on which the spatial variant of Top-kriging is based may be highly appropriate for practical purposes of estimating runoff time series in ungauged catchments.

7.5.5 Potential applications of Top-kriging

Top-kriging uses concurrent runoff data in neighbouring catchments and does not explicitly account for rainfall-runoff processes. This means that there are a number of applications, such as flood forecasting and assessing the effects of land use change on runoff, for which Top-kriging is not the right method and deterministic runoff models are needed. However, there is a range of applications where Top-kriging could be applied to estimate runoff time series in ungauged catchments. The main applications in a simulation mode are estimating the flow duration curve for assessing the hydropower potential of river reaches, estimating temporal flow (and low flow) variability in the context of environmental flow requirements, and producing high resolution maps of runoff variability. These maps could not only involve mean annual runoff, as in the traditional runoff mapping approaches, but also high resolution characteristics such as temporal variance and runoff regimes. Another simulation application is the generation of hydrographs for ungauged catchments to which deterministic rainfall runoff models are calibrated. This may reduce some of the problems with regionalising model parameters (Parajka et al., 2006). There are also potential on-line applications. One of them is near real time monitoring and visualisation of runoff. In Top-kriging, much of the effort is in estimating the kriging weights. Once the weights are known, estimation of runoff for all river

reaches in a region from runoff data is a straightforward linear operation, ideally suited for on-line applications.

The relatively high performance of Top-kriging found in this chapter is because errors of rainfall data and parameter identifiability issues of traditional runoff models are avoided. The model efficiencies obtained by Top-kriging are considerably higher than what can be typically achieved by using a regionalised deterministic model. This means that, for certain applications, Top-kriging is an appealing method for estimating runoff in ungauged catchments. However, a prerequisite of using Top-kriging is a relatively dense stream gauging network. In the Innviertel and Austrian examples, a total of 19 stream gauges over an area of 1500km² and 376 stream gauges over an area of 80 000km², respectively, were available. If the stream gauging network is less dense, the benefits of Top-kriging would probably be less obvious. However, given that the improvement over deterministic models is quite significant, Top-kriging may still be the method of choice.

Chapter 8. Summary of results and conclusions

The aim of this thesis was to analyse the variability and scale effects in geostatistical analyses in hydrology, and to take these effects into account for estimation purposes. This was done both from a general point of view, starting with the theories, and for the more specific case of stream flow variables. This was done in a series of ways, from analyses of synthetically generated data to estimates of observed hydrographs.

8.1 Measurement scale effects

Chapters 2 and 3 covered the theoretical part of sampling scale effects. Through the generation of random fields and different sampling strategies, it was possible to show the effect of the sampling scale triplet (Blöschl and Sivapalan, 1995) on both non-parametric (Chapter 2) and parametric (Chapter 3) estimates of the correlation structure. A large number of two dimensional random fields were generated with a known underlying variance and correlation length. Hypothetical samples were drawn from these random fields, conforming to a certain sampling scale triplet. These samples were then used to estimate the sample mean, spatial variance and integral scale (Chapter 2) and the parameters of a variogram (Chapter 3), and the biases in comparison with the underlying random field were analysed.

The results showed that if the sampling scale is not commensurate with the underlying random field, estimates will be strongly biased by the correlation structure. Estimates of integral scale or correlation length will be overestimated if the spacing between samples is too large, it will be underestimated if the extent of the examined area is too small. Estimates of sill or variance will be underestimated if the extent of the examined area is too small. The sill or variance will be underestimated and the integral scale or correlation length will be overestimated if the samples have a large support. A joint effect can also be found in many cases. If the number of samples is low (e.g. 16 samples), the estimated integral scale is a pure function of the sampling scale. A larger number of samples do not reduce the bias of variance and integral scale if the extent is too short, but it will help to reduce the variance of the estimate.

For the parametric estimation of the variogram, two different estimation methods were compared. The Maximum Likelihood (ML) method of parameter estimation gives the smallest biases for largely spaced random samples, while the Weighted Least Squares (WLS) method gives the smallest biases for largely spaced gridded samples. Non-parametric estimates exhibit smaller random errors but larger biases than estimates from the two parametric methods (ML, WLS). The results were also compared to analytical expressions for the bias of the mean, spatial variance and integral scale as a function of the scale triplet taken from Western and Blöschl (1999), and to analytical expressions for the uncertainty of the same variables. The biases and the uncertainties of the non-parametric estimates agreed closely to the analytical expressions.

All these results are important for geostatistical hydrological analyses. Gelhar (1993), Blöschl (1999) and Western and Blöschl (1999) all observed how the sampling scale influences the analyses of observations. The results of Chapters 2 and 3 give a more coherent description of the qualitative and quantitative biases due to these sampling scale effects. Suggestions are given on how to adjust a monitoring network to the scales of the variables of interest and how to interpret sampling scale effects in environmental data. The preferable situation is to have a monitoring network where the spacing is shorter than the correlation length, and the extent larger than the correlation length. If this is not possible, the chapters suggests how a researcher can back-calculate the underlying correlation structure.

8.2 Characteristic scales

In Chapter 4, some of the results from Chapter 2 and 3 were of importance for analyses of real data. Spatial and temporal variograms of precipitation, runoff, and groundwater levels in Austria were examined in the search for eventual characteristic scales exist, and if they existed, their size. The characteristic scales were then interpreted also in the context of the measurement scale triplet, which for this study the most relevant were the catchments sizes and the extent of Austria in space, and the temporal resolution in time. In time, precipitation and runoff are stationary with characteristic scales on the order of a day and a month, respectively, while groundwater levels are non-stationary. As the estimated characteristic temporal scale of precipitation is only slightly larger than the temporal resolution of the precipitation measurements (1 day), the results from Chapter 2 and 3 indicate that this characteristic scale is overestimated. In space, point precipitation is almost fractal, so no characteristic scales exist. Runoff is non-stationary but also not a fractal as it exhibits a break

in the variograms. A similar break was observed in the variograms for catchment precipitation, and an analysis of these indicates that this break is due to aggregation effects imposed by the catchment area. Catchment runoff does on the other hand exhibit stronger aggregation behaviour than catchment precipitation. A spatial variogram of hypothetical point runoff back-calculated from runoff variograms of three catchment size classes using aggregation statistics (regularisation) is almost stationary and exhibits shorter characteristic space scales than catchment runoff. Although this back-calculation was mathematically possible, the results indicated the need for the spatio-temporal aggregation theories developed in Chapter 5. Groundwater levels are non-stationary in space exhibiting shorter scale variability than precipitation and runoff, but are also not fractal, as there is also here a break in the variogram.

