GEO4-4420

# Stochastic Hydrology

Prof. dr. Marc F.P. Bierkens
Prof. Dr. Frans C. van Geer

*Department of Physical Geography*
*Utrecht University*

**Contents**

# Chapter 1: Introduction

## 1.1 Why stochastic hydrology?

The term "stochastics" derives from the Greek word "Stochasticos" (Στοχαστικός) which in turn is derived from "Stochazesthai" (Στοχάζεσθαι), which is derived from Stochos (Στόχος). The word Stochos means "target", while the word Stochazesthai has the following meanings: (a) to shoot (an arrow) at a target, (b) to guess or conjecture (the target), (c) to imagine, think deeply, bethink, contemplate, cogitate, meditate (after Koutsoyiannis, 2010, p. 951). In the modern sense "stochastic" in stochastic methods refers to the random element incorporated in these methods. Stochastic methods thus aim at predicting the value of some variable at non-observed times or at non-observed locations, while also stating how uncertain we are when making these predictions

But why should we care so much about the uncertainty associated with our predictions? The following example (Figure 1.1) shows a time series of observed water table elevations in a piezometer and the outcome of a groundwater model at this location. Also plotted are the differences between the data and the model results. We can observe two features. First, the model time series seems to vary more smoothly then the observations. Secondly, there are noisy differences between model results and observations. These differences, which are called residuals, have among others the following causes:

- *observation errors*. Is it rarely possible to observe a hydrological variable without error. Often, external factors influence an observation, such as temperature and air pressure variations during observation of water levels;
- *errors in boundary conditions, initial conditions and input*. Hydrological models only describe part of reality, for example groundwater flow in a limited region. At the boundaries of the model values of the hydrological variables (such groundwater heads or fluxes) have to be prescribed. These boundary values cannot be observed everywhere, so there is likely to be some error involved. Also, if a model describes the variation of a hydrological system in time, then the hydrological variables at time step zero must be known as it determines how the system will be evolve in later time steps. Again, the initial values of all the hydrological variables at all locations are not exactly known and are estimated with error. Finally, hydrological models are driven by inputs such as rainfall and evaporation. Observing rainfall and evaporation for larger areas is very cumbersome and will usually be done with considerable error;
- *unknown heterogeneity and parameters*. Properties of the land surface and subsurface are highly heterogeneous. Parameters of hydrological systems such as surface roughness, hydraulic conductivity and vegetation properties are therefore highly variable in space and often also in time. Even if we were able to observe these parameters without error, we cannot possibly measure them everywhere. In many hydrological models parameters are assumed homogeneous, i.e. represented by a single value for the entire (or part of the) model region. Even if models take account of the heterogeneity of parameters, this heterogeneity is usually represented by some interpolated map from a few locations where the parameters have been observed. Obviously, these imperfect representations of parameters lead to errors in model results;

- *scale discrepancy.* Many hydrological models consist of numerical approximations of solutions to partial differential equations using either finite element or finite difference methods. Output of these models can at best be interpreted as average values for elements or model blocks. The outputs thus ignore the within element or within block variation of hydrological variables. So, when compared to observations that represent averages for much smaller volumes (virtually points), there is discrepancy in scale that will yield differences between observations and model outcomes (Bierkens et al., 2000);
- *model or system errors.* All models are simplified versions of reality. They cannot contain all the intricate mechanisms and interactions that operate in natural systems. For instance, saturated groundwater flow is described by Darcy's Law, while in reality it is not valid in case of strongly varying velocities, in areas of partly non-laminar flow (e.g. faults) or in areas of very low permeability and high concentrations of solvents. Another example is when a surface water model uses a kinematic wave approximation of surface water flow, while in reality subtle slope gradients in surface water levels dominate the flow. In such cases, the physics of reality differ from that of the model. This will cause an additional error in model results.

In conclusion, apart from the observation errors, the discrepancy between observations and model outcomes are caused by various error sources in our modelling process.

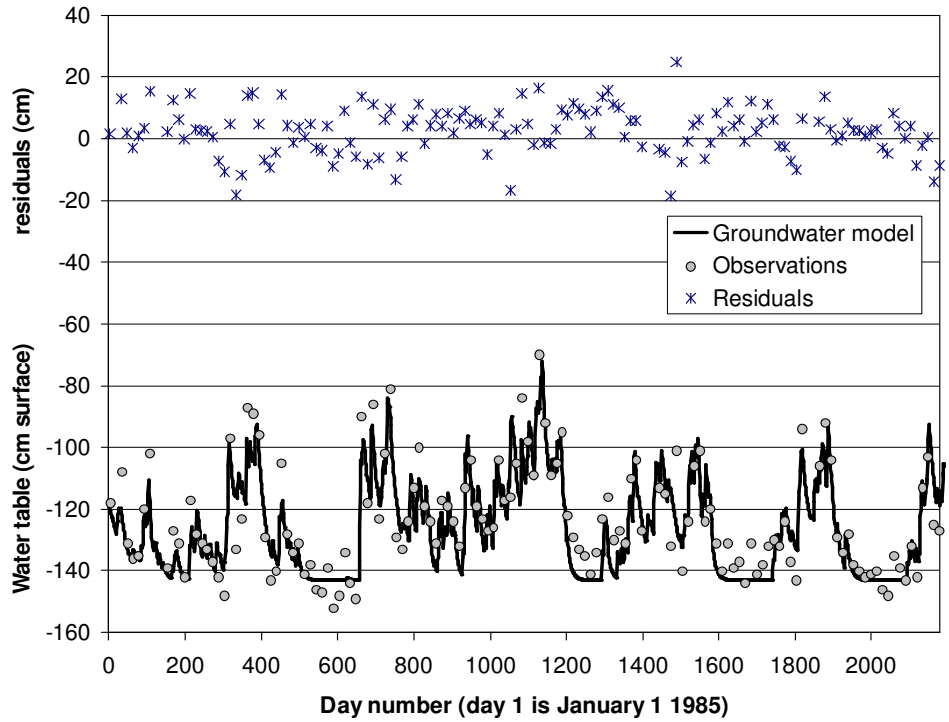



*Figure 1.1 Observed water table depths and water table depths predicted with a groundwater model at the same location. Also shown are the residuals: the differences between model outcome and observations.*

There are two distinct ways of dealing with errors in hydrological model outcomes:

*Deterministic hydrology.* In deterministic hydrology one is usually aware of these errors. They are taken into account, often in a primitive way, during calibration of models. During this phase of the modelling process one tries to find the parameter values of the model (e.g. surface roughness or hydraulic conductivity) such that the magnitude of the residuals is minimized. After calibration of the model, the errors are not explicitly taken into account while performing further calculations with the model. Errors in model outcomes are thus ignored.

*Stochastic Hydrology*. Stochastic hydrology not only tries to use models for predicting hydrological variables, but also tries to quantify the errors in model outcomes. Of course, in practice we do not know the exact values of the errors of our model predictions; if we knew them, we could correct our model outcomes for them and be totally accurate. What we often do know, usually from the few measurements that we did take, is some probability distribution of the errors. We will define the probability distribution more precisely in the next chapters. Here it suffices to know that a probability distribution tells one how likely it is that an error has a certain value.

To make this difference more clear, Figure 1.2 is shown. Consider some hydrological variable $z$, say soil moisture content, whose value is calculated (at some location and at some time) by a unsaturated zone model. The model output is denoted as $\breve{z}$. We then consider the error $e = \breve{z} - z$. Because we do not know it exactly we consider it as a so called *random variable* (chapter 3) $E$ (notice the use of capitals for random variables) whose exact value we do not know but of which we do know the probability distribution. So in case of deterministic hydrology modelling efforts would only yield $\breve{z}$ (upper figure of Figure 1.2a), while stochastic hydrology would yield both $\breve{z}$ and the probability distribution of the (random) error $E$ (lower figure of Figure 1.2a).

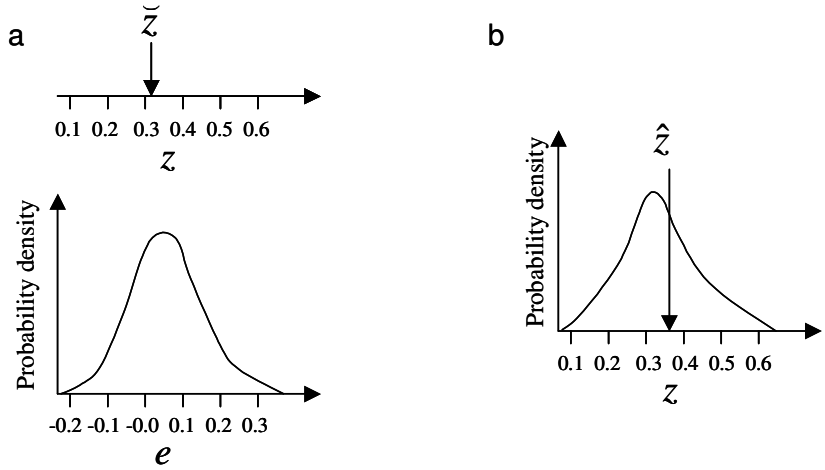



*Figure 1.2 Stochastic Hydrology is about combining deterministic model outcomes with a probability distribution of the errors (Figure 1.2a), or alternatively, considering the hydrological variable as random and determining its probability distribution and some "best prediction"(Figure 1.2b).*

Most of the methods used in stochastic hydrology do not consider errors in model outcomes explicitly. Instead it is assumed that the hydrological variable $z$ itself is a random variable $Z$. This means that we consider the hydrological variable (e.g. soil moisture) as one for which we cannot know the exact value, but for which we can calculate the probability distribution (see Figure 1.2b). The probability distribution of Figure 1.2b thus tells us that although we do not know the soil moisture content exactly, we do know that it is more likely to be around 0.3 then around 0.2 or 0.5. Models that provide probability distributions of target variables instead of single values are called *stochastic models*. Based on the probability distribution it is usually possible to obtain a so called *best prediction* $\hat{z}$, which is the one for which the errors are smallest on average. Incidentally, the value of the best prediction does not have to be the same as the deterministic model outcome $\breve{z}$.

**Box 1. Stochastic models and physics**
A widespread misconception about deterministic and stochastic models is that the former use physical laws (such mass and momentum conservation), while the latter are largely empirical and based entirely on data-analysis. This of course is not true. Deterministic models can be either physically-based (e.g. a model based on Saint-Venant equations for flood routing) and empirical (e.g. a rating curve used as a deterministic model for predicting sediment loads from water levels). Conversely, any physically-based model becomes a stochastic model once its inputs, parameters or outputs are treated as random.

There are a number of clear advantages in taking the uncertainty in model results into account, i.e. using stochastic instead of deterministic models.

- The example of Figure 1.1 shows that model outcomes often give a much smoother picture of reality. This is because models are often based on an idealized representation of reality with simple processes and homogenous parameters. However, reality is usually messy and rugged. This may be a problem when interest is focussed on extreme values: deterministic models typically underestimate the probability of occurrence of extremes, which is rather unfortunate when predicting for instance river stages for dam building. Stochastic models can be used with a technique called "stochastic simulation" (see chapters hereafter) which is able to produce images of reality that are rugged enough to get the extreme statistics right.
- As stated above, the value of the best prediction $\hat{z}$ does not have to be the same as the deterministic model outcome $\breve{z}$. This is particularly the case when the relation between model input (e.g. rainfall, evaporation) or model parameters (e.g. hydraulic conductivity, manning coefficient) and model output is non-linear (this is the case in almost all hydrological models) and our deterministic assessment of model inputs and parameters is not error free (also almost always the case). In this case, stochastic models are able to provide the best prediction using the probability distribution of model outcomes, while deterministic models cannot and are therefore less accurate.
- If we look closely at the residuals in Figure 1 it can be seen that they are correlated in time: a positive residual is more likely to be followed by another positive residual and vice versa. This correlation, if significant, means that there is still some information

present in the residual time series. This information can be used to improve model predictions between observation times, for instance by using time series modelling (chapter 5) or geostatistics (chapter 6). This will yield better predictions than the deterministic model alone. Also, it turns out that if the residuals are correlated, calibration of deterministic models (which assume no correlation between residuals) yield less accurate or even biased (with systematic errors) calibration results when compared with stochastic models that do take account of the correlation of residuals (te Stroet, 1995).

- By explicitly accounting for the uncertainty in our prediction we may in fact be able to make better decisions. A classical example is remediation of polluted soil, where stochastic methods can be used to estimate the probability distribution of pollutant concentration at some non-visited location. Given a critical threshold above which regulation states that remediation is necessary, it is possible to calculate the probability of a false positive decision (we decide to remediate, while in reality the concentration is below the threshold) and that of a false negative (we decide not to remediate while in reality the concentration is above the threshold). Given these probabilities and the associated costs (of remediation and health risk) it is then possible for each location to decide whether to remediate such that the total costs and health risk are minimised.
- There are abundant stochastic methods where a relation is established between the uncertainty in model outcomes and the number of observations in time and space used to either parameterize or calibrate the model. If such a relation exists, it can be used for monitoring network design. For example, in groundwater exploration wells are drilled to perform pumping tests for the estimation of transmissivities and to observe hydraulic heads. The transmissivity observations can be used to make an initial map of transmissivity used in the groundwater model. This initial map can subsequently be updated by calibrating the groundwater model to head observations in the wells. Certain stochastic methods are able to quantify the uncertainty in groundwater head predicted by the model in relation to the number of wells drilled, their location and how often they have been observed (e.g. Bierkens, 2006). These stochastic methods can therefore be used to perform monitoring network optimization: finding the optimal well locations and observation times to minimise uncertainty in model predictions.
- The last reason why stochastic methods are advantageous over deterministic methods is related to the previous one. Stochastic methods enable us to relate the uncertainty in model outcomes to different sources of uncertainty (errors) in input variables, parameters and boundary conditions. Therefore, using stochastic analysis we also know which (error) source contributes the most to the uncertainty in model outcomes, which source comes second etc. If our resources are limited, stochastic hydrology thus can guide us where to spend our money (how many observations for which variable or parameter) to achieve maximum uncertainty reduction at minimum cost. An excellent book on this view on uncertainty is written by Heuvelink (1998).

## 1.2 Scope and content of these lecture notes

These notes aim at presenting an overview of the field of stochastic hydrology at an introductory level. This means that a wide range of topics and methods will be treated, while each topic and method is only treated at a basic level. So, the book is meant as an introduction to the field while showing its breadth, rather than providing an in depth treatise. References are given to more advanced texts and papers for each subject. The book thus aims at teaching the basics to hydrologists who are seeking to apply stochastic methods. It can be used for a one-semester course at third year undergraduate or first year graduate level.

The lecture notes treat basic topics that should be the core of any course on stochastic hydrology. These topics are: descriptive statistics; probability and random variables; hydrological statistics and extremes; random functions; time series analysis; geostatistics; forward stochastic modelling; state prediction and data-assimilation. A number of more advanced topics that could constitute enough material for a second course are not treated. These are, among others: sampling and monitoring; inverse estimation; ordinary stochastic differential equations; point processes; upscaling and downscaling methods, uncertainty and decision making. During the course these advanced topics will be shortly introduced during the lectures. Students are required to study one of these topics from exemplary papers and write a research proposal about it.

## 1.3 Some useful definitions for the following chapters

### *1.3.1 Description of a model according to system's theory*

Many methods in stochastic hydrology are best understood by looking at a hydrological model from the viewpoint of system's theory. What follows here is how a model is defined in system's theory, as well as definitions for state variables, input variables, parameters and constants.

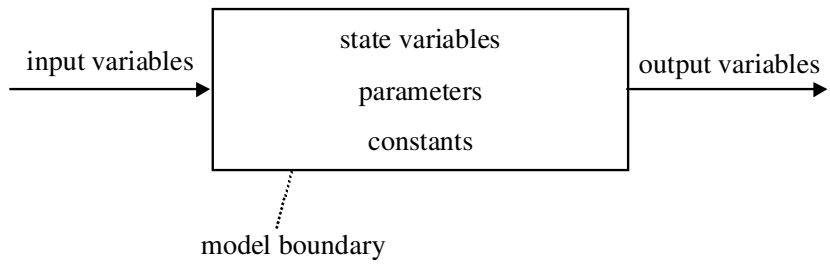



*Figure 1.3 Model and model properties according to system's theory*

Figure 1.3 shows a schematic representation of a model as used in system's theory. A *model* is a simplified representation of part of reality. The *model boundary* separates the part of reality described by the model from the rest of reality. Everything that is to know

about the part of reality described by the model at a certain time is contained in the *state variables*. These are variables because their values can change both in space and time. The variation of the state variables is caused by the variation of one or more *input variables*. Input variables are always observed and originate from outside the model boundary. Consequently, input variables also include boundary conditions and initial conditions such as used when solving differential equations. If the state variables are known, one or more *output variables* can be calculated. An output variable traverses the model boundary and thus influences the part of reality not described by the model. Both input variables and output variables can change in space and time. The state variables are related to the input variables and output variables through *parameters*. Parameters may change in space but are invariant in time. Because they are constant in time, parameters represent the intrinsic properties of the model. Finally, a model may have one or more *constants*. Constants are properties of a model that do not change in both space and time (within the confines of the model). Examples of such constants are the gravity constant and the viscosity of water in density independent groundwater flow at a constant temperature.

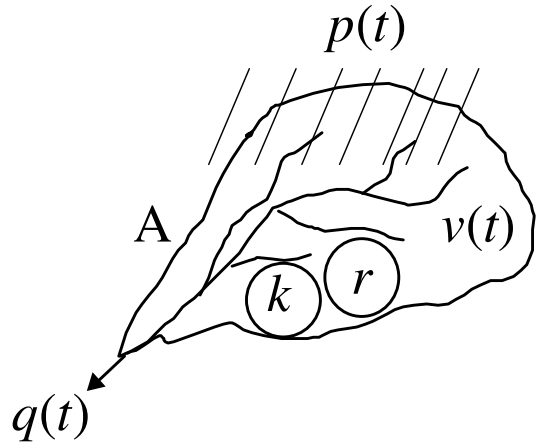



*Figure 1.4 Illustration of model properties following system's theory with a model of a catchment; v(t): state variable, storage surface water in catchment [$L^3$]; q(t): output variable, surface runoff from catchment [$L^3T^{-1}$]; p(t): input variable, precipitation [$LT^{-1}$]; k : parameter, reservoir constant [$T^{-1}$]; r : parameter, infiltration capacity [$LT^{-1}$]; A: constant, area of the catchment [$L^2$].*

Because the description above is rather abstract, we will try to illustrate it with the example shown in Figure 1.4. We consider a model describing the discharge from surface runoff $q$ [$L^3T^{-1}$] from a catchment caused by the average precipitation $p$ [$LT^{-1}$] observed as averages over discrete time steps $\Delta t$, i.e. $q(t)$ and $p(t)$ represent the average discharge and precipitation between $t$-$\Delta t$ and $t$. The model boundary is formed by geographical boundaries such as the catchment boundary (i.e. the divide) on the sides, the catchment's surface below and a few meters above the catchment's surface above, and also by the virtual boundary with everything that is not described by the model such as groundwater flow, soil moisture, chemical transport etc. Obviously, precipitation is the input variable and surface runoff the output variable. The state variable of this model is the amount of water stored on the catchment's surface: $v$ [$L^3$]. The state variable is modelled with the following water balance equation:

$$v(t) = v(t-1) + \left\{ A \cdot [p(t) - r]^{+} - q(t) \right\} \Delta t \tag{1.1}$$

where $r$ [$LT^{-1}$] is the infiltration capacity. The superscript + is added to [$p(t)$-$r$] to denote that if $p(t) < r$ we have [$p(t)$-$r$] = 0. The output variable $q$ is related to the state variable $v$ at the previous time step with the following equation:

$$q(t) = kv(t) \tag{1.2}$$

Through substitution of (1.2) into (1.1) we can calculate the development in time of the state variable directly from the input variable as:

$$v(t) = [1 - k\Delta t] \cdot v(t-1) + A \cdot [p(t) - r]^{+} \Delta t \tag{1.3}$$

Two model parameters can be distinguished: the infiltration capacity of the soil $r$ [$LT^{-1}$] which relates the input variable with the state variable and the catchment parameter $k$ [$T^{-1}$] relating the output variable to the state variable. The constant A [$L^2$] is the area of the catchment.