It is in this chapter suggested that the decrease of spatial characteristic scales from catchment precipitation to runoff and to groundwater is the result of a superposition of small scale variability of catchment and aquifer properties on the atmospheric forcing. For comparison, TDR soil moisture data from a comprehensive Australian data set were examined. These data suggest that, in time, soil moisture is close to stationary with characteristic scales on the order of two weeks while in space soil moisture is non-stationary and close to fractal over the extent sampled. Space-time traces of characteristic scales fit well into a conceptual diagram of Blöschl and Sivapalan (1995). The scaling exponents z in $T \sim L^z$ (where T is time and L is space) are on the order of 0.5 for precipitation, 0.8 for runoff from small catchments, 1.2 for runoff from large catchments, 0.8 for groundwater levels and 0.5 for soil moisture.

8.3 Catchments as space-time filters

Chapter 5 continues the analyses of precipitation and runoff, but now as the outcome of joint spatio-temporal random processes. It is suggested that catchments can be conceptualised as linear space-time filters. Catchment area A is interpreted as the spatial support and the catchment response time T is interpreted as the temporal support of the runoff measurements. These two supports are related by $T = \mu A^{\kappa}$ which embodies the space-time connections of the rainfall-runoff process from a geostatistical perspective. To test the framework, spatio-temporal sample variograms were estimated for three different catchments size classes from about 30 years of quarter hourly precipitation and runoff data from about 500 catchments in Austria. Also a spatio-temporal sample variogram of precipitation was inferred. Similar to the results of Chapter 4, a stronger aggregation effect could be observed for runoff than for

precipitation. As the analyses in this chapter was performed on quarter hourly data, the temporal characteristic scales precipitation were found to be shorter than in Chapter 4. In a first step, spatio-temporal variogram models were fitted to the sample variograms for the three catchment size classes independently. Three different spatio-temporal variogram models were tested, an exponential model, Cressie-Huang model (1999), product-sum model (De Cesare et al. 2001; De Iaco et al. 2001), all of them with a non-stationary part added, in addition to a pure fractal model. All models except for the fractal model could well be fitted to the observed sample variograms.

In a second step, variograms are fitted to all three catchment size classes jointly by estimating the parameters of a point/instantaneous spatio-temporal variogram model and aggregating (regularising) it to the spatial and temporal scales of the catchments. Again, all models except from the fractal model give good fits to the sample variograms of runoff with dimensionless errors ranging from 0.02 to 0.03, and the model parameters are plausible. This indicates that the first order effects of the spatio-temporal variability of runoff are indeed captured by conceptualising catchments as linear space-time filters. The scaling exponent κ is found to vary between 0.3 and 0.4 for different variogram models. With a μ value of around 2, this range of κ gives response times of about 5 hours and one day for catchments of 10 and 1000 km² area, respectively. The estimated range of κ represents the average scaling characteristics of catchments within the study region. Obviously, for an individual catchment, the response time may deviate significantly from the general relationship of Eq. 5.15. Fig. 3 of Merz and Blöschl (2003), for example, shows a map of the deviations from such a general relationship based on an exponent of $\kappa = 0.35$ for Austria. It should be noted that, if known, the deviations of catchment response times from a general relationship could be included in the space-time filter framework proposed here.

It is interesting that the product-sum model almost reduces to a separable model (spatio-temporal correlation C_{st} is a function of the pure spatial correlation C_s and the pure temporal correlation C_t, i.e. $C_{st} = C_s C_t$) with the fitted parameters, i.e. $a_{3s} = a_{3t} = 0.0070$ and $k = 142$, which is very close to the condition for the product-sum model to reduce to the separable model of Rodríguez-Iturbe and Mejía (1974). Separable models are computationally more convenient for some applications (Fuentes, 2006). The spatial variogram fits in this chapter (Fig. 5.10 right column) are as good as or better than those in Chapter 4 (Fig. 4.6 b) where only spatial aggregation was applied. In addition, these variograms describe the temporal aggregation effects well (Fig. 5.10 left column).

8.4 Top-kriging – geostatistics on stream networks

Chapter 6 presented the top-kriging approach, or topological kriging, as a method for estimating streamflow-related variables in ungauged catchments. It takes both the area and the nested nature of catchments into account, and hence uses a spatial filtering principle for the gamma values in the kriging equations. The main appeal of the method is that it is a best linear unbiased estimator (BLUE) adapted for the case of stream networks without any additional assumptions. The concept is built on the work of Sauquet et al. (2000) and extends it in a number of ways. The method is tested for the case of the specific 100-year flood for two Austrian regions. The method provides more plausible and, indeed, more accurate estimates than ordinary kriging. In the example of the Mur region it was showed how the estimates of the tributaries are improved over ordinary kriging. In the example of the Inn region it was showed how the estimates on the main stream are improved. In general, both the estimates on the main stream and the tributaries will be superior to Ordinary Kriging and the relative magnitude of the improvement will depend on the location of the gauges.

In essence, Top-Kriging treats upstream and downstream catchments differently from neighbouring catchments that do not share a subcatchment. This is also what one would do in manual interpolation. Typically, the effect of nesting will differ for upstream and downstream catchments. Larger catchments are generally given larger weights than smaller catchments (Fig. 6.3) and they are also likely to be well correlated with other large neighbours. It is therefore more likely that the nesting will have more impact on the weights of an upstream catchment than on those of a downstream catchment.

For the variable of interest, Top-kriging also provides estimates of the uncertainty. On the main stream the estimated uncertainties are smallest and they gradually increase as one moves towards the headwaters. The IAHS Decade on Predictions in Ungauged Basins (PUB) (Sivapalan et al., 2003) has predictive uncertainty of hydrological variables as its main focus. This is exactly what Top-kriging provides for the most natural case of best linear unbiased estimators. The method as presented here is able to exploit the information contained in short records by accounting for the uncertainty of each gauge.

Top-kriging assumes linear aggregation, as it is a linear estimator. This means that, strictly speaking, the method only applies to variables that are mass conserving over nested catchments. It is still suggested that the method can also be profitably used, as an approximation, for variables that do not aggregate linearly but show a degree of averaging.

8.5 Top-kriging – Estimation of runoff time series

The Top-kriging approach is further developed in Chapter 7, where it is combined with the spatio-temporal filtering assumption presented in Chapter 5. By taking this spatio-temporal filtering effects into account, it is possible to estimate runoff in one catchment as a weighted average of the observed runoff in neighbouring catchments, both nested and non-nested. The assumption was first tested for 19 catchments in the Innviertel region in Austria, then for a larger data set of all Austria. Different variants of the Top-kriging approach were tested. The simplest was to consider runoff to consist of a set of spatially correlated time series. Each time step was then estimated as the weighted average of the runoff of neighbouring stations. To take advantage of the information from more than one time step, spatio-temporal kriging and spatio-temporal cokriging were attempted. A simple routing model was also suggested, to model the advection between stream gauges.

A cross-validation test was performed, where each catchment was assumed to be ungauged, and the time series for a period of 10 years estimated from the neighbouring catchments. The Nash-Sutcliffe efficiency was used as a measure of the model efficiency (ME). From analyses of ME, spatial kriging performed better than spatio-temporal kriging and spatio-temporal cokriging. The results also indicated that it is preferable to include a simple routing model. ME for the catchments in the Innviertel region was relatively high, in the range 0.72-0.96 for hourly runoff. Inspections of the average runoff and the temporal variance of the estimated time series revealed that the method respected the average of the observed time series, but that the variances of the catchments with the largest observed temporal variances were reduced in comparison to the observations. A similar effect can be found in the study by Gottschalk et al. (2006, Fig. 6, 1 hour). Several visual comparisons of the observed and estimated time series also indicated that the estimated time series reflect the true dynamics.

Model efficiencies from the Top-kriging approach were also compared with ME from a deterministic model (Parajka et al., 2005). This comparison indicated that the top-kriging approach is can well be used for regionalization purposes. For the Innviertel area, where top-kriging performed particularly well and the deterministic model slightly poorer than for the rest of Austria, the median ME was about 0.2 higher than for the at site calibrated deterministic model and 0.3 better than the regionalized deterministic model.

There are different factors that control the performance of the top-kriging approach and the deterministic model. For the deterministic model, it is particularly the uncertainty of the input variables that constitutes a limiting factor for the performance. Particularly precipitation

cannot be measured with the necessary spatial resolution. There is also a range of other uncertainties that will accumulate to the precipitation uncertainties, including parameter estimation and model uncertainties. The uncertainties of the Top-kriging approach is on the other hand limited to the assumption of spatial homogeneity of runoff, the spatial distribution of stream gauges and the quality of the measurements. Much of the spatial and temporal variability of the atmospheric forcing has been filtered by the catchments. Both the runoff of our target catchment and the runoff at the neighbouring stream gauges therefore represent aggregated values that have a much larger spatial and temporal correlation than the atmospheric input. They can also be measured with a much higher certainty.

The kriging variance (Fig. 7.16) is an indicator of expected model performance. A closer description of effect of Top-kriging on the kriging variance is given by Skøien et al. (2006). However, for some catchments ME was considerably poorer than what could be expected from the stream gauge density. It might then be a question about the degree of homogeneity of the data series. It can partly be that neighbouring catchments have totally different hydrogeology, and there are no upstream or downstream neighbours to improve the model efficiency. It can also be a question about the quality of the data during extreme events, whether the rating curve is reliable. It is more complicated to assure the degree of homogeneity for runoff time series as it includes both the moments and the dynamics of the runoff. On the other hand, it is not uncommon that there are regions in a random field where the homogeneity assumptions are not fulfilled.

8.6 Application of results

The ultimate aim of this thesis was to develop methods for estimation of high temporal resolution time series. However, the studies presented in the earlier chapters have possible applications that go beyond the methods developed in the last chapter. Chapters 2 and 3 analysed scaling effects of measurement scale. In addition to the impact these analyses had on the later chapters, the results are important for many environmental disciplines. Measurements of precipitation (from gauges or radar), soil moisture (from TDR or satellites), hydraulic conductivities (from pump tests or soil samples) and many other environmental variables will all be influenced by measurement scale effects. The analyses presented in Chapters 2 and 3 will help researchers in optimising the sampling network, in choosing the right method for estimating correlation scale for their problem, and in interpreting their results. The characteristic scales found in Chapter 4 can be used for different hydrological applications,

including back-of-an-envelope calculations, for simplifying equations and for planning measurements. Hydrology is still at a very early stage as compared to other disciplines in taking these characteristic scales into use, and hopefully these analyses will contribute to their wider use in a similar way as for example in fluid dynamics.

The three last chapters mostly focus on how catchment area and network structure can be taken into account. Some possible applications in a simulation mode are estimating the flow duration curve for assessing the hydropower potential of river reaches, estimating temporal flow (and low flow) variability in the context of environmental flow requirements, producing high resolution maps of runoff variability, and producing maps for floods with different return periods. With some modifications the methods developed here can be transferred to many other disciplines where spatial or spatio-temporal estimates are needed from observations with a spatial or spatio-temporal support, such as areal planning and biology.

8.7 Outlook

There are several possible extensions to the work that has been presented in this thesis, particularly to the Chapters 5 and 7. As geostatistical estimation theory is generally developed for estimating residuals, it can also improve the results of the different top-kriging approaches if the top-kriging approach is rather conducted on the residuals of a trend model. Such a trend model can be a result of regression between the variable of interest with some other variable, the outcome of a rainfall-runoff model or a spatial or temporal trend model corresponding to mean annual runoff, seasonal changes or other information about the processes.

Adamowski, K., and C. Bocci (2001) Geostatistical regional trend detection in river flow data, Hydrological Processes, *15* (18), 3331-3341.

Ai-Geostats (2004) Discussion on "Automated variogram modelling" [Mailing-list], msg 1524-1550, Available at: http://groups.yahoo.com/group/new_ai_geostats/messages, (Verified 1. Nov. 2006).