### *1.3.2 Notation*

The concept of random variables and random functions will be explained in detail in the following chapters. However, it is useful to define the notation conventions briefly in the beginning. Readers can thus refer back to this subsection while studying the rest of this book.

Constants are denoted in roman, e.g. the constant g for gravity acceleration, or A for the area.

Variables and parameters are denoted in italics: e.g. $h$ for hydraulic head and $k$ for hydraulic conductivity. The distinction between deterministic and random (stochastic) variables is made by denoting the latter as capital italics. So, $h$ stands for the deterministic groundwater head (assumed completely known) and $H$ for groundwater head as a random variable.

Vectors and matrices are given in bold face notation. Vectors are denoted as lower case, e.g. **h** a vector of groundwater heads at the nodes of a finite difference model, while matrices are denoted as capitals, such as **K** for a tensor with conductivities in various directions. Unfortunately, it is difficult to make a distinction between stochastic and deterministic vectors and matrices. Therefore, if not clear from the context, it will be indicated explicitly in the text whether a vector or matrix is stochastic or not.

Spatial co-ordinates ($x$,$y$,$z$) are denoted with the space vector **x**, while $t$ is reserved for time. Discrete points in space and time are denoted as $\boldsymbol{x}_i$ and $t_k$ respectively. Random

functions of space, time and space-time are thus denoted as (example with $H$): $H(\mathbf{x})$, $H(t)$, $H(\mathbf{x},t)$.

Outcomes from a deterministic model are denoted as (example with $h$): $\breve{h}$. Optimal estimates of deterministic parameters, constants or variables are denoted with a hat (example with $k$): $\hat{k}$, while optimal predictions of realisations of random variable denoted by $\hat{K}$. Note that the term *estimate* is reserved for deterministic variables and *prediction* for random (stochastic) variables.

To denote a spatial or temporal or spatio-temporal average of a function an overbar is used, e.g. $\overline{h}$ if hydraulic head is deterministic and $\overline{H}$ if it is stochastic. So, $\hat{\overline{H}}(\mathbf{x})$ stands for the prediction of the spatial average of the random function $H(\mathbf{x})$.

# Chapter 2: Descriptive statistics

In this chapter and further on in this book we make use of a synthetic but extremely illustrative data set (Walker lake data set) that has been constructed by Journel and Deutsch (1998)[1]. The data set is used to show how some simple statistics can be calculated.

## 2.1 Univariate statistics

Let us assume that we have made 140 observations of some hydrological variable z (e.g. hydraulic conductivity in m/d). Figure 2.1 shows a plot of the sample locations with the grey scale of the dots according to the value of the observation.

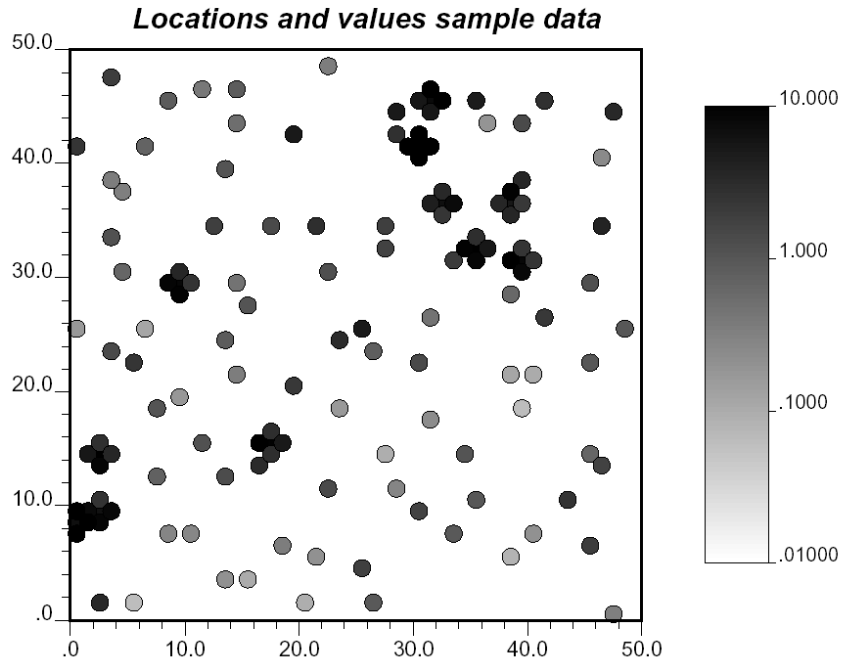



*Figure 2.1 Samples of hydraulic conductivity z*

To obtain insight into our dataset it is good practice to make a *histogram*. To this end we divide the range of value found into a number (say $m$) of classes $z_1$-$z_2$, $z_2$-$z_3$, $z_3$-$z_4$, …, $z_{m-1}$-$z_m$ and counts the number of data values falling into each class. The number of observations falling into a class divided by the total number of observations is called the (relative) *frequency*. Figure 2.2 shows the histogram or *frequency distribution* of the $z$-

[1] All of the larger numerical examples shown in this chapter are based on the Walker-lake data set. The geostatistical analyses and the plots are performed using the GSLIB geostatistical software of Deutsch and Journel (1998).

data. From the histogram we can see how the observations are distributed over the range of values. For instance, we can see that approximately 33% of our data has a value of hydraulic conductivity between 0-1 m/d.

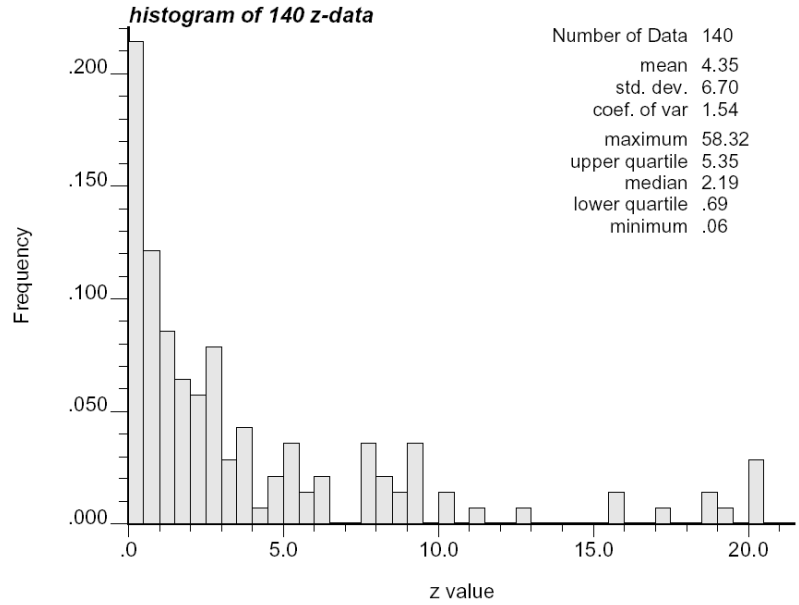



*Figure 2.2 Histogram or frequency distribution of hydraulic conductivity z*

Another way of representing the distribution of data values is by using the *cumulative frequency distribution.* Here we first sort the data in ascending order. Next data are given a rank number $i$, $i$=1,..,$n$, with $n$ the total number of observations (in our case 140). After that, the data values are plotted against the rank number divided by the total number of observations plus on: $i/(n+1)$. Figure 2.3 shows the cumulative frequency distribution of the hydraulic conductivity data.

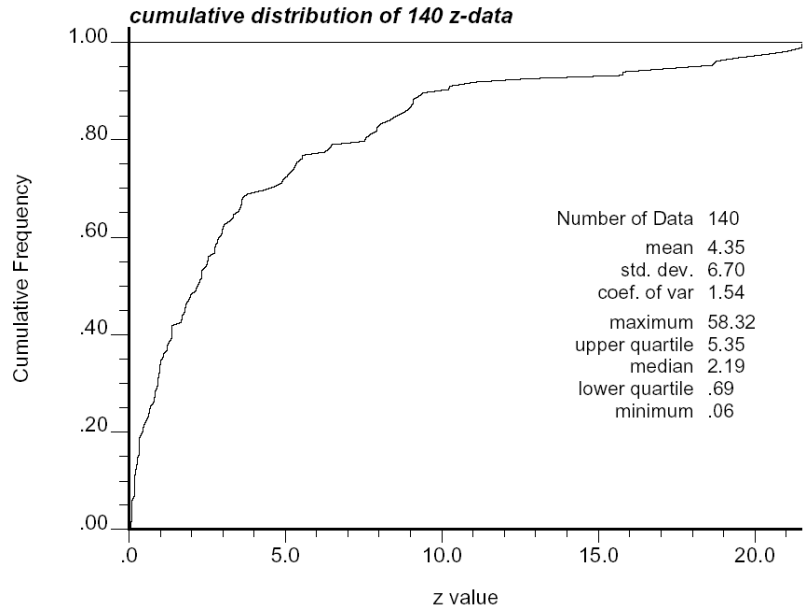



*Figure 2.3 Cumulative frequency distribution of hydraulic conductivity*

The cumulative frequency distribution shows us the percentage of data with values smaller than a given threshold. For instance, from 2.3 we see that 64% of the observations has a value smaller than 5 m/d. Note, that if the 140 samples were taken in such a way that they are representative of the area (e.g. by random sampling) that the cumulative frequency distribution provides an estimate of the fraction of the research area with values smaller or equal to a certain value. This may for instance be relevant when mapping pollution. The cumulative frequency distribution then provides immediately an estimate of the fraction of a terrain with concentrations above critical thresholds, i.e. the fraction that should be remediated.

To make a continuous curve the values between the data points have been linearly interpolated. Figure 2.4 shows the relation between the histogram and the cumulative frequency distribution. It shows that once the cumulative frequency distribution function is constructed from the data (5 data values for this simple example) it can be used to construct a histogram by "differentiation".

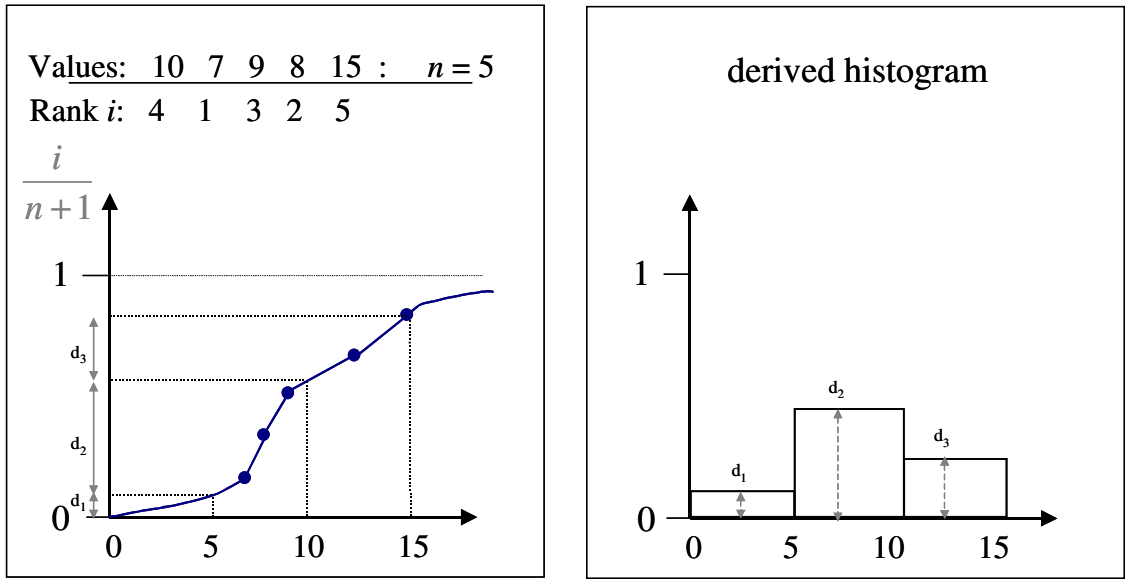



*Figure 2.4 The relation between the Cumulative frequency distribution (left) and the histogram*

To describe the form of frequency distribution a number of measures are usually calculated.

*The mean m* is the average value of the data and is a measure of *locality*, i.e. the centre of mass of the histogram. With $n$ the number data and $z_i$ the value of the $i$th observation we have

$$m_z = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{2.1}$$

*The variance* $s_z^2$ is a measure of the *spread* of the data and is calculated as:

$$s_z^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - m_x)^2 = \frac{1}{n}\sum_{i=1}^{n} z_i{}^2 - m_z^2 \tag{2.2}$$

The larger the variance the wider is the frequency distribution. For instance in Figure 2.5 two histograms are shown with the same mean value but with a different variance.

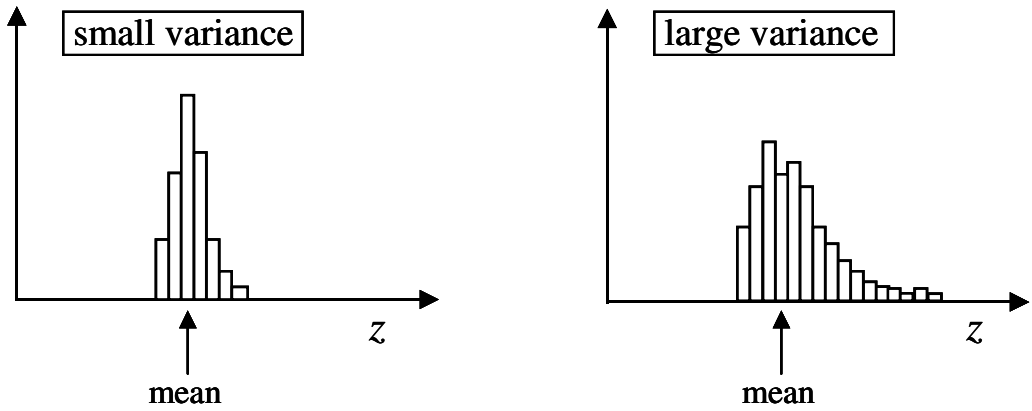



*Figure 2.5 Two histograms of datasets with the same mean value but with different variances*

*Standard deviation*
The standard deviation is also a measure of spread and has the advantage that is has the same units as the original variable. It is calculated as the square-root of the variance:

$$s_z = \sqrt{s_z^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_x)^2} \tag{2.3}$$

*Coefficient of variation*
To obtain a measure of spread that is relative to the magnitude of the variable considered the coefficient of variation is often used:

$$CV_z = \frac{s_z}{m_z} \tag{2.4}$$

Note that this measure only makes sense for variables with strictly positive values (e.g. hydraulic conductivity, soil moisture content, discharge).

*Skewness*
The skewness of the frequency distribution tells us whether it is symmetrical around its central value or whether it is asymmetrical with a longer tail to the left (<0) or to the right (>0)

$$CS_z = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)^3}{s_z^3} \tag{2.5}$$

Figure 2.6 shows two histograms with the same variance, where one is negatively and one is positively skewed.

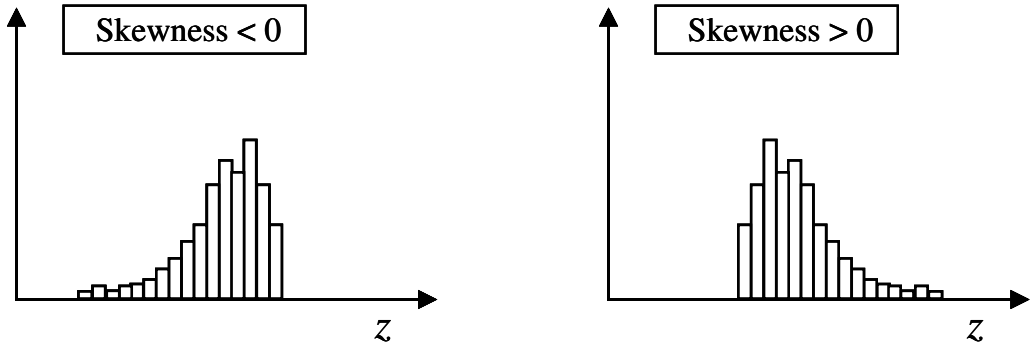



*Figure 2.6 Two frequency distributions with the same variances but with different coefficients of skewness.*

*Curtosis*

The curtosis measures the "peakedness" of the frequency distribution (see Figure 2.7) and is calculated from the data as:

$$CC_z = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)^4}{s_z^4} - 3 \qquad (2.6)$$

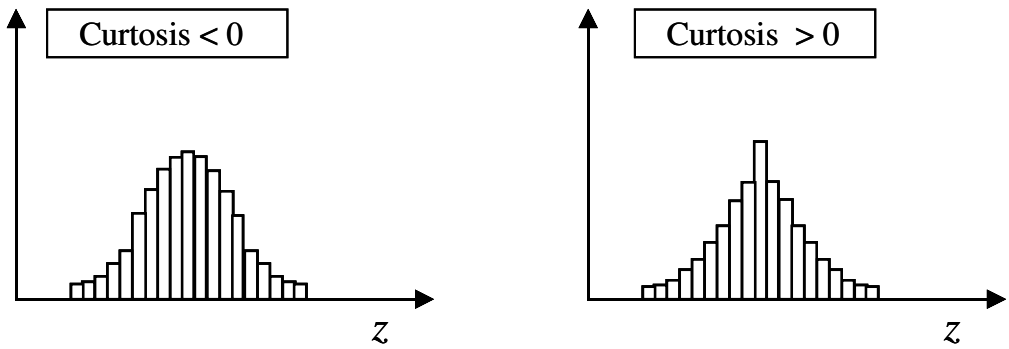



*Figure 2.7 Frequency distributions with positive and negative curtosis*

The value 3 is deducted in Equation (2.6) because for a normal (Gaussian) distribution (see also chapter 3), the first part of Equation (2.6) is exactly equalt to 3. So by $CC_z$ we we compared the peakedness of the distribution with that of a normal distribution, being more peaked when larger than zero and flatter when smaller than zero.