Ambroise, B. (1999) *La dynamique du cycle de l'eau dans un bassin versant.*, 200 pp., Editions HGA, Bucharest, Romania.

Anderson, M. G., and T. P. Burt (1990) Subsurface runoff, in *Process Studies in Hillslope Hydrology*, edited by M. G. Anderson, and T. P. Burt, pp. 365-400, Wiley, Chichester.

Anderson, M. P. (1997) Characterization of geological heterogeneity, in *Subsurface flow and transport: a stochastic approach*, edited by G. Dagan, and S. P. Neuman, pp. 23-43, University Press, Cambridge.

Austin, P. M., and R. A. Houze (1972) Analysis of the structure of precipitation patterns in New England, Journal of Applied Meteorology, *11*, 926-935.

Beckie, R. (1996) Measurement scale, network sampling scale, and groundwater model parameters, Water Resources Research, *32* (1), 65-76.

Bellehumeur, C., and P. Legendre (1998) Multiscale sources of variation in ecological variables: modeling spatial dispersion, elaborating sampling designs, Landscape Ecology, *13* (1), 15-25.

Bierkens, M. F. P., P. A. Finke, and P. de Willigen (2000) *Upscaling and downscaling methods for environmental research*, 190 pp., Kluwer Academic Press, Dordrecht.

Blöschl, G. (1996) *Scale and scaling in hydrology*, Wiener Mitteilungen Wasser, Abwasser, Gewässer, 132, 346 pp., Vienna University of Technology, Vienna, Austria.

Blöschl, G. (1999) Scaling issues in snow hydrology, Hydrological Processes, *13* (14-15), 2149-2175.

Blöschl, G. (2001) Scaling in hydrology, Hydrological Processes, *15*, 709-711.

Blöschl, G. (2005) Rainfall-runoff modelling of ungauged catchments. Article 133, in *Encyclopedia of Hydrological Sciences*, vol. 3, edited by M. G. Anderson, pp. 2061-2080, Wiley, Chicester

Blöschl, G., R. B. Grayson, and M. Sivapalan (1995) On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, Hydrological Processes, *9*, 313-330.

Blöschl, G., R. Merz, D. Gutknecht, and M. Bachhiesl (2000) Regional hydrological types - a tool for improved runoff forecasts, paper presented at the XXth Conference of the Danubian Countries on Hydrological Forecasting and the Hydrological Basis of Water Management, SHMÚ Bratislava, Bratislava, Slovakia, Sep. 4-8, 2000. CD-Rom, cl120.pdf.

Blöschl, G., and M. Sivapalan (1995) Scale issues in hydrological modelling - a review, Hydrological Processes, *9* (3-4), 251-290.

Blöschl, G., and M. Sivapalan (1997) Process controls on regional flood frequency: Coefficient of variation and basin scale, Water Resources Research, *33* (12), 2967-2980.

BMLFUW (2003) Hydrologischer Atlas Österreichs (Hydrological Atlas of Austria), Collection of maps, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft and Österreichischer Kunst und Kulturverlag, Vienna, Austria, ISBN 3-85437-250-7.

Bochner, S. (1955) *Harmonic analysis and the theory of probability*, 176 pp., University of California Press, Berkeley and Los Angeles.

Bogaert, P. (1996) Comparison of kriging techniques in a space-time context, Mathematical Geology, *28* (1), 73-86.

Buchholz, O. (2000) *Hydrologische Modelle - Theorie der Modellbildung und Beschreibungssystematik*, Mitteilungen des Lehrstuhls für Wasserbau und Wasserwirtschaft, RWTH Aachen, 300 pp., Verlag Mainz, Wissenschaftsverlag Aachen.

Caeiro, S., M. Painho, P. Goovaerts, H. Costa, and S. Sousa (2003) Spatial sampling design for sediment quality assessment in estuaries, Environmental Modelling & Software, *18* (10), 853-859.

Castellarin, A., L. Brandimarte, A. Montanari, A. Brath, and G. Galeati (2004) Regional flow-duration curves: Reliability for ungauged basins, Advances in Water Resources, *27* (10), 953-965.

Charney, J. G. (1948) On the scale of atmospheric motions, Geofysiske Publikasjoner, Norwegian Academy of Sciences, Oslo, *17* (2), 1-17.

Chilés, J.-P., and P. Delfiner (1999) *Geostatistics. Modelling spatial uncertainty*, 695 pp., Wiley, New York.

Christakos, G. (2000) *Modern Spatiotemporal geostatistics*, International Association for Mathematical Geology Studies in Mathematical Geology, No. 6, 288 pp., Oxford University Press, Oxford.

Cintoli, S., S. P. Neuman, and V. Di Federico (2005) Generating and scaling fractional Brownian motion on finite domains, Geophysical Research Letters, *32* (8), L08404, doi:10.1029/2005GL022608.

Clark, I. (1998) Geostatistical estimation and the lognormal distribution. Geocongress, Pretoria, RSA., Available at: http://uk.geocities.com/drisobelclark/resume/papers/Geocongress1998.zip (verified 01. Nov. 2006).

Clark, I., K. L. Basinger, and W. V. Harper (1987) MUCK - a novel approach to co-kriging, paper presented at The conference on Geostatistical, sensivity, and uncertainty methods for ground-water flow and radionuclide transport modeling, San Fransisco, California edited by B. E. Baxton, Batelle Press, Columbus, OH, pp. 473-493.

Corradini, C., F. Melone, and V. P. Singh (1995) Some remarks on the use of GIUH in the hydrologic practice, Nordic Hydrology, *26*, 297-312.

Cressie, N. (1985) Fitting variogram models by weighted least squares, Mathematical Geology, *17* (5), 563-586.

Cressie, N. (1991) *Statistics for spatial data*, xxii+900 pp., Wiley, New York, NY.

Cressie, N., and D. Hawkins (1980) Robust estimation of the variogram, I, Journal of the International Association for Mathematical Geology, *12*, 115-125.

Cressie, N., and H. C. Huang (1999) Classes of nonseparable, spatio-temporal stationary covariance functions, Journal of the American Statistical Association, *94* (448), 1330-1340.

Cushman, J. H. (1984) On unifying the concepts of scale, instrumentation, and stochastics in the development of multiphase transport theory, Water Resources Research, *20* (11), 1668-1676.