Figure 2.8 shows some additional measures of locality and spread for the cumulative frequency distribution function.

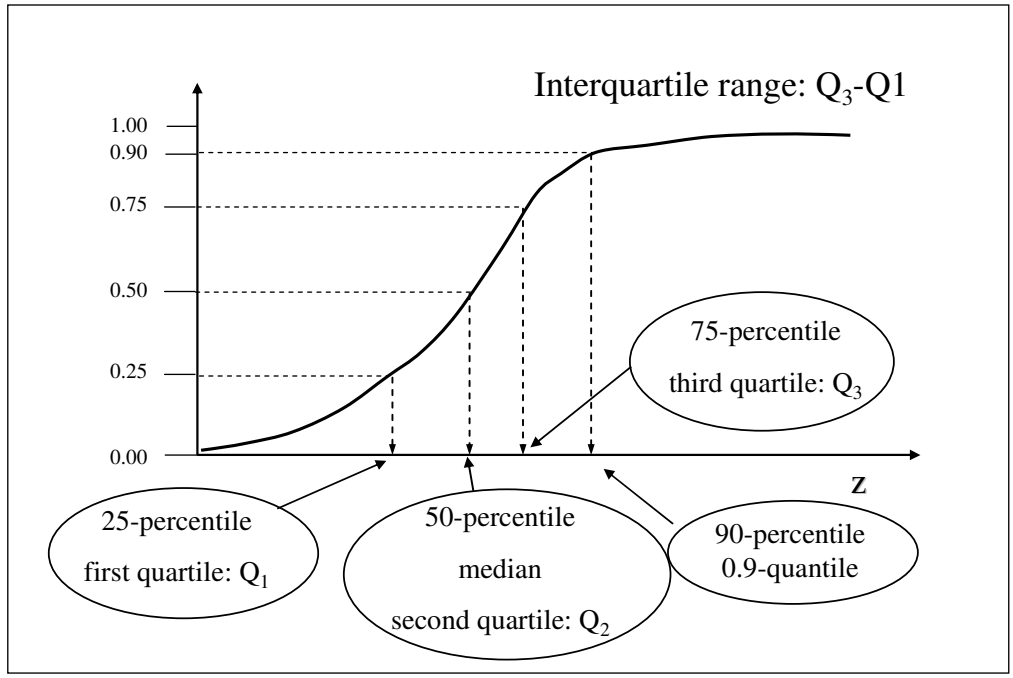



*Figure 2.9 Some additional measures of locality and spread based on the cumulative distribution function.*

The *f-percentile (or f/100-quantile)* of a frequency distribution is the value that is larger than or equal to *f* percent of the data values.

The 50-percentile (or 0.5-quantile) is also called the *median*. It is often used as an alternative measure of locality to the mean in case the frequency distribution is positively skewed. The mean is not a very robust measure in that case as it is very sensitive to the largest (or smallest) values in the dataset.

The 25-percentile, 50-percentile and 75-percentile are denoted as the first, second and third *quartiles* of the frequency distribution: $Q_1$, $Q_2$, $Q_3$ respectively.

*The interquartile range* $Q_3$-$Q_1$ is an alternative measure of spread to the variance that is preferably used in case of skewed distributions. The reason is that the variance, like the mean, is very sensitive to the largest (or smallest) values in the dataset.

An efficient way of displaying locality and spread statistics of a frequency distribution is making a Box-and-whisker plot. Figure 2.10 shows an example. The width of the box provides the interquartile range, its sides the first and third quartile. The line in the middle represents the median and the cross the mean. The whiskers length's are equal to the minimum and the maximum value (circles) as long as these extremes are within 1.5 times the interquartile range (e.g. lower whisker in Figure 2.10), otherwise the whisker is set equal to 1.5 times the interquartile range (e.g. upper whisker in Figure 2.10). Observations lying outside 1.5 times the interquartile range are often identified as outliers. Box-and-whisker plots are a convenient way of viewing statistical properties, especially when comparing multiple groups or classes (see Figure 2.11 for an example of observations of hydraulic conductivity for various texture classes).

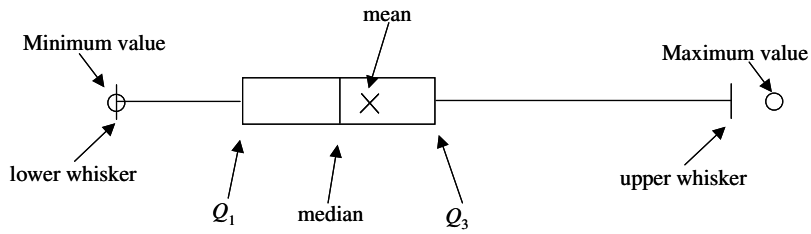



*Figure 2.10 Components of a box-and-whisker plot*

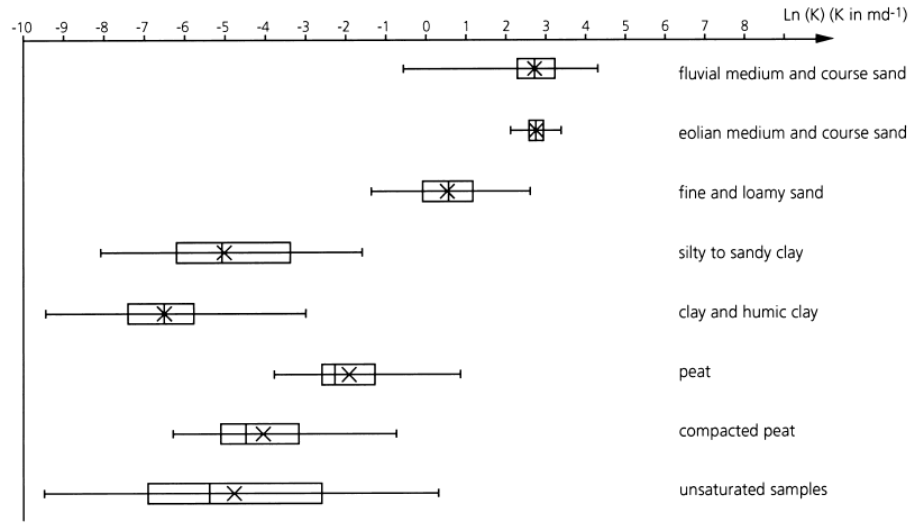



*Figure 2.11 Box-and-whisker plots are a convenient way to compare the statistical properties of multiple groups or classes (from Bierkens, 1996)*

## 2.2 Bivariate statistics

Up to know we have considered statistical properties of a single variable: *univariate* statistical properties. In this section statistics of two variables are considered, i.e. *bivariate* statistics. In case we are dealing with two variables measured simultaneously at a single location or at a single time, additional statistics can be obtained that measure the degree of co-variation of the two data sets, i.e. the degree to which high values of one variable are related with high (or low) values of the other variable.

*Covariance*
The covariance measures linear co-variation of two datasets of variables $z$ and $y$. It is calculated from the data as:

$$C_{zy} = \frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)(y_i - m_y) = \frac{1}{n}\sum_{i=1}^{n} z_i y_i - m_z m_y \tag{2.7}$$

*Correlation coefficient*
The covariance depends on the actual values of the variables. The correlation coefficient provides a measure of linear co-variation that is normalized with respect to the magnitudes of the variables $z$ and $y$:

$$r_{zy} = \frac{C_{zy}}{s_z s_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)(y_i - m_y)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - m_z)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - m_y)^2}} \tag{2.8}$$

A convenient way of calculating the correlation coefficient is as follows:

$$r_{zy} = \frac{n\sum_{i=1}^{n} z_i y_i - \sum_{i=1}^{n} z_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \tag{2.9}$$

So, one calculates $\Sigma z_i$, $\Sigma y_i$, $\Sigma z_i^2$, $\Sigma y_i^2$ and $\Sigma z_i y_i$ and evaluates (2.9). Figure 2.12 shows a so called scatterplot between the $z$-values observed at the 140 locations of Figure 2.1 and the $y$-values also observed there (e.g. z could for instance be hydraulic conductivity and y sand fraction in %). The correlation coefficient between the $z$- and $y$-values equals 0.57.

Figure 2.13 shows examples of various degrees of correlation between two variables, including negative correlation (large values of one exist together with small values of the other). Beware that the correlation coefficient only measures the degree of *linear* co-variation (i.e. linear dependence) between two variables. This can also be seen in Figure 2.13 (lower right figure), where obviously there is strong dependence between z and y, although the correlation coefficient is zero.

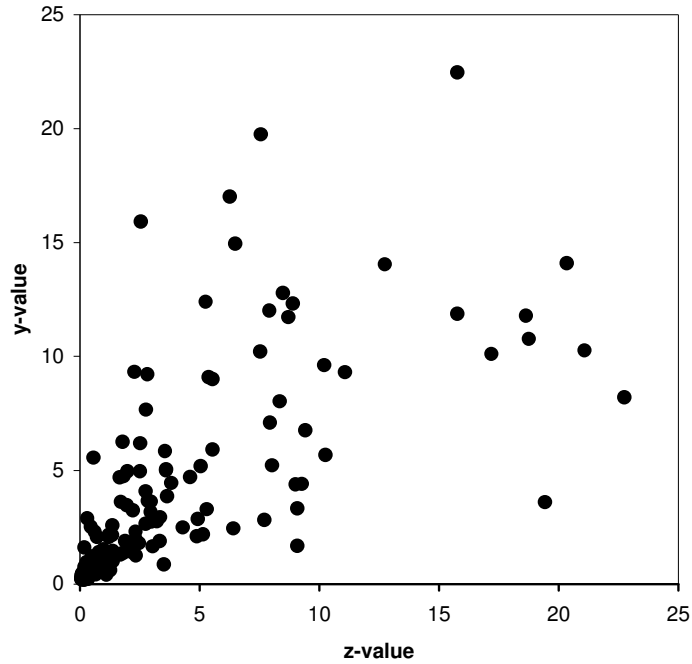



*Figure 2.12 Scatter plot of z- and y-data showing covariation. The correlation coefficient equals 0.57*

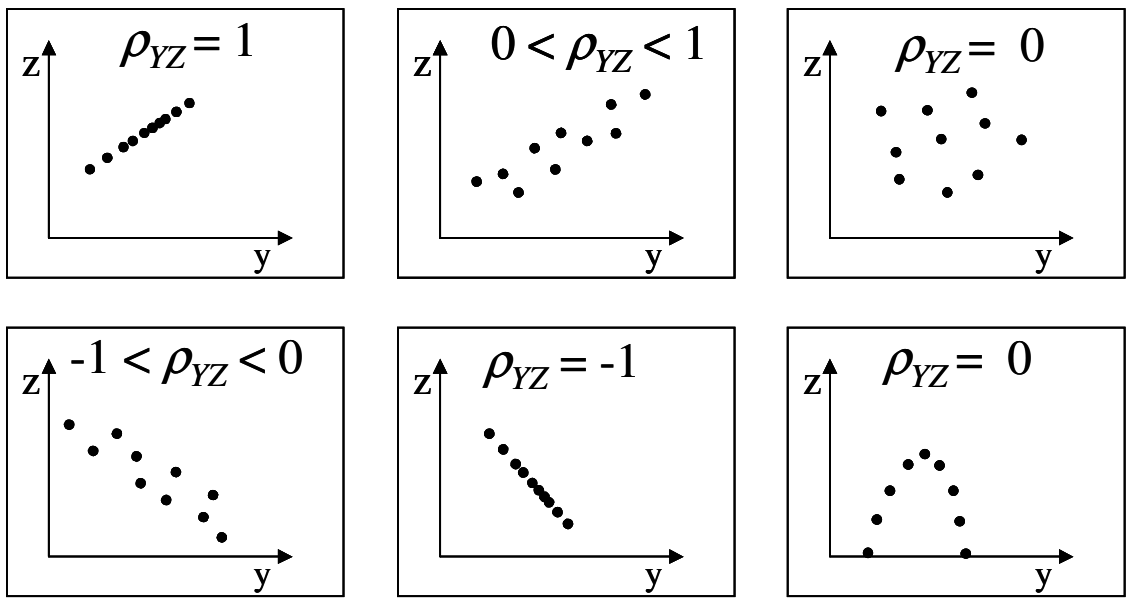



*Figure 2.13 Scatter plots showing covariation and the associated correlation coefficients between two sets of variables that have been observed simultaneously.*

## 2.3 Exercises

Consider the following data set:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *n* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *z* | 1.7 | 6.26 | 7.56 | 7.92 | 0.96 | 2.47 | 2.55 | 0.28 | 1.34 | 0.71 |
| *y* | 1.3 | 17.02 | 19.74 | 12.01 | 0.66 | 1.8 | 15.91 | 0.62 | 2.15 | 2.07 |
| | | | | | | | | | | |
| *n* | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| *z* | 1.66 | 2.99 | 8.71 | 0.09 | 0.62 | 0.99 | 10.27 | 2.96 | 5.54 | 3.61 |
| *y* | 4.68 | 2.74 | 11.72 | 0.24 | 2.3 | 0.52 | 5.67 | 3.17 | 5.92 | 5.03 |

1. Make a histogram of *z* with class-widths of 5 units. What fraction of the data has values between 5 and 10?
2. Construct the cumulative frequency distribution of *z* and *y*
3. Calculate the mean, the variance, the skewnes, the quantiles, the medium and the interquantile range of *z* and *y*.
4. Draw a box-and-whisker plot of the z- and y-values. Are there any possible outliers?
5. Suppose that *z* is the concentration of some pollutant in the soil (mg/kg). Suppose that the samples have been taken randomly from the site of interest. If the critical concentration is 5 mg/kg and the site is 8000 $m^2$. Approximately what area of the site has be cleaned up?
6. Calculate the correlation coefficient between *z* and *y*?
7. What fraction of the data has a z-value smaller than 5 *and* a y-value smaller than 10?
8. What fraction of the data has a z-value smaller than 5 *or* a y-value smaller than 10?

# Chapter 3. Probability and random variables

## 3.1 Random variables and probability distributions

A random variable is a variable that can have a set of different values generated by some probabilistic mechanism. We do not know the value of a stochastic variable, but we do know the probability with which a certain value can occur. For instance, the outcome of throwing a die is not known beforehand. We do however know the probability that the outcome is 3. This probability is 1/6 (if the die is not tampered with). So the outcome of throwing a die is a random variable. The same goes for the outcome of throwing two dice. The probability of the outcome being 3 is now 1/18. A random variable is usually written as a capital (e.g. *D* for the unknown outcome of throwing two dice) and an actual outcome (after the dice have been thrown) with a lower case (e.g. *d*).
The "expected value" or "mean" of a random variable can be calculated if we know which values the random variable can take and with which probability. If *D* is the outcome of throwing two dice, the probability distribution Pr(*d*) is given in the following table:

*Table 3.1 Probabilities of outcomes of throwing two dice*

| D | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pr(d) | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

The mean or expected value is calculated as ($N_d$ the number of possible outcomes and $d_i$ outcome $i$):

$$\mathrm{E}[D] = \sum_{i=1}^{N_d} d_i \Pr[d_i] = 2 \cdot 1/36 + 3 \cdot 2/36 + ..... + 12 \cdot 1/36 = 7 \tag{3.1}$$

That the expected value equals 7 means that if we were to throw the two dice a very large number of times and calculate the average outcomes of all these throws we would end up with a number very close to 7. This means that we could take a sample of $n$ outcomes $d_j$ of a random variable $D$ and estimate its mean with an equation such as (2.1):

$$\hat{\mathrm{E}}[D] = \frac{1}{n} \sum_{j=1}^{n} d_j \tag{3.2}$$

The mean is the centre of mass of the probability distribution and tells us what would be the average of generating many outcomes. The variance is a measure of spread. It tells us something about the width of the probability distribution. Also, it tells us how different the various generated outcomes (throws of the dice) are. A larger variance means that the probability distribution is wide, the variation among outcomes is large and therefore we are more uncertain about the outcome of the random variable. Figure 2.5 shows two probability distributions with the same mean, but with different variances. The variance of a random variable is calculated from the probability distribution as:

$$\begin{aligned}\mathrm{VAR}[D] &= E[(D - E[D])^2] = \sum_{i=1}^{N_d} (d_i - E[D])^2 \Pr[d_i] \\ &= (2-7)^2 \cdot 1/36 + (3-7)^2 \cdot 2/36 + ..... + (12-7)^2 \cdot 1/36 \\ &= 5.8333\end{aligned} \tag{3.3}$$

The variance can be estimated from a random sample of $n$ outcomes ($n$ throws of two dice) $d_j$ as:

$$\hat{\mathrm{VAR}}[D] = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \hat{\mathrm{E}}[\mathrm{D}])^2 \tag{3.4}$$

When we compare equation 3.4 with the variance formula given in chapter 2 (Equation 2.2) we see that here we divide by $n$-1 instead of $n$. This is because in this case we provide an estimator of the variance in case the mean is not known and must be estimated from the data. To obtain an unbiased estimate for the variance (without systematic error) we have to account for the uncertainty about the mean. Hence we divide by $n$-1, leading to a slightly larger variance. The number *n-1* is also called the degrees of freedom. Another way of looking at this is that we have to hand in one degree of freedom as we already used it to estimate the mean!

Instead of the variance, one often uses its square root as a measure of spread. This square root is called the standard deviation. Greek symbols used for the mean, variance and standard deviation are $\mu, \sigma^2$ and $\sigma$ respectively.

The concept of a random variable is used to express uncertainty. If we are uncertain about the actual value of some property (e.g. the concentration of a pollutant or the number of individuals in a population), this property is "modelled" as a random variable. The more uncertain we are about the actual but unknown value, the larger the variance of the probability distribution of this random variable.

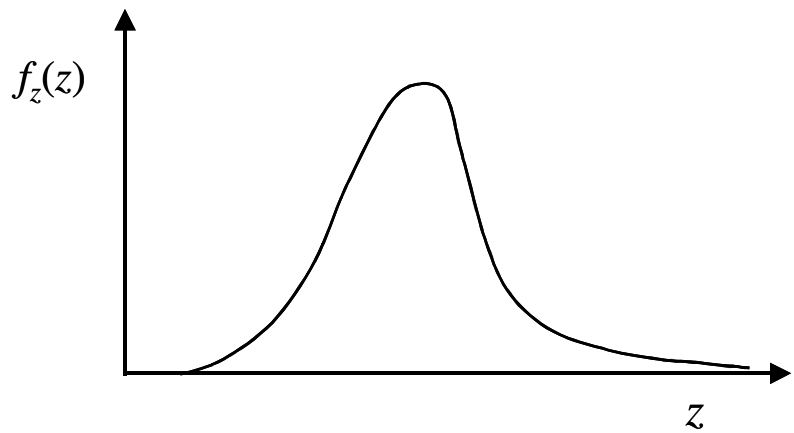



*Figure 3.1. A probability density function*

The outcome of throwing dice is a discrete property. It can only take a limited number of countable values. If the property is continuous it can take any real value between certain bounds (e.g. altitude, hydraulic conductivity, concentration). To describe the probability of a certain outcome of real valued random variable $Z$, instead of a (discrete) probability distribution, a continuous function called the probability density function $f_Z(z)$ is used (see Figure 3.1). The probability density itself does not provide a probability. For instance, we *cannot* say $\Pr[Z=z_1] = f_z(z_1)$! Instead, the probability density gives the *probability mass per unit z*. So, the probability that $Z$ lies between two boundaries *can* be calculated from the probability density by taking the integral:

$$\Pr[z_1 < Z \leq z_2] = \int_{z_1}^{z_2} f_Z(z)dz \tag{3.5}$$

Equation (3.5) can now be used to arrive at a more formal definition of probability density by taking the following limit:

$$f_Z(z) = \lim_{dz \to 0} \frac{\Pr[z < Z \leq z + dz]}{dz} \tag{3.6}$$

An additional condition necessary for $f_Z(z)$ to be a probability density function (pdf) is that the area under it is equal to 1:

$$\int_{-\infty}^{\infty} f_Z(z)dz = 1 \tag{3.7}$$

The probability that $Z$ is smaller than a certain value $z$ is given by the cumulative probability distribution function (cpdf), also simply called distribution function:

$$F_Z(z) = \Pr[Z \leq z] = \int_{-\infty}^{z} f_Z(z)dz \tag{3.8}$$

From 3.8 it also follows that the pdf is the derivative of the cpdf:

$$f_Z(z) = \frac{dF_Z(z)}{dz} \tag{3.9}$$

In risk analysis one is often interested in calculating the probability that a certain critical threshold $z_c$ is exceeded. This can be calculated from both the pdf and the cpdf as:

$$\Pr[Z > z_c] = \int_{z_c}^{\infty} f_Z(z)dz = 1 - F_Z(z_c) \tag{3.10}$$

Similarly, the probability that $Z$ is in between two values can be calculated with the pdf (Equation 3.5), but also with the cpdf:

$$\Pr[z_1 < Z \leq z_2] = F_Z(z_2) - F_Z(z_1) \tag{3.11}$$

## 3.2 Elements of probability theory

The basic rules used in stochastic analysis stem from elementary probability theory. Logically, we would like to start out with a definition of probability. As it turns out this is not straightforward as there exist different notions about probability. A first subdivision that can be made is between objectivistic and subjectivistic notions of probability (e.g. Christakos, 1992).

### 3.2.1 Objectivistic definitions

There are three different definitions here. Central to these definitions is the notion of some event $A$ (e.g. an event can be the die falling on 5, a flood occurring or the value of conductivity being in the 5-10 m/d range).

The classical definition
This is the oldest notion of probability and it can for instance be used to deduce the probability distributions of throwing two dice. The probability Pr($A$) of an event $A$ is determined a priori (without experimentation) with the ratio:

$$\Pr(A) = \frac{N_A}{N} \tag{3.12}$$

with $N$ the number of possible outcomes and $N_A$ all the outcomes resulting in event $A$, provided that all outcomes are *equally likely*. A problem with this definition of course is that it is not always possible to deduce $N$ (especially is $N$ is infinite such as in continuous valued events). Moreover, the definition contains the term equally likely, which is itself a probability statement.

The relative frequency definition
This notion of probability uses the following definition. The probability Pr($A$) of an event $A$ is the limit of performing probabilistic experiments:

$$\Pr(A) = \lim_{n \to \infty} \frac{n_A}{n} \tag{3.13}$$

where $n_A$ the number of occurrences of $A$ and $n$ the number of trials. This frequentistic view of probability is intuitively appealing because it provides a nice link between

probability and the relative frequency described in chapter 2. However, there are some problems, such as the fact that it is in practice not possible to perform infinite trials.

The axiomatic definition

This definition, which can be attributed to Kolmogorov († 1933), uses set theory to define probability. We imagine an experiment, in which the event $A$ is the outcome of a trial. The set of all possible outcomes of a trial is called the sampling space or the certain event $S$. The union $\{A \cup B\}$ of two events $A$ and $B$ is the event that $A$ or $B$ occurs. The axiomatic definition of probability is based entirely on the following three postulates:

1. The probability of an event is a positive number assigned to this event:

$$\Pr(A) \geq 0 \tag{3.14}$$

2. The probability of the certain event (the event is equal to all possible outcomes) equals 1:

$$\Pr(S) = 1 \tag{3.15}$$

3. If the events A and B are *mutually exclusive* then:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \tag{3.16}$$

Figure 3.2 shows schematically using so called Venn diagrams the certain event $S$ with events A and B that are mutually exclusive (left figure) and not mutually exclusive (right figure). Some more derived rules based on the axiomatic probability definition will be given hereafter.

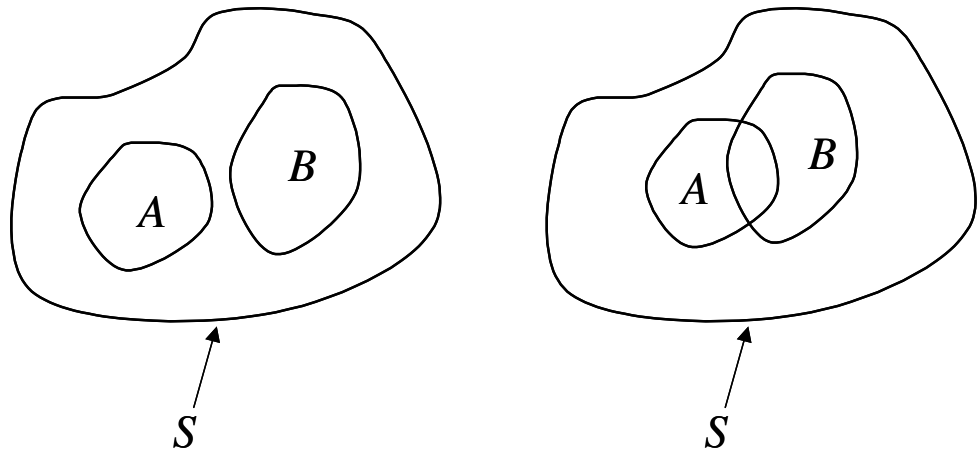



*Figure 3.2 Example of Venn diagrams showing two mutually exclusive events A and B and two events that are not mutually exclusive.*

Generally, the axiomatic notion of probability is deemed superior to the others. For an extensive description on the subtle differences and peculiarities of the various definitions of probability we refer to Papoulis (1991).

### 3.2.2 Subjectivistic definition

In the subjectivistic definition, probability measures our "confidence" about the value or a range of values of a property whose value is unknown. The probability distribution thus reflects our uncertainty about the unknown but true value of a property. The probability density function then measures the *likelihood* that the true but unknown value is between certain limits. So, in this subjectivistic definition of probability we do not have to think about frequencies, population sizes or events. We are faced with some property that is not known exactly, either because we can only measure it with some (random) measurement error or because we cannot measure it at all, or only partly. Think for instance about hydraulic conductivity in a heterogeneous geological formation. It is impossible to measure it everywhere at reasonable costs, so in practice we can only measure it at a limited number of locations (often with measurement error, because taking undisturbed sediment cores and perform Darcy experiments is very difficult in practice). If we have an aquifer with no observations, but we do know that it consists of sands, we know that the true value at some location is more likely to be close to 10 $md^{-1}$ than 0.0001 $md^{-1}$ or 1000 $md^{-1}$. Based on this experience from elsewhere (observations in other aquifers) we can then define an *a priori* probability distribution that measures the likelihood of the various possible values at our unknown location. What we do in the back of our mind is collecting all the information we have on sandy aquifers in the rest of the world and propose that their conductivities are similar to the one at hand. We can then use observations from these other aquifers to construct a prior distribution function. If subsequently observations are being collected that are specific to the aquifer at hand, we may use these observations to narrow the a priori probability distribution down, by incorporating the observed values. What results is a so called *a posteriori* probability distribution that has a smaller variance, such that we are more certain about the unknown conductivity at an unobserved location then we were before the observations.

The subjectivistic probability does not need any observation to define it. It can be defined from the top of our head, thus expressing *our* uncertainty or confidence about an unknown value. This way of viewing probability and the possibility to update such a probability with observations is called Bayesian statistics (see hereafter) and has led to much debate and controversy in the statistical community, especially between people who accept Bayesian statistics and people who view probability as axiomatic or frequentistic. In stochastic hydrology, which is an applied scientific discipline, the various notions of probability have never been a real issue, but insights have been borrowed from the various probability concepts:

- probability is mostly viewed as subjectivistic (except maybe Hydrological statistics (chapter 4) which is more frequentistic in nature);
- a priori probability distributions are often not really subjectivistic but based on observations taken at other times or locations in the same area of interest;
- updating of the a priori probability distributions to a posteriori distributions makes use of Bayes' theorem, which is in fact best defined using axiomatic probability rules.

**Box 2: About uncertainty and reality**
Often one can read in papers statements like: "hydraulic conductivity is uncertain", or "the uncertain behaviour of river discharge is modelled as.." . Such statements seem to suggest that reality itself is random. Whether this is true or not is a rather philosophical question. The most common view is that nature is deterministic, except maybe at the sub-atomic level. We will adhere to this view in this book and use the following notion of reality and uncertainty, which relates to a subjectivistic view on probability: Reality is completely deterministic. However, we do not have perfect knowledge of reality, because we only have limited information on it. We can only observe it partly, observe it with error or do not exactly know the underlying process description. It is because of this that we may perceive (parts of) reality as random and find that a random variable or random process and the associated concepts of probability constitute a useful model of reality. Therefore, randomness is not a property of reality but a property of the stochastic *model* that we use to describe reality and our uncertainty about it.

### 3.2.3 Brief review of elementary probability theory

Even though the definition of probability may be a subjectivistic one, to really perform calculations with probabilities requires rules derived from the axiomatic definition. Here we will review some of these rules. This review is based on Vanmarcke (1983).

The basic axioms of probability are given by 3.14 and 3.16. As stated above, the *union of events A* and *B* is the event that either *A* or *B* occurs and is denoted as $\{A \cup B\}$. The *joint event* $\{A \cap B\}$ is the event that both *A* and *B* occur. From the Venn diagram Figure 3.3 it follows directly that the probability of the union of events and the joint event are related as follows:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \tag{3.17}$$

If events A and B are mutually exclusive (Figure 3.2 left figure) it can be seen that $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ and $\Pr(A \cap B) = 0$. If the multiple events $A_1$, $A_{2...,}$ $A_M$ are mutually exclusive, then probability of the union of these events is the sum of their probabilities:

$$\Pr(A_1 \cup A_2 \cup ... \cup A_M) = \sum_{i=1}^{M} \Pr(A_i) \tag{3.18}$$

In the special case that all events in *S* are mutually exclusive and that they constitute all possible events (they are said to be *collectively exhaustive*) then it follows that their probabilities sum to 1:

$$\sum_{i=1}^{M} \Pr(A_i) = 1 \tag{3.18}$$

Mutually exclusive and collectively exhaustive events are also called *simple events*. For $M$=2 simple events imply that $\Pr(A^c) = 1 - \Pr(A)$ with $A^c$ the *complement* of $A$

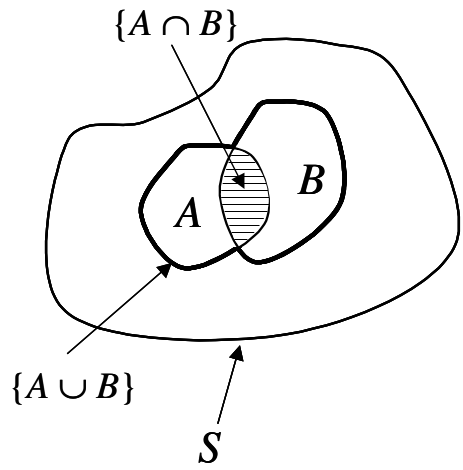



*Figure 3.3 Venn diagram showing the relation between the union of events and joint events.*

The degree of probabilistic dependence between two events is measured by the so called *conditional probability* of A given B:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{3.19}$$

Of course $A$ and $B$ can be interchanged so that

$$\Pr(A \cap B) = \Pr(A \mid B)P(B) = \Pr(B \mid A)P(A) \tag{3.20}$$

Two events A and B are said to be (statistically) independent if the probability of their joint event is equal to the product of the probabilities of the individual events:

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{3.21}$$

This also implies that $\Pr(A \mid B)\Pr(B) = \Pr(A)$ and $\Pr(B \mid A)\Pr(A) = \Pr(B)$. This means that knowledge about B does not have an effect on the uncertainty about A and vice versa.

Finally, if we consider a set of simple events $A_i$ intersected by an event $B$, we can deduce from the Venn diagram in Figure 3.4 and Equation (3.20) the following relationship:

$$\Pr(B) = \sum_{i=1}^{M} \Pr(A_i \cap B) = \sum_{i=1}^{M} \Pr(B \mid A_i)\Pr(A_i) \tag{3.22}$$

This shows that the probability of $B$ is the weighted sum of the probability of $B$ given $A_i$ with the probability of $A_i$ as weight. This relationship is known as the *total probability theorem*. An example on how to use this is as follows: suppose that we have from previous data-analyses for each texture class, i.e. sand, clay, silt and peat, the probability distribution of hydraulic conductivity. Then, if we have estimated at some unvisited location the probabilities on sand, clay, silt and peat from borehole data (see for instance chapter 7), we are able to derive probabilities of hydraulic conductivity from these using (3.22).

The conditional probability of $A_i$, given B can be calculated by combining Equations (3.19) and (3.22):

$$\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i)\Pr(A_i)}{\sum_{j=1}^{M} \Pr(B \mid A_j)\Pr(A_j)} \tag{3.23}$$

This relationship is known as Bayes' theorem. As explained before, it can be used to update a priori distributions using data. For instance, suppose that we have from information elsewhere the *a priori* probability of soil moisture content at some non-observed location, say $\Pr(A_i)$. Let $B$ represent the outcomes of observations around the non-observed location. The probability $\Pr(A_i|B)$ is called the *a posteriori* probability, i.e. the probability of soil moisture content at the unobserved location given the observed values around it. To calculate the a posteriori probability we need the so called *likelihood* $\Pr(B|A_i)$, i.e. the probability of observing $B$ given that soil moisture content $A_i$ is true.

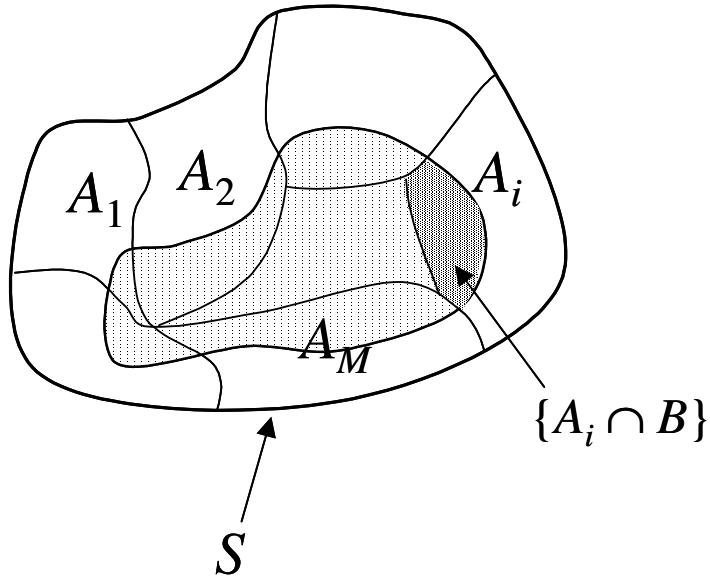



*Figure 3.4 Venn diagram showing the intersection between an event B and a set of mutually exclusive and collectively exhaustive events $A_i$, $i$=1,..,$M$.*

## 3.3 Measures of (continuous) probability distributions

In Chapter 2 we introduced a number of measures of frequency distributions, which are related to datasets and their histogram form. Similar to a histogram, the locality and form of probability density functions can be described by a number of measures. These measures are like Equations (3.1) and (3.3), but as we are now working with continuous variables, the sums are replaced by integrals. Before introducing these measures we start with the definition of the *expected value*.

Let $g(Z)$ be a function of some random variable $Z$. The expected value of $g(Z)$ is defined as:

$$\mathrm{E}[g(Z)] = \int_{-\infty}^{\infty} g(z)\, f_Z(z) dz \tag{3.24}$$

For discrete random variables $D$ the expected value $g(D)$ is defined as

$$\mathrm{E}[g(D)] = \sum_i g(d)\, p_D(d_i) \tag{3.25}$$

where $p_D(d_i)$ is the *probability mass function* of a discrete random variable (e.g. the probabilities in Table 3.1). So we see that it can be viewed as the weighted sum of $g(Z)$ over the domain of $Z$ with the probability density of $Z$ as weight. If we take $g(Z) = Z$ we obtain the *mean* or *expected* value of $Z$ (the continuous version of 3.1).

$$\mu_Z = \mathrm{E}[Z] = \int_{-\infty}^{\infty} z\, f_Z(z) dz \tag{3.26}$$

If we take $g(Z) = (Z-\mu_Z)^2$ we obtain the *variance* (continuous version of 3.3):

$$\sigma_Z^2 = \mathrm{E}[(Z-\mu)^2] = \int_{-\infty}^{\infty} (z-\mu)^2\, f_Z(z) dz \tag{3.27}$$

The estimators of the mean and the variance are the same as in Equations (3.2) and (3.4) with $d_j$ replaced with $z_j$. The standard deviation is given by $\sigma_Z = \sqrt{\sigma_Z^2}$ and the coefficient of variation by

$$CV_Z = \frac{\sigma_Z}{\mu_Z} \tag{3.28}$$

The following rules apply to mean and variance (if $a$ and $b$ are deterministic constants):

$$\mathrm{E}[a + bZ] = a + b\,\mathrm{E}[Z] \tag{3.29}$$

$$\mathrm{VAR}[a + bZ] = b^2\ \mathrm{VAR}[Z] \tag{3.30}$$

The *skewness* is defined as:

$$CS_Z = \frac{\mathrm{E}[(Z-\mu)^3]}{\sigma_Z^3} = \frac{1}{\sigma_Z^3}\int_{-\infty}^{\infty}(z-\mu)^3\, f_Z(z)dz \tag{3.31}$$

and the *curtosis* as:

$$CC_Z = \frac{\mathrm{E}[(Z-\mu)^4]}{\sigma_Z^4} - 3 = \frac{1}{\sigma_Z^4}\int_{-\infty}^{\infty}(z-\mu)^3\, f_Z(z)dz - 3 \tag{3.32}$$

Skewness and curtosis can be estimated with equations 2.5 and 2.6 with $n$ replaced by $n$-1 if the mean and the variance have been estimated as well.