Cushman, J. H. (1987) More on stochastic models, Water Resources Research, *23*, 750-752.

Daviau, J.-L., K. Adamowski, and G. G. Patry (2000) Regional flood frequency analysis using GIS, L-moment and geostatistical methods, Hydrological Processes, *14* (15), 2731-2753.

De Cesare, L., D. E. Myers, and D. Posa (2001) Estimating and modeling space-time correlation structures, Statistics & Probability Letters, *51*, 9-14.

De Cesare, L., D. E. Myers, and D. Posa (2002) FORTRAN programs for space-time modeling, Computers & Geosciences, *28*, 205-212.

De Iaco, S., D. E. Myers, and D. Posa (2001) Space-time analysis using a general product-sum model, Statistics & Probability Letters, *52*, 21-28.

de Marsily, G. (1986) *Quantitative hydrogeology*, 440 pp., Academic Press Inc., London.

Deutsch, C. V., and A. G. Journel (1992) *GSLIB - Geostatistical softare library and user's guide*, 340 pp., Oxford University Press, New York.

Di Federico, V., and S. P. Neuman (1997) Scaling of random fields by means of truncated power variograms and associated spectra, Water Resources Research, *33* (5), 1075-1085.

Dooge, J. C. I. (1973) Linear theory of hydrologic systems, Technical Bulletin No. 1468, 315 pp., Agric. Research Service, US Dep. Of Agriculture, Washington.

Dooge, J. C. I. (1986) Looking for hydrologic laws, Water Resources Research, *22*, 46S-58S.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992) Effective and Efficient Global Optimization for Conceptual Rainfall-runoff Models, Water Resources Research, *28* (4), 1015-1031.

Dunne, T. (1978) Field studies of hillslope flow processes, in *Hillslope Hydrology*, edited by M. J. Kirkby, pp. 227-293, Wiley, Chichester.

Eaton, B., M. Church, and D. Ham (2002) Scaling and regionalization of flood flows in British Columbia, Canada, Hydrological Processes, *16* (16), 3245-3263.

Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. Liu, and A. Namkhai (2000) Temporal and spatial scales of observed soil moisture variations in the extratropics, Journal of Geophysical Research, *105* (D9), 11865-11877.

Faures, J.-M., D. C. Goodrich, D. A. Woolhiser, and S. Sorooshian (1995) Impact of small-scale spatial rainfall variability on runoff modeling, Journal of Hydrology, *173* (1-4), 309-326.

Fischer, H. B., E. J. List, R. C. Y. Koh, J. Imberger, and N. H. Brooks (1979) *Mixing in inland and coastal waters*, 483 pp., Academic Press, San Diego, California.

Fleming, S. W., A. M. Lavenue, A. H. Aly, and A. Adams (2002) Practical applications of spectral analysis to hydrologic time series, Hydrological Processes, *16* (2), 565-574.

Fortak, H. (1982) *Meterologie*, 293 pp., Dietrich Reimer, Berlin.

Foufoula-Georgiou, E., and V. Vuruputur (2000) Patterns and organisation in precipitation, in *Spatial Patterns in Catchment Hydrology: Observations and Modelling*, edited by R. B. Grayson, and G. Blöschl, pp. 82-104, Cambridge University Press, Cambridge.

Franklin, R. B., and A. L. Mills (2003) Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field, FEMS Microbiology Ecology, *44*, 335-346.

Fuentes, M. (2006) Testing for separability of spatial-temporal covariance functions, Journal of Statistical Planning and Inference, *136*, 447-466.

Gallant, J. C., I. D. Moore, M. F. Hutchinson, and P. E. Gessler (1994) Estimating fractal dimension of profiles: a comparison of methods, Mathematical Geology, *26*, 455-481.

Gelhar, L. W. (1993) *Stochastic subsurface hydrology*, 390 pp., Prentice-Hall, Englewood Cliffs, NJ.

Ghosh, B. (1951) Random distances within a rectangle and between two rectangles, Bull. Calcutta Math. Soc., *43*, 17-24.

Gippel, C. J. (2005) Environmental flows: managing hydrological environments. Article 191, in *Encyclopedia of Hydrological Sciences*, vol. 5, edited by M. G. Anderson, pp. 2953-2971, Wiley, Chicester

Goovaerts, P., A. Auchincloss, and A. V. Diez-Roux (2006) Performance comparison of spatial and space-time interpolation tehniques for prediction of air pollutant concentrations in the Los Angeles area, paper presented at IAMG 6th International Congress, Liège, Belgium, September 4-8, 2006.

Gorsich, D. J., and M. G. Genton (2000) Variogram model selection via nonparametric derivative estimation, Mathematical Geology, *32* (2), 249-270.

Gottschalk, L. (1993a) Correlation and covariance of runoff, Stochastic hydrology and hydraulics, *7*, 85-101.

Gottschalk, L. (1993b) Interpolation of runoff applying objective methods, Stochastic hydrology and hydraulics, *7*, 269-281.

Gottschalk, L., I. Krasovskaia, E. Leblois, and E. Sauquet (2006) Mapping mean and variance of runoff in a river basin, Hydrology and earth system sciences, *10*, 469-484.

Grayson, R. B., A. W. Western, F. H. S. Chiew, and G. Blöschl (1997) Preferred states in spatial soil moisture patterns: local and non-local controls, Water Resources Research, *33*, 2897-2908.

Gupta, V. K., and E. C. Waymire (1990) Multiscaling properties of spatial rainfall and river flow distributions, Journal of Geophysical Research, *95* (D3), 1999-2009.

Haltiner, G. J., and R. T. Williams (1980) *Numerical prediction and dynamic meteorology*, 477 pp., John Wiley and Sons, New York.

Harte, J. (1988) *Consider a Spherical Cow, A Course in Environmental Problem Solving*, 283 pp., University Science Books, Sausalito, California.

Hatcher, B. G., J. Imberger, and S. V. Smith (1987) Scaling analysis of coral reef systems: an approach to problems of scale, Coral Reefs, *5*, 171-181.

Haugen, D. A. (1978) Effects of sampling rates and averaging periods on meteorological measurements, paper presented at In: Proceedings, 4th Symposium on Meteorological Observations and Instrumentation, American Meteorological Society, Boston, pp. 15-18.