## 3.4 Moments

The $k$th moment $\mu_k$ of a random variable $Z$ is defined as:

$$\mu_k = \mathrm{E}[Z^k] = \int_{-\infty}^{\infty} z^k\, f_Z(z)dz \tag{3.33}$$

Often, one works with the central moments defined as:

$$M_k = \mathrm{E}[(Z-\mu_Z)^k] = \int_{-\infty}^{\infty}(Z-\mu_Z)^k\, f_Z(z)dz \tag{3.34}$$

Moments and central moments are related to the more standard measures of probability distributions as:

$$\begin{aligned}
&\mu_Z = \mu_1 \\
&\sigma_Z^2 = M_2 = \mu_2 - (\mu_1)^2 \\
&CS_Z = \frac{M_3}{\sigma_Z^3} \\
&CC_Z = \frac{M_4}{\sigma_Z^4} - 3
\end{aligned} \tag{3.35}$$

## 3.5 Characteristic functions

There are a number of transformations of probability distributions that are useful when working with random variables. We start we the *moment generating function*, which is defined as:

$$M_Z(s) = \mathrm{E}[e^{sZ}] = \int_{-\infty}^{\infty} e^{sz} f_Z(z) dz \tag{3.36}$$

The moment-generating function is related to the Laplace transform. The moment generating function can be used to calculate the moments as:

$$\mu_k = \mathrm{E}[Z^k] = \left. \frac{d^k M_Z(s)}{ds^k} \right|_{s=0} \tag{3.37}$$

Take for instance the negative exponential distribution:

$$f_Z(z) = \lambda e^{-\lambda z}, z \geq 0 \tag{3.38}$$

The moment generating function of this distribution is:

$$M_Z(s) = \lambda \int_0^{\infty} e^{sz} e^{-\lambda z} dz = \lambda \int_0^{\infty} e^{-(\lambda - s)z} dz = \frac{\lambda}{\lambda - s} \tag{3.39}$$

From this we can calculate the moments:

$$\mu_1 = \left. \frac{d}{ds} \left( \frac{\lambda}{\lambda - s} \right) \right|_{s=0} = \left. \frac{\lambda}{(\lambda - s)^2} \right|_{s=0} = \frac{1}{\lambda} \tag{3.40}$$

$$\mu_2 = \left. \frac{d^2}{ds^2} \left( \frac{\lambda}{\lambda - s} \right) \right|_{s=0} = \left. \frac{2\lambda}{(\lambda - s)^3} \right|_{s=0} = \frac{2}{\lambda^2} \tag{3.41}$$

So the variance is equal to $1/\lambda^2$.

Another transformation often used is the *characteristic function*:

$$\varphi_Z(s) = \mathrm{E}[e^{isZ}] = \int_{-\infty}^{\infty} e^{isz} f_Z(z) dz \quad \text{with } \sqrt{i} = -1 \tag{3.42}$$

The characteristic function is akin to the Fourier transform. The inverse of (3.42) can also be defined:

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isz}\varphi_Z(s)ds \tag{3.43}$$

This means that if two random variables have the same characteristic function they are identically distributed. Like the moment generating function the characteristic function can be used to calculate the moments:

$$\mu_k = \mathrm{E}[Z^k] = \frac{1}{i^k}\frac{d^k\varphi_Z(s)}{ds^k}\bigg|_{s=0} \tag{3.44}$$

If we expand the exponential exp(*isZ*) in a Taylor series around *Z*=0 we obtain:

$$e^{isZ} = 1 + isZ + \frac{1}{2}(isZ)^2 + \frac{1}{6}(isZ)^3 + \ldots \tag{3.45}$$

By taking expectations on both sides we obtain an expression relating the characteristic function to moments of *Z*:

$$\begin{aligned}\varphi_Z(s) = E[e^{isZ}] &= 1 + isE[Z] + \frac{1}{2}(is)^2E[Z^2] + \frac{1}{6}(is)^3E[Z^3] + \ldots \\ &= 1 + is\mu_1 + \frac{1}{2}(is)^2\mu_2 + \frac{1}{6}(is)^3\mu_3 + \ldots.\end{aligned} \tag{3.46}$$

Or written more generally:

$$\varphi_Z(s) = \sum_{k=1}^{\infty}\frac{(is)^k}{k!}\mu_k \tag{3.47}$$

I can be proven that the pdf of *Z* is completely defined by its characteristic function. From (3.47) it can also be seen that if all moments exist and if (3.47) converges, that the characteristic function and (through 3.43) also the pdf is completely defined. This is the case for most of the pdfs encountered in practice. This means that for all practical purposes one can approximate the pdf through a sufficient number of moments.

Some additional properties of the characteristic function: If $Z_1$ and $Z_2$ are two independent random variables we have (Grimmet and Stirzaker, 1982):

$$\varphi_{Z_1+Z_2}(s) = \varphi_{Z_1}(s)\varphi_{Z_2}(s) \tag{3.48}$$

The same relation holds for the moment generating function. Also we have that for a variable $Y$= a+b$Z$ the characteristic function becomes (Grimmet and Stirzaker, 1982):

$$\varphi_Y(s) = e^{is\mathrm{b}}\varphi_Z(\mathrm{a}s) \tag{3.49}$$

From (3.48) we can also deduce that if we have the sum of $M$ identically distributed variables with characteristic function $\varphi_Z(s)$

$$Y = Z_1 + Z_2 + ... + Z_M = \sum_{k=1}^{M} Z_k, \tag{3.50}$$

the characteristic function of $Y$ is given by:

$$\varphi_Y(s) = \left[\varphi_Z(s)\right]^M. \tag{3.51}$$

The form of (3.51) stimulates the introduction of the logarithm of the characteristic function. This is called the *cumulant function* and is defined as:

$$K_Z(s) = \ln \varphi_Z(s) \tag{3.52}$$

From (3.51) and (3.52) then follows that the cumulant of the sum $Y$ of $M$ identically distributed variables with cumulant function $K_Z(s)$ is by:

$$K_Y(s) = MK_Z(s) \tag{3.53}$$

The series expansion of the cumulant function is given by:

$$K_Z(s) = \sum_{n=1}^{\infty} \frac{(is)^n}{n} \kappa_n \tag{3.54}$$

where $\kappa_n$ are called the *cumulants* which are related to the cumulant function as:

$$\kappa_n = \frac{1}{i^n} \frac{d^k K_Z(s)}{ds^k}\bigg|_{s=0} \tag{3.55}$$

The cumulants are conveniently related to the moments of the pdf, such that we can calculate moments from cumulants and vice versa:

$$
\begin{aligned}
k_1 &= \mu_1 \\
k_2 &= \mu_2 - \mu_1^2 = \sigma^2 \\
k_3 &= \mu_3 - 3\mu_2\mu_1 + 3\mu_1^3 \\
k_4 &= \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4
\end{aligned}
\tag{3.56}
$$

Up to know we have only talked about continuous random variables. The moment generating function and the characteristic function also exist for discrete random variables. In this case we have (with $p_D(d_n)$ the probability mass function):

$$M_D(s) = \mathrm{E}[e^{sD}] = \sum_n e^{sd_n} p_D(d_n) \tag{3.57}$$

$$\varphi_D(s) = \mathrm{E}[e^{isD}] = \sum_n e^{isd_n} p_D(d_n) \tag{3.58}$$

Apart from these functions, discrete random variables can also be characterized with using the *probability generating function*:

$$G_D(s) = \mathrm{E}[s^D] = \sum_n s^{d_n} p_D(d_n) \tag{3.59}$$

This transformation is related to the Z-transform and only exists for discrete variables. Note that $G_D(0) = \Pr(d = 0)$ and $G_D(1) = 1$. Some useful properties (Grimmet and Stirzaker, 1982):

$$G_{D_1+D_2}(s) = G_{D_1}(s) G_{D_2}(s) \tag{3.60}$$

$$E[D] = \left. \frac{dG_D(s)}{ds} \right|_{s=1} \tag{3.61}$$

$$E[D^2] - E[D] = \left. \frac{d^2 G_D(s)}{ds^2} \right|_{s=1} \tag{3.62}$$

## 3.6 Some well known probability distributions and their properties

There are many different models for probability distributions. Which model to choose for which variable depends on its type. Many hydrological variables are strictly positive (e.g. hydraulic conductivity, rainfall intensity) and require therefore probability density functions (pdfs) for positive variables. Also, certain variables, such as the number of rain

storms arriving in a fixed interval, are discrete, while others are continuous. In this section we will provide an overview of a number of probability density functions and their properties. Table 3.2 gives the pdfs and expressions for the mean and the variance in terms of the distribution parameters. Also given in the last column are a number of hydrological variables that may be described with the various pdfs. Figure 3.5 provides plots for a number of the continuous pdfs of Table 3.2 and Table 3.3 gives expressions for the associated generating functions.

Some words should be spent on the most famous of distributions: the normal or Gaussian distribution. This is the distribution that naturally arises for random variables that are themselves the result of the sum of a large number of independent events. The underlying rule is called the *Central Limit Theorem* and it reads:

Let $Z_1$, $Z_2$,…, $Z_N$ be a set of $N$ independent random variables that have an arbitrary probability distribution with mean $\mu_i$ and variance $\sigma_i^2$. Then the normal form random variable

$$Y_{\text{norm}} = \frac{\sum_{i=1}^{N} Z_i - \sum_{i=1}^{N} \mu_i}{\sqrt{\sum_{i=1}^{N} \sigma_i^2}} \tag{3.63}$$

has a limiting cumulative distribution function that approaches the normal (standard Gaussian) distribution

Typically error distributions, very relevant to stochastic hydrology, are Gaussian distributed, because errors are often the sum of many independent error sources. If $N$ is very large and the individual variables are mildly dependent then it turns out in practice that the summed variable is approximately Gaussian. An example of a hydrological variable that can often be described with a Gaussian distribution is a freely fluctuating groundwater level $H_t$ that fluctuates under pulses of precipitation surplus $P_t$ (precipitation minus evapotranspiration). Using a simple water balance of the soil column it is possible to write the groundwater level at some time t as the sum of precipitation surplus events (Knotters and Bierkens, 1999):

$$h_t = \sum_{k=0}^{M} \alpha^k P_{t-k} \tag{3.64}$$

If we view the rainfall surplus series as random variables, then the groundwater level will be approximately Gaussian distributed if $M$ and $\alpha$ are large enough. Table 3.4 provides a cumulative distribution table $F\chi(x) = \Pr[\chi \leq x]$ for the standard normal random variable $\chi$, with mean zero ($\mu_Z$=0) and standard deviation equal to 1 ($\sigma_Z$=1). A number of often used quantiles of the distribution are given in Table 3.5.

Another distribution often used in hydrology that is worth mentioning is the lognormal distribution. A variable $Z$ has a lognormal or logGaussian distribution if its natural logarithm $Y=\ln Z$ is Gaussian distributed. A well-known example is hydraulic conductivity. When sampled randomly in space, hydraulic conductivity obeys a lognormal distribution (Freeze, 1975). This assumption has been reconfirmed by many observations thereafter. Some useful transformation formulae between the means and variances of the normal and the lognormal distribution are:

$$\mu_Z = e^{\mu_Y + \sigma_Y^2/2} \tag{3.65}$$

$$\sigma_Z^2 = e^{2\mu_Y + \sigma_Y^2}(e^{\sigma_Y^2} - 1) \tag{3.66}$$

*Table 3.2 Some well known discrete and continuous probability density functions*

| **Distribution** | **Probability density/mass** | **Expected value** | **Variance** | **Example of Hydrological application** |
|---|---|---|---|---|
| Binomial $B(N,p)$ | $\binom{N}{n} p^n (1-p)^{N-n}$ $n = 0,1,2,..,N$ | $Np$ | $Np(1-p)$ | The number $n$ of flood events with probability $p$ occurring in $N$ time steps |
| Geometric $G(p)$ | $(1-p)^{n-1} p$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ | The number of time steps until a flood event with probability $p$ occurs. |
| Poisson $P(\lambda)$ | $\frac{e^{-\lambda}\lambda^n}{n!}$ | $\lambda$ | $\lambda$ | The number of rain storms occurring in a given time period. |
| Uniform $U(a,b)$ | $\frac{1}{b-a}$ $a \geq z \geq b$ | $\frac{b-a}{2}$ | $\frac{(b-a)^2}{12}$ | (Non-informative) prior distribution of a hydrological parameter provided to a parameter estimation method |
| Exponential $E(\lambda)$ | $\lambda e^{-\lambda z}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | The time between two rain storms |
| Gaussian/Normal $N(\mu,\sigma)$ | $\frac{1}{\sqrt{2\pi}\sigma} e^{-[\frac{1}{2}(z-\mu)^2/\sigma^2]}$ | $\mu$ | $\sigma^2$ | Many applications: prior distribution for parameter optimisation; modelling of errors; likelihood functions |
| logNormal $L(\mu,\sigma)$ | $\frac{1}{\sqrt{2\pi}\sigma_z} e^{-[\frac{1}{2}(\ln z-\mu)^2/\sigma^2]}$ | $\mu$ | $\sigma^2$ | Hydraulic conductivity |
| Gamma $\Gamma(n,\lambda)$ (note: $n \in \Re$) | $\frac{\lambda^n}{\Gamma(n)} z^{n-1} e^{-\lambda z}$ | $\frac{n}{\lambda}$ | $\frac{n}{\lambda^2}$ | Sum of $n$ independent random variables that are exponentially distributed with parameter $\lambda$; Instantaneous unit hydrograph of $n$ linear reservoirs in series; pdf of travel times in a catchment; |

| | | | | |
|---|---|---|---|---|
| | | | | very flexible distribution for strictly positive variables. |
| Beta $\beta(p,q)$ | $\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} z^{p-1}(1-z)^{q-1}$<br>$p>0, q>0, \quad 0 \le z \le 1$ | $\frac{p}{p+q}$ | $\frac{pq}{(p+q)^2(p+q+1)}$ | Very flexible distribution for variables with upper and lower boundaries; used as a priori distribution in Bayesian analysis and parameter estimation |
| Gumbel $G(a,b)$ (Extreme value distribution Type I) | $be^{-b(z-a)}\exp(-e^{-b(z-a)})$ | $a+\frac{0.5772}{b}$ | $\frac{\pi^2}{6b^2}$ | Yearly maximum discharge used for design of dams and dikes |
| Weibull $W(\lambda,\beta)$ (Extreme value distribution type III) | $\lambda^\beta \beta z^{\beta-1}\exp[-(\lambda x)^\beta]$ | $\frac{1}{\lambda}\Gamma(1+\frac{1}{\beta})$ | $\frac{1}{\lambda^2}(A-B)$<br>$A=\Gamma(1+\frac{2}{\beta})$<br>$B=\left(\Gamma(1+\frac{1}{\beta})\right)^2$ | Yearly minimum discharge used in low flow analysis. |

*Table 3.3 Characteristic functions for a number of probability distributions*

| **Distribution** | **Probability generating function** | **Moment generating function** | **Characteristic function** |
|---|---|---|---|
| Binomial $B(n,p)$ | $(1-p+ps)^n$ | $(1-p+pe^s)^n$ | $(1-p+pe^{is})^n$ |
| Geometric $G(p)$ | $\frac{ps}{1-(1-p)s}$ | $\frac{pe^s}{1-(1-p)e^s}$ | $\frac{pe^{is}}{1-(1-p)e^{is}}$ |
| Poisson $P(\lambda)$ | $e^{\lambda(s-1)}$ | $e^{\lambda(e^s-1)}$ | $e^{\lambda(e^{is}-1)}$ |
| Uniform $U(a,b)$ | - | $\frac{e^{bs}-e^{as}}{s(b-a)}$ | $\frac{e^{ibs}-e^{ais}}{is(b-a)}$ |
| Exponential $E(\lambda)$ | - | $\frac{\lambda}{\lambda-s}$ | $\frac{\lambda}{\lambda-is}$ |
| Gaussian/normal $N(\mu,\sigma)$ | - | $e^{\mu s+\frac{1}{2}\sigma^2 s^2}$ | $e^{i\mu s-\frac{1}{2}\sigma^2 s^2}$ |
| Gamma $\Gamma(n,\lambda)$ | - | $\left(\frac{\lambda}{\lambda-s}\right)^n$ | $\left(\frac{\lambda}{\lambda-is}\right)^n$ |

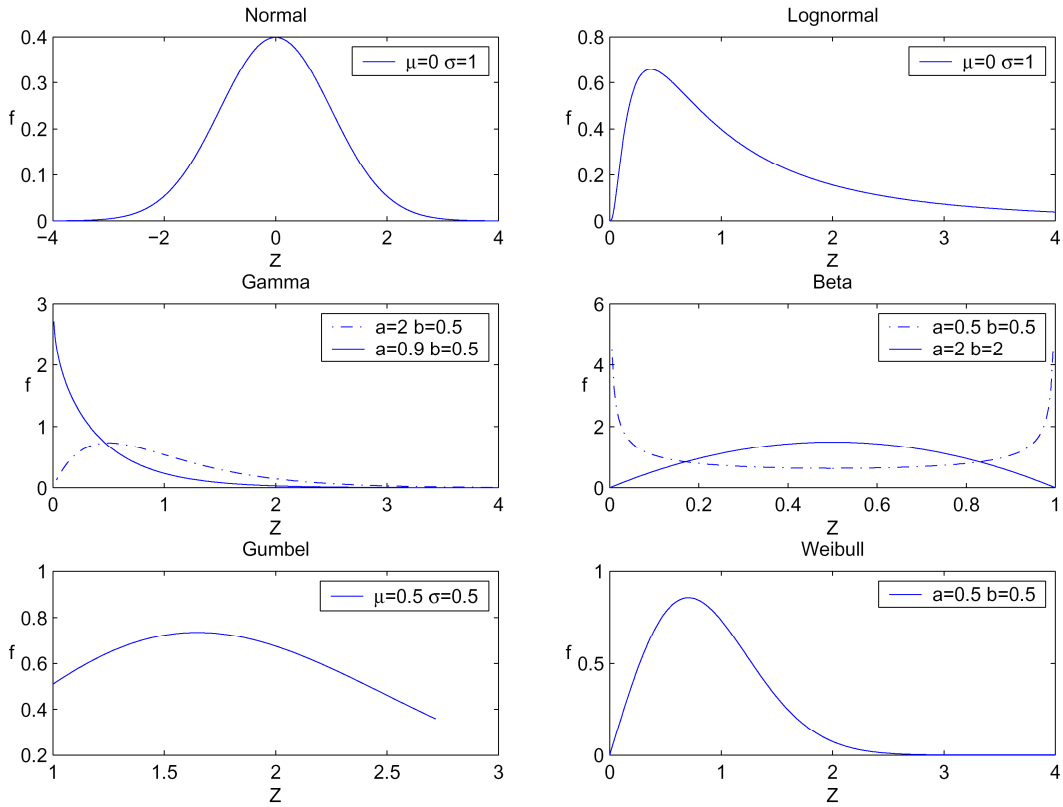



*Figure 3.5 Plots for some well known probability density functions*

## 3.7 Two or more random variables

If two random variables $Z$ and $Y$ are simultaneously considered (e.g. hydraulic conductivity and porosity) we are interested in the bivariate probability density function $f_{ZY}(z,y)$ that can be used to calculate the probability that both $Z$ and $Y$ are between certain limits:

$$\Pr[z_1 < Z \le z_2 \cap y_1 < Y \le y_2] = \int_{y_1}^{y_2}\int_{z_1}^{z_2} f_{ZY}(z,y)\,dz\,dy \tag{3.67}$$