Hewitt, J. E., S. F. Thrush, V. J. Cummings, and S. J. Turner (1998) The effect of changing sampling scales on our ability to detect effects of large-scale processes on communities, Journal of Experimental Marine Biology and Ecology, *227* (2), 251-264.

Hopmans, J. W., and G. H. Schoups (2005) Soil water flow at different spatial scales. Chapter 74, in *Encyclopedia of hydrological sciences*, vol. 2, edited by M. G. Anderson, Wiley, Chichester, England

Hurst, H. E. (1951) Long-term storage capacity of reservoirs, Trans. Am. Soc. Civ. Eng., *116*, 770-808.

Isaaks, E. H., and R. M. Srivastava (1989) *An introduction to applied geostatistics*, 561 pp., Oxford University Press, New York.

James, B. R., and R. A. Freeze (1993) The worth of data in predicting aquitard continuity in hydrogeological design, Water Resources Research, *29*, 2049-2065.

Jost, G., G. B. M. Heuvelink, and A. Papritz (2005) Analysing the space-time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging, Geoderma, *128* (3-4), 258-273.

Journel, A. G., and C. J. Huijbregts (1978) *Mining geostatistics*, 600 pp., Academic Press, London, UK.

Jurado-Expósito, M., F. López-Granados, J. L. González-Andújar, and L. García-Torres (2004) Spatial and temporal analysis of Convolvulus arvensis L. populations over four growing seasons, European Journal of Agronomy, *21* (3), 287-296.

Katsev, S., and I. L'Heureux (2003) Are Hurst exponents estimated from short or irregular time series meaningful?, Computers & Geosciences, *29* (9), 1085-1089.

Kirchner, J. W., X. Feng, and C. Neal (2000) Fractal Stream Chemistry and its Implications for Contaminant Transport in Catchments, Nature, *403*, 524-527.

Kitanidis, P. K. (1983) Statistical estimation of polynomial generalized covariance functions and hydrologic applications, Water Resources Research, *19* (4), 909-921.

Kitanidis, P. K., and R. W. Lane (1985) Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, Journal of Hydrology, *79*, 53-71.

Klemeš, V. (1974) The Hurst phenomenon: A puzzle?, Water Resources Research, *10*, 675-688.

Koltermann, C. E., and S. M. Gorelick (1996) Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches, Water Resources Research, *32*, 2617-2658.

Krige, D. G. (2004) Some practical aspects of the use of lognormal models for confidence limits and block distributions in South African gold mines, Journal of The South African Institute of Mining and Metallurgy, *104* (5), 285-289.

Kyriakidis, P. C. (2004) A geostatistical framework for area-to-point spatial interpolation, Geographical Analysis, *36* (3), 259-289.

Kyriakidis, P. C., and A. G. Journel (1999) Geostatistical space-time models: A review, Mathematical Geology, *31* (6), 651-684.

Lark, R. M. (2000) Estimating variograms of soil properties by the method-of-moments and maximum likelihood, European Journal of Soil Science, *51*, 717-728.

Leopold, L. B., and T. Maddock (1953) *The hydraulic geometry of stream channels and some physiographic implications*, US Geol. Surv. Prof. Pap., 252, 57 pp.

Leuangthong, O., and C. V. Deutsch (2004) Transformation of residuals to avoid artifacts in geostatistical modelling with a trend, Mathematical Geology, *36* (3), 287-305.

Lopes, V. L. (1996) On the effect of uncertainty in spatial distribution of rainfall on catchment modelling, Catena, *28*, 107-119.

Lovejoy, S., and D. Schertzer (1985) Generalized scale invariance and fractal models of rain, Water Resources Research, *21*, 1233-1250.

Mantoglou, A., and J. L. Wilson (1981) Simulation of random fields with the turning bands method, Report No. 264, 199 pp., Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Mantoglou, A., and J. L. Wilson (1982) The turning bands method for simulation of random fields using line generation by a spectral method, Water Resources Research, *18* (5), 1379-1394.

Matheron, G. (1965) *Les variables regionalisées et leur estimation*, 305 pp., Masson, Paris, France.

McBratney, A. B. (1992) On variation, uncertainty and informatics in environmental soil management, Australian Journal of Soil Research, *30*, 913-935.

Melone, F., C. Corradini, and V. P. Singh (2002) Lag prediction in ungauged basins: an investigation through actual data of the upper Tiber River valley, Hydrological Processes, *16*, 1085-1094.

Menke, W. (1989) *Geophysical data analysis: Discrete inverse theory*, pp., Academic Press Inc., New York, NY.

Merz, R., and G. Blöschl (2003) A process typology of regional floods, Water Resources Research, *39* (12), SWC51-SWC520.

Merz, R., and G. Blöschl (2005) Flood frequency regionalisation - Spatial proximity vs. catchment attributes, Journal of Hydrology, *302* (1-4), 283-306.

Merz, R., G. Blöschl, and J. Parajka (2006a) Raum-zeitliche Variabilität von Ereignisabflussbeiwerten in Österreich (Spatial-temporal variability of event runoff coefficients in Austria), Hydrologie und Wasserbewirtschaftung, *50*, 2-11.

Merz, R., G. Blöschl, and J. Parajka (2006b) Regionalisation methods in rainfall-runoff modelling using large catchment samples, in *Large sample basin experiments for hydrological model parameterization: Results of the model parameter exmperiment - MOPEX (Proceedings of the Paris (2004) and Foz de Iguacu (2005) workshops*, IAHS Publ. 307. In Press.

Miller, E. E., and R. D. Miller (1956) Physical theory of capillary flow phenomena, Journal of Applied Physics, *27*, 324-332.

Milly, P. C. D., and R. T. Wetherald (2002) Macroscale water fluxes 3. Effects of land processes on variability of monthly river discharge, Water Resources Research, *38* (11), 1235, doi:10.1029/2001WR000761.

Mitchell, M. W., M. G. Genton, and M. L. Gumpertz (2005) Testing for separability of space-time covariances, Environmetrics, *16*, 819-831.

Mockus, A. (1998) Estimating dependencies from spatial averages, Journal of Computational and Graphical Statistics, *7* (4), 501-513.

Mohanty, B. P., J. S. Famiglietti, and T. H. Skaggs (2000) Evolution of soil moisture spatial structure in a mixed vegetation pixel during the Southern Great Plains 1997 (SGP97) Hydrology Experiment, Water Resources Research, *36* (12), 3675-3686.

Montanari, A. (2005a) Deseasonalisation of hydrological time series through the normal quantile transform, Journal of Hydrology, *313* (3-4), 274-282.

Montanari, A. (2005b) Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, Water Resources Research, *41* (8), 1-13, DOI: 10.1029/2004WR003826.

Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang (1995) The uncertainty of simple spatial averages using rain gauge networks, Water Resources Research, *31* (8), 2011-2017.

Nash, J. E., and J. V. Sutcliffe (1970) River flow forecasting through conceptual models. Part I—a discussion of principles, Journal of Hydrology, *10* (3), 282-290.

Niehoff, D. (2002) Modellierung des Einflusses der Landnutzung auf die Hochwasserentstehung in der Mesoskala, Dissertation, Universität Potsdam.

Nunan, N., K. Wu, I. M. Young, J. W. Crawford, and K. Ritz (2002) In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil, Microbial Ecology, *44*, 296-305.

Nyquist, H. (1924) Certain factors affecting telegraph speed., Bell Systems Technical Journal, *3*, 324-346.

Oliver, M. A., and A. L. Khayrat (2001) A geostatistical investigation of the spatial variation of radon in soil, Computers & Geosciences, *27*, 939-957.

Omre, H. (1984) The variogram and its estimation, in *Geostatistics for natural resources characterization*, edited by G. Verly et al., pp. 107-125, D. Reidel Publishing Company, Dordrecht, Holland.

Orlanski, I. (1975) A rational subdivision of scales for atmospheric processes, Bull. Am. Met. Soc., *56*, 527-530.

Ortiz C., J., and C. V. Deutsch (2002) Calculation of uncertainty in the variogram, Mathematical Geology, *34* (2), 169-183.

Parajka, J., G. Blöschl, and R. Merz (2006) Regional calibration of catchment models - potential for ungauged catchments, Water Resources Research, *submitted.*

Parajka, J., R. Merz, and G. Blöschl (2005) A comparison of regionalisation methods for catchment model parameters, Hydrology and earth system sciences, *9*, 157-171.

Pardo-Igúzquiza, E. (1998) Maximum likelihood estimation of spatial covariance parameters, Mathematical Geology, *30* (1), 95-108.

Pardo-Igúzquiza, E., and P. Dowd (2001) Variance-covariance matrix of the experimental variogram: Assessing variogram uncertainty, Mathematical Geology, *33* (4), 397-419.

Pilgrim, D. H., editor, (1987) *Australian rainfall and runoff*, The Institution of Engineers, Barton, ACT, Australia.

Plate, E. J. (1993) *Statistik und angewandte Wahrscheinlichkeitslehre für Bauingenieure*, 685 pp., Ernst & Sohn Verlag für Architektur und technische Wissenschaften, Berlin.

Prandtl, L. (1925) Bericht über Untersuchungen zur ausgebildeten Turbulenz, Z. angew. Math. Mech, *5*, 136-139.

Priestley, M. B. (1981) *Spectral analysis and time series*, Probability and mathematical statistics, 890 pp., Academic Press Ltd, London.

Qi, Y., and J. Wu (1996) Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices, Landscape Ecology, *11* (1), 39-49.

Quattrochi, D. A. (1997) GCIP studies in the large scale area (lsa) east (1997-1998): a discussion paper, pp., GEWEX Continental-Scale International Project, Natl. Oceanic and Atmos. Admin, Silver Spring, Md.

R Development Core Team (2004) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rahel, F. J. (1990) The hierarchical nature of community persistence: a problem of scale, The American Naturalist, *136* (3), 328-344.

Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and DMIP Participants (2004) Overall distributed model intercomparison project results, Journal of Hydrology, *298* (1-4), 27-60.

Renard, P., H. Demougeot-Renard, and R. Froidevaux, editors, (2005) *Geostatistics for Environmental Applications*, Springer, Berlin Heidelberg, New York.

Ribeiro Jr., P. J., and P. J. Diggle (2001) geoR: A package for geostatistical analysis, R-NEWS, *1* (2), 15-18.

Rinaldo, A., A. Marani, and R. Rigon (1991) Geomorphological dispersion, Water Resources Research, *27* (4), 513-525.

Rodríguez-Iturbe, I. (1986) Scale of fluctuation of rainfall models, Water Resources Research, *22*, 15S-37S.

Rodríguez-Iturbe, I., and J. M. Mejía (1974) The design of rainfall networks in time and space, Water Resources Research, *10* (4), 713-728.

Rodríguez-Iturbe, I., G. K. Vogel, R. Rigon, D. Entekhabi, F. Castelli, and A. Rinaldo (1995) On the spatial organization of soil moisture fields, Geophysical Research Letters, *22*, 2757-2760.

Root, T. L., and S. H. Schneider (1995) Ecology and climate: Research strategies and implications, Science, *269*, 334-340.

Roth, K., H. Flühler, W. A. Jury, and J. C. Parker, editors, (1990) *Field-scale water and solute flux in soils*, Birkhäuser, Basel.

Rouhani, S., and H. Wackernagel (1990) Multivariate geostatistical approach to space-time data analysis, Water Resources Research, *26* (4), 585-591.

Russo, D., and W. A. Jury (1987) A theoretical study of the estimation of the correlation scale in spatially variable fields 1. stationary fields., Water Resources Research, *23* (7), 1257-1268.

Sauquet, E., L. Gottschalk, and E. Leblois (2000) Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme, Hydrological Sciences Journal, *45* (6), 799-815.

Schmid, H. P. (2002) Footprint modeling for vegetation atmosphere exchange studies: a review and perspective, Agricultural and Forest Meteorology, *113*, 159-183.

Schulze, R. (2000) Transcending scales of space and time in impact studies of climate and climate change on agrohydrological responses, Agriculture, Ecosystems and Environment, *82*, 185-212.

Seed, A. W., R. Srikanthan, and M. Menabde (1999) A space and time model for design storm rainfall, Journal of Geophysical Research, *104* (D24), 31623-31630.