A more formal definition of the bivariate pdf is:

$$f_{ZY}(z,y) = \lim_{\substack{dz\to 0\\ dy\to 0}} \frac{\Pr[z_1 < Z \le z_2 \cap y_1 < Y \le y_2]}{dz\,dy} \tag{3.68}$$

The bivariate cumulative distribution function is $F_{ZY}(z,y)$ is given by:

$$F_{ZY}(z,y) = \Pr[Z \le z \cap Y \le y] \tag{3.69}$$

The density function and the distribution function are related as:

$$F_{ZY}(z,y) = \int_{-\infty}^{y}\int_{-\infty}^{z} f_{ZY}(z,y)dzdy \tag{3.70}$$

$$f_{ZY}(z,y) = \frac{\partial^2 F_{ZY}(z,y)}{\partial z \partial y} \tag{3.71}$$

The *marginal distribution* of $Z$ can be obtained from the bivariate distribution by integrating out the $Y$ variable:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{ZY}(z,y)dy \tag{3.72}$$

The *conditional probability* can be obtained from the distribution function as:

$$F_{Z|Y}(z \mid y) = \Pr\{Z \le z \mid Y = y) \tag{3.73}$$

which thus provides the probability that Z is smaller than $z$ given that $Y$ takes the value of $y$. The conditional pdf can be derived from this by differentiation:

$$f_{Z|Y}(z \mid y) = \frac{dF_{Z|Y}(z \mid y)}{dz} \tag{3.74}$$

The conditional density satisfies:

$$\int_{-\infty}^{\infty} f_{Z|Y}(z \mid y)dz = 1 \tag{3.75}$$

The relation between the bivariate pdf and the conditional pdf is given by (see also 3.2.3):

$$f_{ZY}(z,y) = f_{Z|Y}(z \mid y) f_Y(y) = f_{Y|Z}(y \mid z) f_Y(z) \tag{3.76}$$

The *total probability theorem* in terms of density functions reads:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|Z}(y \mid z) f_Z(z)dz \tag{3.77}$$

and *Bayes' theorem* becomes:

$$f_{Z|Y}(z \mid y) = \frac{f_{ZY}(z,y)}{f_Y(y)} = \frac{f_{Y|Z}(y \mid z) f_Z(z)}{\int_{-\infty}^{\infty} f_{Y|Z}(y \mid z) f_Z(z) dz} \tag{3.78}$$

A measure of *linear* statistical dependence between random variables $Z$ and $Y$ is the covariance is defined as:

$$\mathrm{COV}[Z,Y] = \mathrm{E}[(Z-\mu_Z)(Y-\mu_Y)] = \int_{=\infty}^{\infty}\int_{-\infty}^{\infty}(z-\mu_Z)(y-\mu_Y) f_{ZY}(z,y) dz dy \tag{3.79}$$

The covariance between two data sets can be estimated using Equation (2.7), where we have to replace the number of observations $n$ by $n$-1 if the respective mean values of $Z$ and $Y$ have been estimated too. The following relations between variance and covariance exist ($a$ and $b$ constants):

$$\begin{aligned} \mathrm{VAR}[aZ+bY] &= a^2\,\mathrm{VAR}[Z] + b^2\,\mathrm{VAR}[Y] + 2ab\,\mathrm{COV}[Z,Y] \\ \mathrm{VAR}[aZ-bY] &= a^2\,\mathrm{VAR}[Z] + b^2\,\mathrm{VAR}[Y] - 2ab\,\mathrm{COV}[Z,Y] \end{aligned} \tag{3.80}$$

Often the correlation coefficient is used as a measure of linear statistical dependence:

$$\rho_{ZT} = \frac{\mathrm{COV}[Z,Y]}{\sigma_Z\,\sigma_Y} \tag{3.81}$$

The following should be noted. If two random variables are statistically independent they are also uncorrelated: COV[$Y$,$Z$]=0 and $\rho_{YZ}$ = 0. However a zero correlation coefficient does not necessarily mean that $Y$ and $Z$ are statistically independent. The covariance and correlation coefficient only measure *linear* statistical dependence. If a non-linear relation exists, the correlation may be 0 but the two variables may still be statistically dependent, as is shown in the lower right figure of Figure 2.13.

Figure 3.6 shows surface plots and isoplots of the *bivariate Gaussian Distribution*:

$$f_{ZY}(z,y) = \frac{1}{2\pi\sigma_Z\sigma_Y\sqrt{1-\rho_{ZY}^2}} \cdot \exp\left(-\left[\frac{1}{2(1-\rho_{ZY}^2)}\right] \cdot \left[\left(\frac{Z-\mu_Z}{\sigma_Z}\right)^2 + \left(\frac{Z-\mu_Y}{\sigma_Y}\right)^2 - 2\left(\frac{Z-\mu_Z}{\sigma_Z}\right)\left(\frac{Z-\mu_Y}{\sigma_Y}\right)\right]\right) \tag{3.82}$$

where the left plots show the situation for which $\rho_{ZT} = 0$ and the right plots for which $\rho_{ZT} = 0.8$. We can see that the isolines form an ellipse whose form is determined by the ratio $\sigma_Z / \sigma_Y$ and its principle direction by $\rho_{ZT}$.

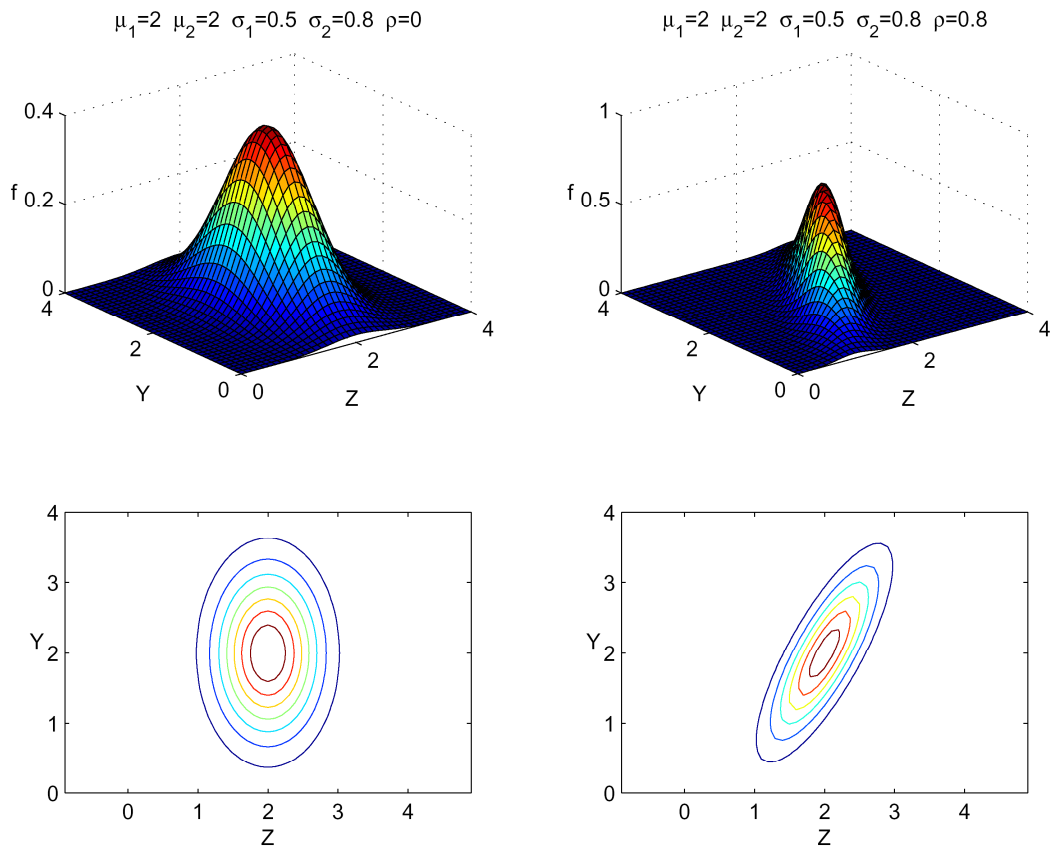



*Figure 3.6 Surface plots and isoplots of the bivariate Gaussian distribution of independent (left) and dependent random variables Z and Y.*

From the relationship $f_{Z|Y}(z \mid y) = f_{ZY}(z, y) / f_Y(y)$ we can derive that the conditional Gaussian density $f_{Z|Y}(z \mid Y = y)$ has a Gaussian distribution $N(\mu, \sigma)$ with

$$\mu = \mu_Z + \rho_{ZY} \frac{\sigma_Z}{\sigma_Y}(y - \mu_Y) \text{ and } \sigma^2 = \sigma_Z^2(1 - \rho_{ZY}^2) \tag{3.83}$$

From these expressions we learn that if we have two dependent random variables and we measure one of them (in this case Y) that our *a priori* distribution is updated to a new *a posteriori* distribution and that our uncertainty about $Z$ (through the variance) has decreased. Of course, if both variables are independent we see that $\mu = \mu_Z$ and $\sigma^2 = \sigma_Z^2$.

Finally, a useful property of the Gaussian distribution is that any linear combination of Gaussian random variables (with $a_k$ deterministic weights)

$$Y = \sum_{i=1}^{N} a_i Z_i \tag{3.84}$$

with mean $\mu_i$, $i$=1,..,$N$, variance $\sigma_i^2$, $i$=1,..,$N$ and $\rho_{ij}$, $i,j$=1,..,$N$ correlation coefficients between random variables $i$ and $j$, is itself Gaussian distributed $N(\mu_Y, \sigma_Y)$ with mean and variance given by:

$$\mu_Y = \sum_{i=1}^{N} a_i \mu_i \tag{3.85}$$

$$\sigma_Y^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \rho_{ij} \sigma_i \sigma_i \tag{3.86}$$

We will end this chapter with some notes on multivariate distributions. All the relationships given here for bivariate distributions can be easily generalised to probability distributions of multiple variables (multivariate distributions): $f_{Z_1..Z_N}(z_1,..,z_N)$.

A distribution often used in stochastic hydrology to parameterise multivariate distributions is the *multivariate Gaussian distribution*. It can be defined as follows: Let $Z_1,Z_2,\ldots,Z_N$ be a collection of $N$ random variables that are collectively Gaussian distributed with mean $\mu_i$, $i$=1,..,$N$, variance $\sigma_i^2$, $i$=1,..,$N$ and $\rho_{ij}$, $i,j$=1,..,$N$ correlation coefficients between variables $i$ and $j$. We define a stochastic vector $\mathbf{z} = (Z_1,Z_2,\ldots,Z_N)^{\mathrm{T}}$ and a vector of mean values $\boldsymbol{\mu} = (\mu_1,\mu_2,\ldots,\mu_N)^{\mathrm{T}}$ (superscript $^{\mathrm{T}}$ stands for transpose). The covariance matrix $\mathbf{C_{zz}}$ is defined as $E[(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}]$. The covariance matrix is a $N\times N$ matrix of covariances. Element $C_{ij}$ of this matrix is equal to $\rho_{ij}\sigma_i\sigma_j$. The multivariate Gaussian probability density function is given by:

$$f_{Z_1\ldots Z_N}(z_1,\ldots,z_N) = \frac{1}{(2\pi)^{N/2} |\mathbf{C_{zz}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \mathbf{C_{zz}}^{-1}(\mathbf{z}-\boldsymbol{\mu})} \tag{3.87}$$

with $|\mathbf{C_{zz}}|$ the determinant of the covariance matrix and $\mathbf{C_{zz}}^{-1}$ the inverse.

*Table 3.4 Cmulative distribution table for the standard normal (Gaussian) distribution N(0,1);* $F\chi(x) = \Pr[\chi \leq x]$, *e.g.* $F\chi(0.61)=0.7921$; *note* $F\chi(-x) = 1 - F\chi(x)$

| $z$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |

*Table 3.5 Selected quantiles of the standard normal distribution N(0,1);note that* $q_{1-p} = -q_p$

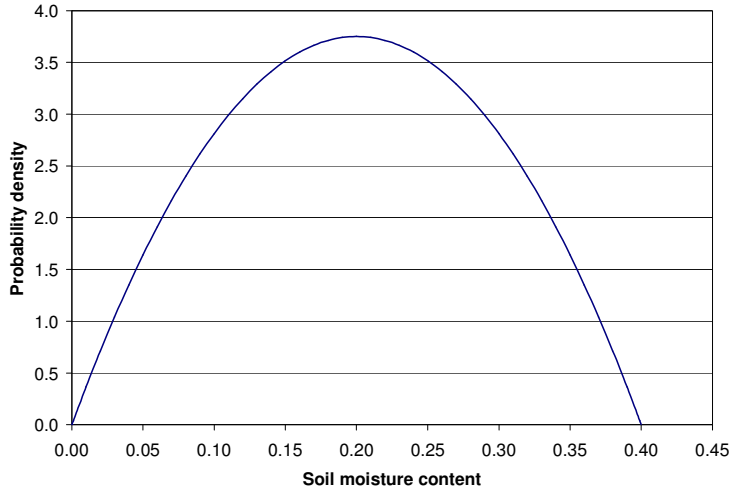
| $p$ | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 |
|---|---|---|---|---|---|---|
| $q_p$ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |

## 3.8 Questions

3.1 Consider the intensity of one-hour rainfall which is assumed to follow an exponential distribution: $f_Z(z) = \lambda e^{-\lambda z}$. With λ=0.1, calculate: Pr[Z>20].

3.2 Consider the following probability density function describing average soil moisture content $Z$ in the root zone of some soil (see also the Figure):

$$f_Z(z) = -93.75z^2 + 37.5z \qquad 0 \le z \le 0.4$$




a) Give the expression for the cumulative probability distribution.
b) Calculate the probability that average soil moisture exceeds 0.30.
c) Calculate the mean $\mu_Z$ and the variance $\sigma_Z^2$ of soil moisture content.

3.3 Hydraulic conductivity at some unobserved location is modelled with a log-normal distribution. The mean of $Y$=ln$K$ is 2.0 and the variance is 1.5. Calculate the mean and the variance of $K$?

3.4 Hydraulic conductivity for an aquifer has a lognormal distribution with mean 10 m/d and variance 200 $m^2/d^2$. What is the probability that at a non-observed location the conductivity is larger than 30 m/d?

3.5 Based on a geological analysis we extracted the following probabilities of texture classes occurring in some aquifer: Pr[sand]=0.7, Pr[clay]=0.2, Pr[peat]=0.1. The following table shows the probability distributions of conductivity classes for the three textures:

*Table: probabilities of conductivity classes (m/d) for three texture classes*

| **Texture** | **$10^{-3}$** | **$10^{-2}$** | **$10^{-1}$** | **$10^{0}$** | **10** | **20** | **50** | **100** |
|---|---|---|---|---|---|---|---|---|
| Sand | 0 | 0 | 0 | 0.1 | 0.4 | 0.3 | 0.1 | 0.1 |
| Clay | 0.3 | 0.4 | 0.2 | 0.1 | 0 | 0 | 0 | 0 |
| Peat | 0.1 | 0.3 | 0.3 | 0.2 | 0.1 | 0 | 0 | 0 |

Calculate the probability distribution of conductivity for the entire aquifer (use the total probability theorem for this).

3.6 Consider two random variables $Z_1$ and $Z_2$ with mean 10 and 25 and variances 300 and 450 respectively. The correlation coefficient between both variables equals 0.7.

a. Calculate the covariance between $Z_1$ and $Z_2$.
b. Calculate the expected value of $Y = Z_1 + Z_2$.
c. Calculate the variance of $Y = Z_1 + Z_2$.

3.7 For the same two variables of 3.5: Assume that they are bivariate Gaussian distributed and:

a. Calculate the probability $Pr[Z_1 < 30]$
b. Calculate the probability $Pr[Z_1 + Z_2 < 50]$
c. Write the expression for the probability $Pr[Z_1 < 30 \cap Z_2 < 40]$
d. Write the expression for the probability $Pr[Z_1 < 30 \cup Z_2 < 40]$

3.7* Derive equations (3.48), (3.49), (3.51) and (3.54).

# 4. Hydrological statistics and extremes

## 4.1 Introduction

The field of statistics is extensive and all of its methods are probably applicable to hydrological problems. Some elementary *descriptive statistics* was already treated in chapter 2, while geostatistics will be treated in chapter 7. The field of hydrological statistics is mainly concerned with the *statistics of extremes*, which will be the main topic in this chapter. In the first part we will mainly look at the analysis of maximum values, in particular to flood estimation. In the second part we will cover some of the statistical tests that should be applied to series of maxima before any further analysis on extremes is possible.

## 4.2 The analysis of maximum values

### 4.2.1 Flow duration curve

We start out with a time series of more then 100 years of daily averaged flow data of the River Rhine at Lobith (approximately) the location where the Rhine enters the Netherlands). Figure 4.1 shows a plot of this time series. To say something about the *flow frequency* a so -called *flow duration curve* can be made. Such a curve is shown in Figure 4.2. The flow frequency may be informative in certain cases, e.g. for 5 % of the time the discharge will be above 5000 $m^3/s$. However, when it comes to building dams or dikes, only the frequency of individual floods or discharge peaks are important. To elaborate: if one needs to find the required height of a dike, then the maximum height of a peak is most important and to a lesser extent its duration. So our goal is to analyse flood peaks. Generally, two different methods are used to convert a series of discharges into a series of peaks: one is based on identification of the maximum discharge per year and the other on analysing all discharge values above a given threshold. We will concentrate on the first method here, and briefly treat the second.

### 4.2.2 Recurrence times

To obtain a series of maximum values we simply record for each year the largest discharge measured. This results in the same number of maximum discharges as recorded years. Sometimes, if maximum discharges occur in one part of the season (winter), it may be wise to work with hydrological years (April 1 to March 31$^{st}$ in Northern Europe). This has been done with the Rhine discharges. The plot with maxima is shown in Figure 4.2. To further analyse these maximum values the following assumptions are made:

1. the maximum values are realisations of independent random variables;
2. there is no trend in time;
3. the maximum values are identically distributed.