Seyfried, M. S., and B. P. Wilcox (1995) Scale and the nature of spatial variability: Field examples having implications for hydrological modeling, Water Resources Research, *31* (1), 173-184.

Shafer, J. M., and M. D. Varljen (1990) Approximation of confidence limits on sample semivariograms from single realisations of spatially correlated random fields, Water Resources Research, *26* (8), 1787-1802.

Sivapalan, M. (1986) Scale Problems in Rainfall, Infiltration and Runoff Production, PhD Dissertation, Dept. of Civil Engng, Princeton University.

Sivapalan, M., and G. Blöschl (1998) Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves, Journal of Hydrology, *204* (1-4), 150-167.

Sivapalan, M., K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. J. McDonnell, E. M. Mendiondo, P. E. O'Connell, T. Oki, J. W. Pomeroy, D. Schertzer, S. Uhlenbrook, and E. Zehe (2003) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, Hydrological Sciences Journal, *48* (6), 857-880.

Skaugen, T. (1997) Classification of rainfall into small- and large-scale events by statistical pattern recognition, Journal of Hydrology, *200*, 40-57.

Smagorinsky, J. (1974) Global atmospheric modelling and the numerical simulation of climate, in *Weather and Climate Modification*, edited by W. N. Hess, pp. 633-686, Wiley, New York.

Snepvangers, J. J. J. C., G. B. M. Heuvelink, and J. A. Huisman (2003) Soil water content interpolation using spatio-temporal kriging with external drift, Geoderma, *112*, 253-271.

Sokal, R. R., and F. J. Rohlf (1995) *Biometry*, 887 pp., W. H. Freeman, New York.

Stenger, R., E. Priesack, and F. Beese (2002) Spatial variation of nitrate-N and related soil properties at the plot-scale, Geoderma, *105* (3-4), 259-275.

Stommel, H. (1963) Varieties of oceanographic experience, Science, *139*, 572-576.

Stull, R. B. (1988) *An Introduction to Boundary Layer Meteorology*, Dordrecht, 666 pp., Kluwer Academic Publ.

Taylor, G. I. (1921) Diffusion by continuous movements, Proc. London Math. Soc., *20*, 196-211.

Taylor, G. I. (1938) The spectrum of turbulence, Proc. R. Soc. Lond. A., *164*, 476-490.

Theocharopoulos, S. P., G. Wagner, J. Sprengart, M.-E. Mohr, A. Desaules, H. Muntau, M. Christou, and P. Quevauviller (2001) European soil sampling guidelines for soil pollution studies, The Science of The Total Environment, *264* (1-2), 51-62.

Thompson, S. K. (2002) *Sampling*, Wiley series in probability and statistics, 367 pp., Wiley, New York, USA.

Uchida, T., K. Kosugi, and T. Mizuyama (2001) Effects of pipeflow on hydrological process and its relation to landslide: a review of pipeflow studies in forested headwater catchments, Hydrological Processes, *15*, 2151-2174.

Vanmarcke, E. (1983) *Random Fields: Analysis and Synthesis*, 382 pp., The MIT Press, Cambridge, MA.

Wallace, C. S. A., J. M. Watts, and S. R. Yool (2000) Characterizing the spatial structure of vegetation communities in the Mojave Desert using geostatistical techniques, Computers & Geosciences, *26* (4), 397-410.

Webster, R., and M. A. Oliver (1992) Sample adequately to estimate variograms of soil properties, Journal of Soil Science, *43*, 177-192.

Webster, R., and M. A. Oliver (2001) *Geostatistics for environmental scientists*, Statistics in Practice, 271 pp., Wiley, Chichester.

Western, A. W., and G. Blöschl (1999) On the spatial scaling of soil moisture, Journal of Hydrology, *217*, 203-224.

Western, A. W., G. Blöschl, and R. B. Grayson (1998) Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment, Journal of Hydrology, *205*, 20-37.

Western, A. W., and R. B. Grayson (1998) The Tarrawarra data set: Soil moisture patterns, soil characteristics and hydrological flux measurements, Water Resources Research, *34* (10), 2765-2768.

Western, A. W., R. B. Grayson, and G. Blöschl (2002) Scaling of soil moisture: a hydrologic perspective, Annu. Rev. Earth Planet. Sci., *30*, 149-180.

Western, A. W., R. B. Grayson, G. Blöschl, G. R. Willgoose, and T. A. McMahon (1999) Observed spatial organisation of soil moisture and its relation to terrain indices, Water Resources Research, *35*, 797-810.

Western, A. W., S.-L. Zhou, R. B. Grayson, T. A. McMahon, G. Blöschl, and D. J. Wilson (2004) Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, Journal of Hydrology, *286* (1-4), 113-134.

Wiens, J. A. (1989) Spatial scaling in ecology, Functional Ecology, *3*, 385-397.

Wilby, R. L. (1997) *Contemporary Hydrology*, 354 pp., John Wiley, Chichester.

Woodbury, A. D., and E. A. Sudicky (1991) The geostatistical characteristics of the Borden Aquifer, Water Resources Research, *27* (4), 533-546.

Woods, R., R. B. Grayson, A. W. Western, M. Duncan, D. J. Wilson, R. I. Young, R. P. Ibbitt, R. D. Henderson, and T. A. McMahon (2001) Experimental Design and Initial Results from the Mahurangi River Variability Experiment: MARVEX, in *Land Surface Hydrology, Meteorology and Climate*, edited by V. Lakshmi, J. D. Albertson, and J. Schaake, pp. 201-213, American Geophysical Union.

Woods, R., and M. Sivapalan (1999) A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing, Water Resources Research, *35* (8), 2469-2485.

Woods, R., M. Sivapalan, and M. Duncan (1995) Investigating the representative elementary area concept: An approach based on field data., Hydrological Processes, *9*, 291-312.

Yamamoto, J. K. (2000) An alternative measure of the reliability of ordinary kriging estimates, Mathematical Geology, *32* (4), 489-509.

Zehe, E., and G. Blöschl (2004) Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, Water Resources Research, *40*, W10202, doi:10.1029/2003WR002869.

Zimmerman, D. L., and M. B. Zimmerman (1991) A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors, Technometrics, *33* (1), 77-91.