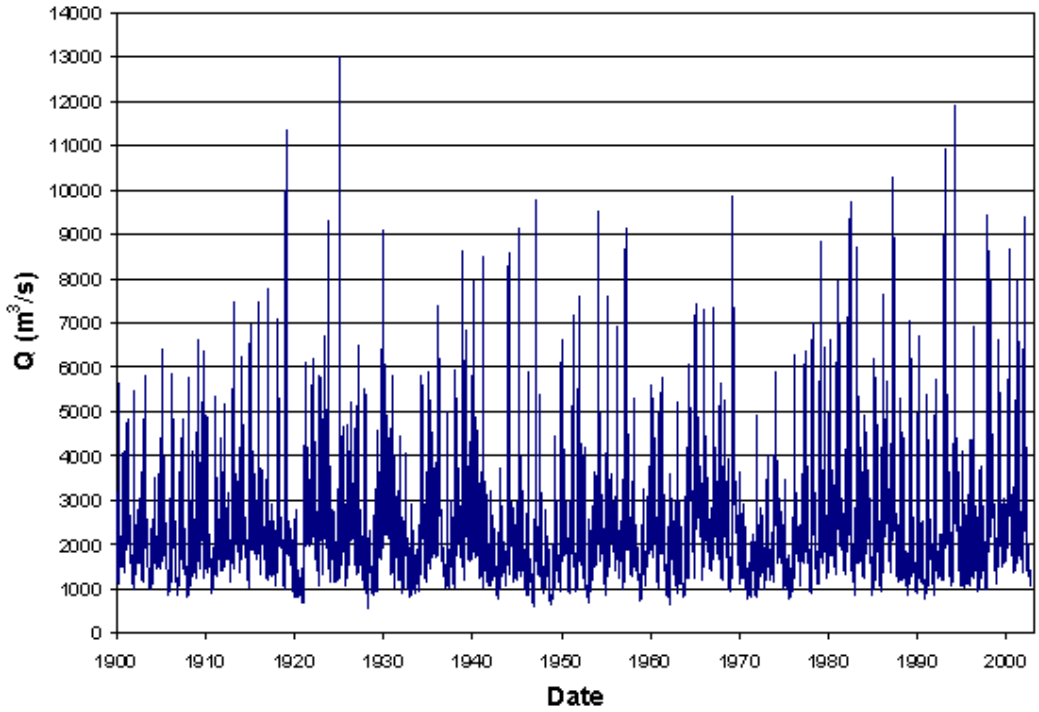



*Figure 4.1 Daily average discharges ($m^3/s$) of the Rhine river at Lobith*

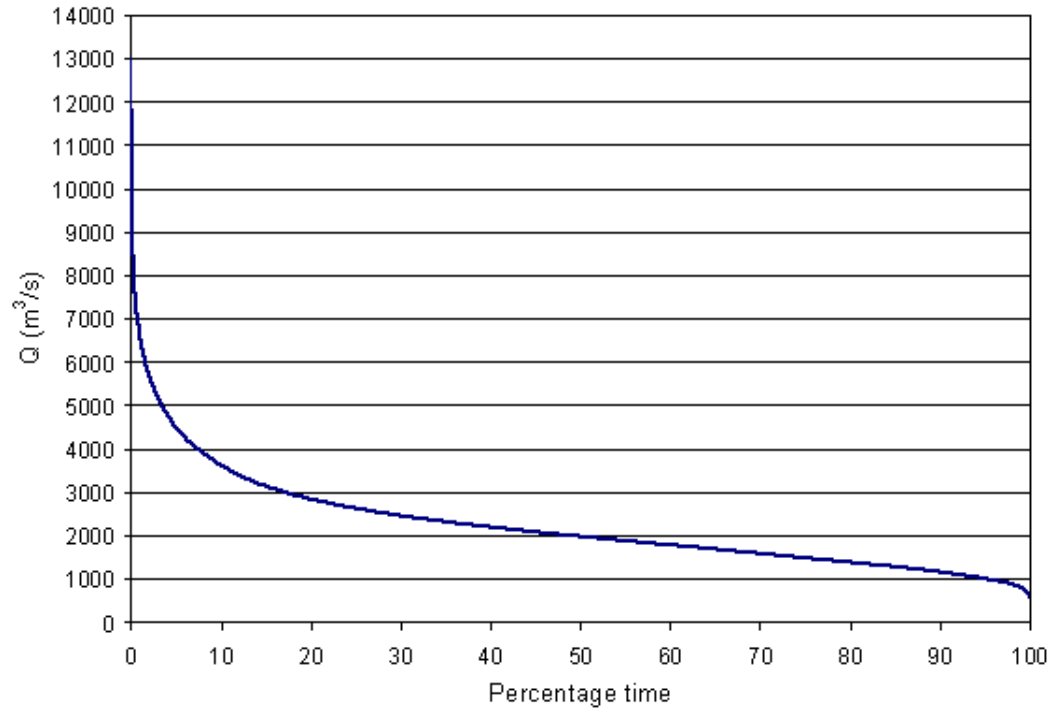



*Figure 4.2 Flow duration curve of the Rhine river at Lobith*

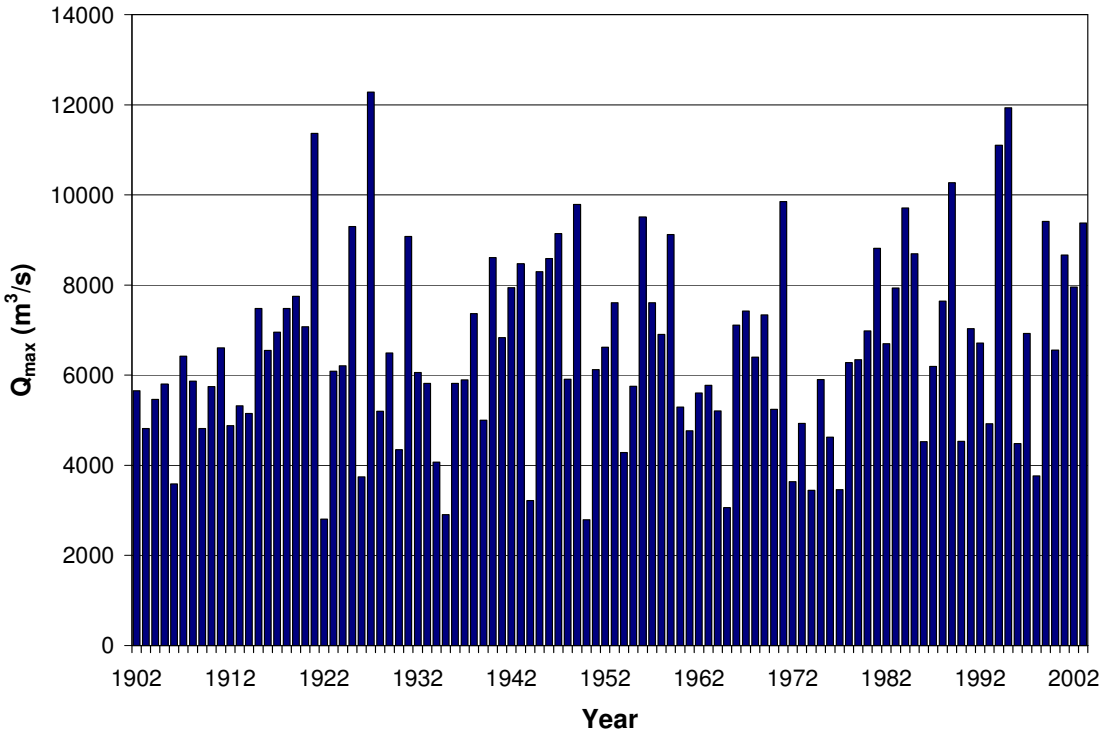



*Figure 4.3 yearly maximum values of daily average discharge of the Rhine river at Lobith*

In section 4.3 we will describe some statistical tests to check whether these assumptions are likely to be valid.

If the maximum values are indeed independent random variables, the first step of an analysis would be to calculate the cumulative frequency distribution and use this as an estimate for the (cumulative) distribution function $F(y) = \Pr(Y \leq y)$. The probability that a certain value $y$ is exceeded by the maximum values is given by the function $P(y) = \Pr(Y > y) = 1 - F(y)$. Finally, the *recurrence time* or *return period* $T(y)$ when applied to yearly maximum values is the mean (or expected) number of years between two flood events that are larger than $y$ and is calculated using either $F(y)$ or $P(y)$ as:

$$T(y) = \frac{1}{P(y)} = \frac{1}{1 - F(y)} \tag{4.1}$$

To estimate the recurrence time, first the cumulative distribution function must be estimated. The most common way of doing this is by arranging the values in ascending order and assigning rank numbers from small to large: $y_i, i=1,..,N$, with $N$ the number of years. Ordinates of the cumulative distribution function are then estimated by:

$$\hat{F}(y_i) = \frac{i}{N+1} \tag{4.2}$$

There have been other estimators proposed (e.g. Gringorten: $\hat{F}(y_i) = (i-0.44)/(i+0.12)$), but for larger $N$ differences between these are small. Table 4.1 shows (part of) the analysis performed on the maximum discharge values for the Rhine river. As can be seen, the maximum recurrence period that can be analysed with the raw data is $N$+1 years (103 in the example). However, many dams and dikes are designed at larger return periods. For instance, river dikes in the Netherlands are based on 1:1250 year floods. To be able to predict flood sizes belonging to these larger recurrence times we need to somehow extrapolate the record. One way of doing this is to fit some probability distribution to the data and use this probability distribution to predict the magnitudes of floods with larger recurrence times. As will be shown in the next section, the Gumbel distribution is a good candidate for extrapolating maximum value data.

*Table 4.1 Analysis of maximum values of Rhine discharge at Lobith for recurrence time*

| $Y$ | Rank | $F(y)$ | $P(y)$ | $T(y)$ |
|---|---|---|---|---|
| 2790 | 1 | 0.00971 | 0.99029 | 1.0098 |
| 2800 | 2 | 0.01942 | 0.98058 | 1.0198 |
| 2905 | 3 | 0.02913 | 0.97087 | 1.0300 |
| 3061 | 4 | 0.03883 | 0.96117 | 1.0404 |
| 3220 | 5 | 0.04854 | 0.95146 | 1.0510 |
| 3444 | 6 | 0.05825 | 0.94175 | 1.0619 |
| 3459 | 7 | 0.06796 | 0.93204 | 1.0729 |
| . | . | . | . | . |
| . | . | . | . | . |
| 9140 | 90 | 0.87379 | 0.12621 | 7.9231 |
| 9300 | 91 | 0.88350 | 0.11650 | 8.5833 |
| 9372 | 92 | 0.89320 | 0.10680 | 9.3636 |
| 9413 | 93 | 0.90291 | 0.09709 | 10.3000 |
| 9510 | 94 | 0.91262 | 0.08738 | 11.4444 |
| 9707 | 95 | 0.92233 | 0.07767 | 12.8750 |
| 9785 | 96 | 0.93204 | 0.06796 | 14.7143 |
| 9850 | 97 | 0.94175 | 0.05825 | 17.1667 |
| 10274 | 98 | 0.95146 | 0.04854 | 20.6000 |
| 11100 | 99 | 0.96117 | 0.03883 | 25.7500 |
| 11365 | 100 | 0.97087 | 0.02913 | 34.3333 |
| 11931 | 101 | 0.98058 | 0.01942 | 51.5000 |
| 12280 | 102 | 0.99029 | 0.00971 | 103.0000 |

### 4.2.2 Maximum values and the Gumbel distribution

The added assumption with regard to the record of maximum values is that they are random variables that follow a Gumbel distribution, i.e. the following pdf and distribution function:

$$f_Z(z) = be^{-b(z-a)} \exp(-e^{-b(z-a)}) \tag{4.3}$$

$$F_Z(z) = \exp(-e^{-b(z-a)}) \tag{4.4}$$

Here we will show that the distribution of the maximum values is likely to be a Gumbel distribution. Let $Z_1$, $Z_2$,…,$Z_N$ be $N$ independent and identically distributed variables with distribution function $F_Z(z)$. Let $Y$ be the maximum of this series $Y = \max(Z_1, Z_2,\ldots,Z_N)$. The distribution function of $Y$ is then given by:

$$\begin{aligned} F_Y(y) = \Pr(Y \le y) &= \Pr(Z_1 \le y, Z_2 \le y, .., Z_N \le y) \\ &= \Pr(Z_1 \le y) \cdot \Pr(Z_2 \le y) \cdots \Pr(Z_N \le y) \\ &= \left(F_Z(y)\right)^N \end{aligned} \tag{4.5}$$

We cannot derive the probability distribution from (4.5) alone, because $\left(F_Z(y)\right)^N$ is a so called degenerative distribution: if $N \to \infty$ then $\left(F_Z(y)\right)^N \to 0$. To obtain a non-degenerative distribution of $Y$ we need to reduce and normalize the maximum values. Now suppose that the $Z$ variables have an exponential distribution: $F_Z(z) = 1 - \exp(-bz)$ with $b > 0$. We apply the following transformation of the maximum $Y$:

$$X = b\left(Y - \frac{\ln N}{b}\right) \tag{4.6}$$

The distribution function of this variable becomes:

$$\begin{aligned} \Pr(X \le x) &= \Pr(b(Y - \frac{\ln N}{b}) \le x) = \Pr(Y \le \frac{1}{b}(x + \ln N)) \\ &= \left(F_Z(\frac{1}{b}(x + \ln N))\right)^N = \left(1 - \exp(-(x + \ln N))\right)^N \\ &= \left(1 - \frac{\exp(-x)}{N}\right)^N \end{aligned} \tag{4.7}$$

Taking the limit yields:

$$\lim_{N\to\infty} \Pr(X \le x) = \lim_{N\to\infty}\left(1 - \frac{\exp(-x)}{N}\right)^N = \exp(-e^{-x}) \tag{4.8}$$

which is the normalized Gumbel distribution. So the limit distribution of $X$ if $Z$ has exponential distribution is $\exp(-e^{-x})$. If we define $a = \log N / b$ then $X$=$b$($Y$-a). So for large $N$ we have:

$$\begin{aligned} &\Pr[X \le x] = \exp(-e^{-x}) \\ &\quad \Rightarrow \Pr[b(Y-a) \le x] = \exp(-e^{-x}) \\ &\quad \Rightarrow \Pr[Y \le y] = \Pr[b(Y-a) \le b(y-a)] = \exp(-e^{-b(y-a)}) \end{aligned} \tag{4.9}$$

So, finally we have shown that the maximum $Y$ of $N$ independently and identically exponentially distributed random variables $Z_I$ has a Gumbel distribution with parameters $a$ and $b$. For finite $N$ this distribution is also used, where $a$ and $b$ are found through fitting the distribution to the data.

It can be shown in a similar way as for the exponential distribution that the Gumbel limit distribution is also found for the maximum of independent variables with the following distributions: Gaussian, logGaussian, Gamma, Logistic and Gumbel itself. This is the reason why the Gumbel distribution has been found to be a suitable distribution to model probabilities of maximum values, such as shown in Figure 4.3. Of course, we have to assume that these maximum values themselves are obtained from independent variables $Z$ within a year. This is clearly not the case as is shown in Figure 4.1. However, as long as the maximum values are approximately independent of other maximum values, it turns out that the Gumbel distribution provides a good model if $a$ and $b$ can be fitted.

### 4.2.3 Fitting the Gumbel distribution

To be able to use the Gumbel distribution to predict larger recurrence times it has to be fitted to the data. There are several ways of doing this: 1) using Gumbel probability paper; 2) linear regression; 3) the methods of moments; 4) maximum likelihood estimation. The first three methods will be shown here.

*Gumbel probability paper*
Using equation (4.4) it can be shown that the following relation holds between a given maximum $y$ and the distribution function $F(y)$

$$y = a - \frac{1}{b}\ln(-\ln F(y)) \tag{4.10}$$

So by plotting -ln(-ln$F(y)$) against the maximum values $y$ we should obtain a straight line if the maximum values follow a Gumbel distribution. If Gumbel probability paper is used the double log-transformation of the distribution function is already implicit in the scale of the abscissa. On this special paper plotting maximum values $y_i$ with rank $i$ against $T(y_i) = 1 - 1/[i/(N+1)]$ will yield a straight line if they are Gumbel distributed. The Rhine discharge maximum values of table 4.1 have been plotted in Figure 4.4. Once these data have been plotted a straight line can be fitted by hand, from which the parameters $a$ and $1/b$ can calculated. The fitted line is also shown in Figure 4.4.

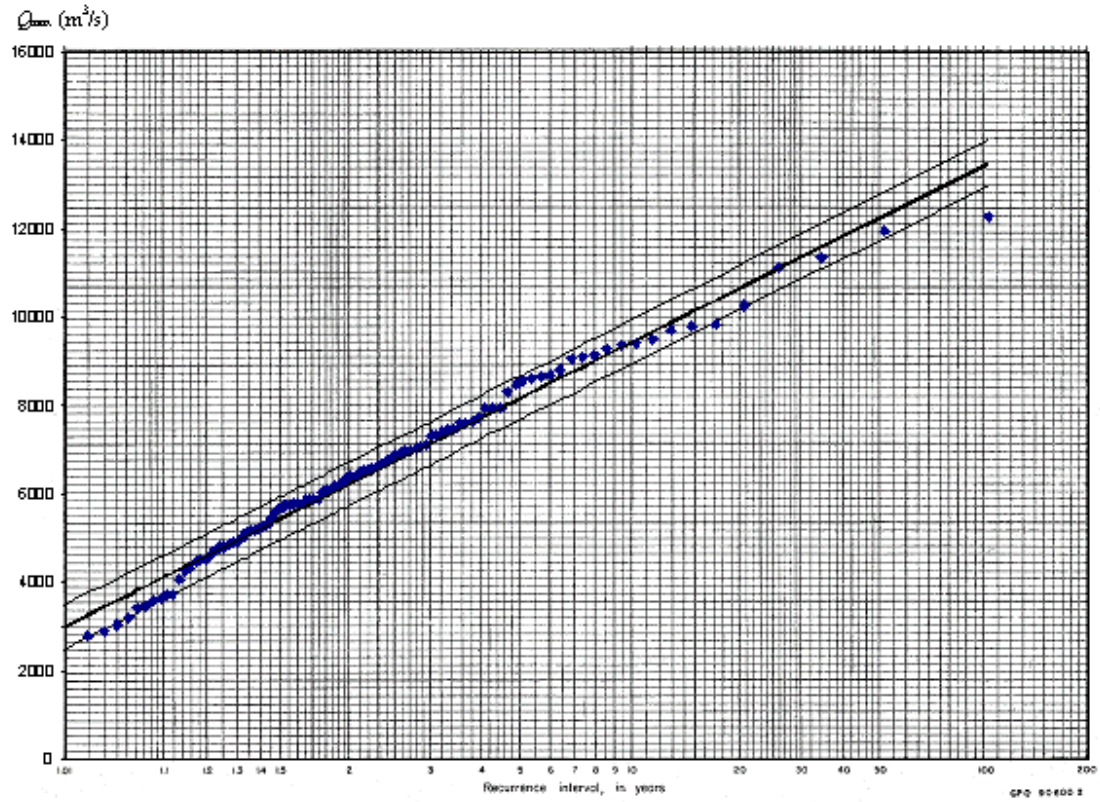



*Figure 4.4 Gumbel probability plot of Rhine discharge maximum values and a fitted linear relation;* $\hat{a} = 5604$, $\hat{b} = 0.0005877$

*Linear Regression*

Alternatively, one can plot $y_i$ against $-\ln[-\ln(i/(N+1))]$ on a linear scale. This will provide a straight line in case the $y_i$ are Gumbel distributed. Fitting a straight line with linear regression will yield the intercept $a$ at $-\ln[-\ln(i/(N+1))] = 0$ and the slope $1/b$.

*Methods of moments*

The easiest way to obtaining the Gumbel parameters is by the method of moments. From the probability density function it can be derived that the mean $\mu_Y$ and variance $\sigma_Y^2$ are given by (see also Table 3.3):

$$\mu_Y = a + \frac{0.5772}{b} \tag{4.11}$$

$$\sigma_Y^2 = \frac{\pi^2}{6b^2} \tag{4.12}$$

Using standard estimators for the mean ($m_Y = 1/n \sum Y_i$) and the variance ($s_Y^2 = 1/(n-1)\sum (Y_i - m_Y)^2$) the parameters can thus be estimated as:

$$\frac{1}{\hat{b}} = \sqrt{\frac{6s_Y^2}{\pi^2}} \tag{4.13}$$

$$\hat{a} = m_Y - \frac{0.5772}{\hat{b}} \tag{4.14}$$

Figure 4.5 shows a plot of the recurrence time versus maximum discharge with parameters fitted with the method of moments.

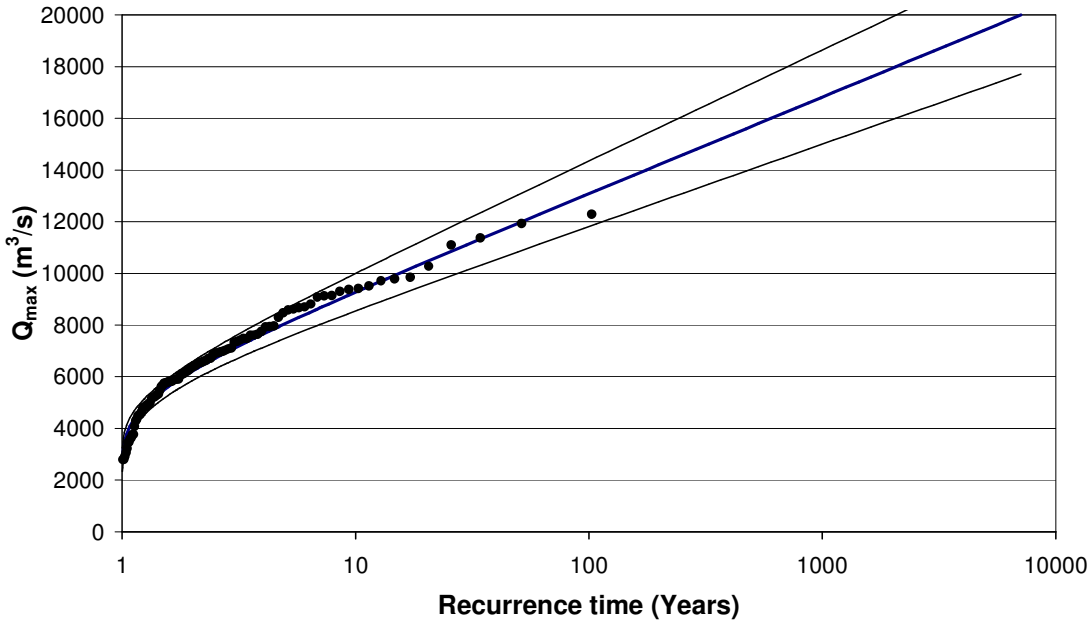



*Figure 4.5: Recurrence time versus yearly maximum of daily averaged discharge of the Rhine river at Lobith:* $\hat{a} = 5621$, $\hat{b} = 0.00061674$

The method of moments yields a biased estimate of the parameters. More sophisticated estimators are unbiased and more accurate are the probability weighted moments method (Landwehr at al, 1979) and debiased maximum likelihood methods (e.g. Hosking, 1985). These estimation methods are however much more complex, while the results are not always very different.

### 4.2.4 Estimation of the T-year event and its confidence limits

Once the parameters of the Gumbel-distribution have been estimated we can estimate magnitude the *T*-year event, which is the basis for our design. This event is estimated by (note this is a biased estimate):

$$\hat{y}_T = \hat{a} - \frac{1}{\hat{b}}\ln(-\ln(\frac{T-1}{T})) \tag{4.15}$$

Alternatively, one could of course read it directly from the Gumbel plot on probability paper. When applied to the Rhine dataset and the parameters obtained from the method of moments we obtain for instance for $T$=1250 (this is the current design norm) $y_{1250} = 5621 - 1621 \cdot \ln(-\ln(0.9992)) = 17182 \text{ m}^3/s$. With the parameters obtained from linear regression we obtain 17736 m$^3$/s.

When parameters are obtained through linear regression the 95%-confidence limits of $y_T$ can be obtained through:

$$\hat{y}_T - t_{95} s_{\hat{Y}}(T) \le y_T \le \hat{y}_T + t_{95} s_{\hat{Y}}(T) \tag{4.16}$$

with $t_{95}$ the 95-point of the student's t-distribution, and $s_{\hat{Y}}$ the standard error of the (regression) prediction error (taking account of parameter uncertainty) which is estimated as:

$$s^2_{\hat{Y}}(T) = \left(1 + \frac{1}{N} + \frac{(x(T) - \bar{x})^2}{\sum (x(T_i) - \bar{x})^2}\right) \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{4.17}$$

with $x(T_i) = -\ln[-\ln((T_i - 1)/T_i)]$ the transform of recurrence interval $T_i$ of event with rank $i$ from the $N$ pairs ($T_i$, $y_{Ti}$) in the Gumbel plot. In Figure 4.4 the 95% confidence limits are also given. The 95% confidence limits of the 1250 years event are {17183, 18289} m$^3$/s.

An approximate estimate of the estimation variance of $y_T$ in case the parameters are estimated with the method of moments is given by (Stedinger et al., 1993):

$$\hat{\text{Var}}(\hat{y}_T) \approx \frac{(1.11 + 0.52x + 0.61x^2)}{\hat{b}^2 N} \tag{4.18}$$

with $x = -\ln[-\ln(T-1)/T)]$. In figure 4.5 the 95%-confidence limits are given assuming the estimation error of $\hat{y}_T$ to be Gaussian distributed: $\hat{y}_T \pm 1.96\sqrt{\text{Var}(\hat{y}_T)}$. For $T$=1250 years these limits are: {15862,19611} m$^3$/s. In Van Montfort (1969) confidence limits are given for the case when the parameters are obtained with maximum likelihood estimation.

*The number of T-year events in a given period*

It should be stressed that the $T$-year flood event does not exactly arise every $T$ years! If we had a very long period of observations (say $\tau$ years, with $\tau$ in the order of 100000

years or so) then the $T$-year event would occur *on average* $N/T$ times. The probability that in $N$ years a $T$-year flood event occurs $n$ times follows a binomial distribution:

$$\Pr(n \text{ events } y_T \text{ in } N \text{ years}) = \binom{N}{n} p^n (1-p)^{N-n} \tag{4.19}$$

with $p = 1/T$. So the probability of a 100 year flood event occurring in the next year is 0.1. The probability of *exactly* one flood even occurring in the coming 10 years is

$$\Pr(1 \text{ events } y_{100} \text{ in } 10 \text{ years}) = \binom{10}{1} 0.01 \cdot (1-0.01)^9 = 0.0914 \tag{4.20}$$

The probability that *one or more* flood events occur in 10 years is 1-Pr(no events) = $1 - 0.99^{10} = 0.0956$. Going back to our Rhine example: the probability of at least one $T$=1250 years event occurring in the next 100 years is $1 - 0.9992^{100} = 0.0769$, which is still almost 8%!

*The time until the occurrence of a T-year event*

The number of years $m$ until the next $T$-year event is a random variable with a geometric distribution with $p = 1/T$:

$$\Pr(m \text{ years until event } y_T) = (1-p)^{m-1} p \tag{4.21}$$

The recurrence time $T$ is the expected value of this distribution: $E[m] = 1/p = T$. The probability of a $T$-year flood occurring in the next year is $p$ (as expected).

### 4.2.6 Pot data

An alternative way of obtaining flood statistics from time series of discharges is through *partial duration series* or *peak over threshold (pot) data*. The idea is to choose a threshold above which we imagine that discharge peaks would be called a flood event. We then consider only the discharges above this threshold. Figure 4.6 shows the pot-data for the Rhine catchment for a threshold of 6000 $m^3/s$. If there are $n$ exceedances and if the magnitude of the exceedances $Z$ are independent and identically distributed with distribution function $F_Z(z)$ than the probability of the maximum $Y$ of these exceedances is given by ( $y \geq 0$ and $y = 0$ if $n = 0$)

$$\begin{aligned} \Pr[Y \leq y] &= \Pr[Z_1 \leq y \cap Z_2 \leq y \cap \ldots \cap Z_n \leq y) = \\ &\Pr[Z_1 \leq y] \cdot \Pr[Z_2 \leq y] \cdots \Pr[Z_n \leq y] = \left(F_Z(y)\right)^n \end{aligned} \tag{4.22}$$

Because the *number* of exceedences $N$ is also a random variable (from Equation 3.22) we find:

$$\Pr[Y \le y] = \sum_{n=0}^{\infty} \left(F_Z(y)\right)^n \Pr[N = n] \tag{4.23}$$

It turns out that if the threshold is large enough that the number of exceedances has a Poisson distribution. Thus we have for (4.23):

$$\Pr[Y \le y] = \sum_{n=1}^{\infty} \left(F_Z(z)\right)^n \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda[1-F_Z(y)]} \tag{4.24}$$

If the exceedances obey an exponential distribution then it follows that the maximum of the Pot-data has a Gumbel distribution:

$$\begin{aligned} \Pr[Y \le y] &= e^{-\lambda[1-F_Z(y)]} = \exp[-\lambda(1-1+e^{-by})] \\ &= \exp[-e^{\ln\lambda} e^{-by}) = \exp[-e^{-b(y-\frac{\ln\lambda}{b})}] \end{aligned} \tag{4.25}$$

So in conclusion: *the maximum value of a Poisson number of independent and exponentially distributed exceedances follows a Gumbel distribution.*

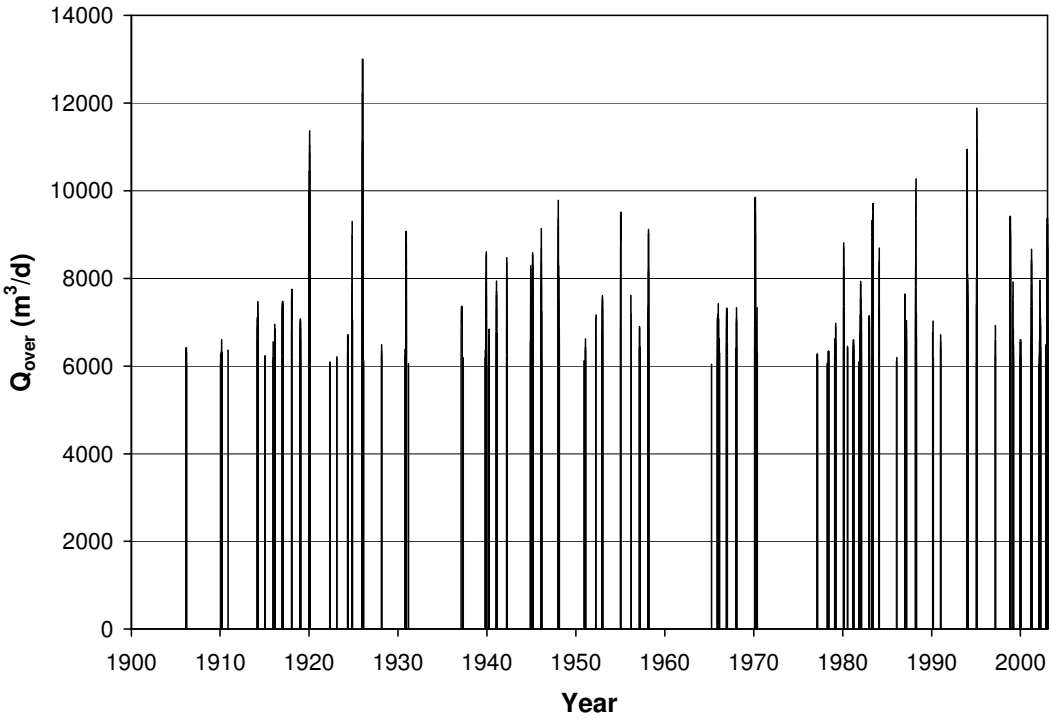



*Figure 4.6 Peak over threshold data of daily average Rhine discharge at Lobith*

The analysis of the pot data is rather straightforward. The *T*-year exceedance is given by:

$$\hat{y}_T = \frac{\ln \lambda}{b} - \frac{1}{b}\ln(-\ln(\frac{T-1}{T})) \tag{4.26}$$

with $\lambda$ the mean number of exceedances per time unit $\bar{n}$ and $1/b$ the average magnitude of the exceedances $\bar{z}$ which can both be estimated from the pot data. The pot data of the Rhine data at Lobith yield: $\bar{n} = 5.786$ and $\bar{z} = 1228\,\mathrm{m^3/s}$. The 1250 year exceedance is thus given by:

$$\hat{y}_T = 1228\left\{\ln 5.786 - \ln(-\ln(\frac{1249}{1250}))\right\} = 10910\,\mathrm{m^3/s} \tag{4.27}$$

The 1:1250 year event then is obtained by adding the threshold: 16910 $\mathrm{m^3/s}$.

### 4.2.7 Other distributions for maximum values

The Gumbel distribution is not the only probability distribution that is used to model maximum values. In fact, the Gumbel distribution is a specific case of the so called General Extreme Value distribution (GEV):

$$F(y) = \begin{cases} \exp\left\{-\left[1 - b\theta(y-a)\right]^{1/\theta}\right\} & \theta \neq 0 \\ \exp\{-\exp(-b(y-a)\} & \theta = 0 \end{cases} \tag{4.28}$$

Where $\theta$ is a form parameter. As can be seen the GEV reverts to the Gumbel distribution for $\theta$=0. The Gumbel distribution therefore is often called the Extreme value distribution of type I (EV type I). If $\theta$ >0 we obtain a EV type III distribution (or Weibull distribution) which has a finite upper bound. If $\theta$ < 0 it is called the EV type II distribution whose right hand tail is thicker than the Gumbel distribution, such that we have a larger probability of larger floods. Figure 4.7 shows the three types of extreme value distributions on Gumbel probability paper. As can be seen, the three distributions coincide for T = 1/(1-*e*) ≈1.582 years. Apart from the GEV distribution other distributions used to model maximum values are the lognormal distribution (see 3.2) and the log-Pearson type 3 distribution. The parameters of the lognormal distribution can be obtained through for instance the method of moments by estimating the mean and the variance of the log-maximum values $\mu_{\ln Y}$ and $\sigma^2_{\ln Y}$. The Pearson type 3 distribution does not have a closed form solution but is tabulated. The *p-quantile* $y_p$ of the probability distribution (value of $y$ for which $\Pr(Y \leq y) = p$) is also tabulated using so called frequency factors $K_p(CS_Y)$:

$$y_p = \mu_Y + \sigma_Y K_p(CS_Y) \tag{4.29}$$

with $\mu_Y, \sigma_Y, CS_Y$ the mean, standard deviation and coefficient of skewness respectively. The Log-Pearson type 3 distribution is obtained by using the mean, variance and skewness of the logarithms: $\mu_{\ln Y}, \sigma_{\ln Y}, CS_{\ln Y}$. Differences between extrapolations with these distributions can be substantial if the time series is length is limited, in which case the choice of the distribution may be very critical. To be safe, sometimes several distributions are fitted and the most critical taken as design norm. Figure 4.8 shows two different distributions fitted to the Rhine data.

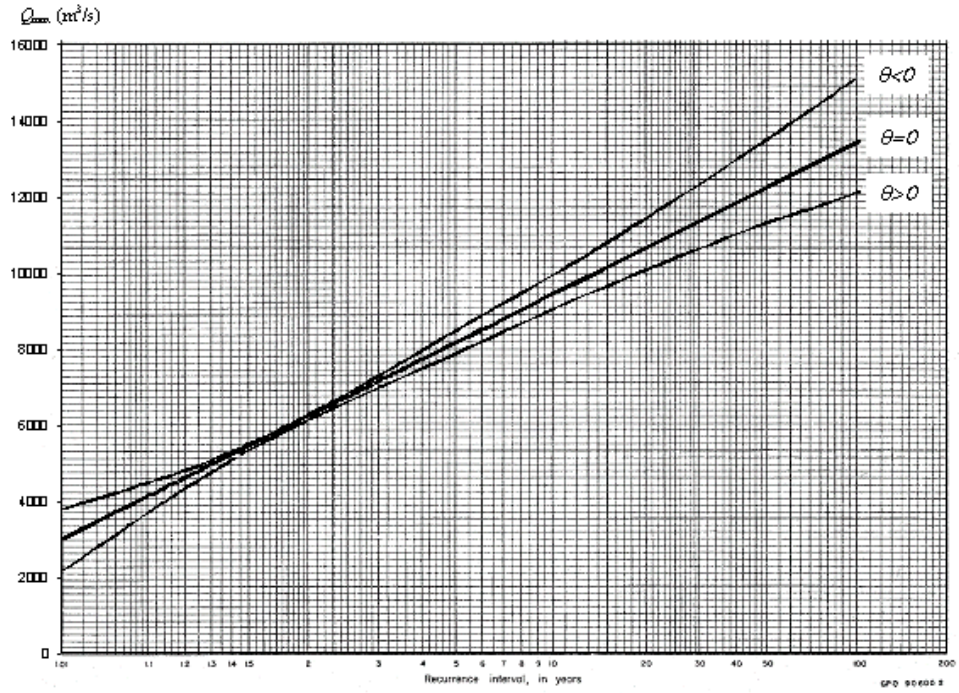



*Figure 4.7 GEV-distributions on Gumbel probability paper.*

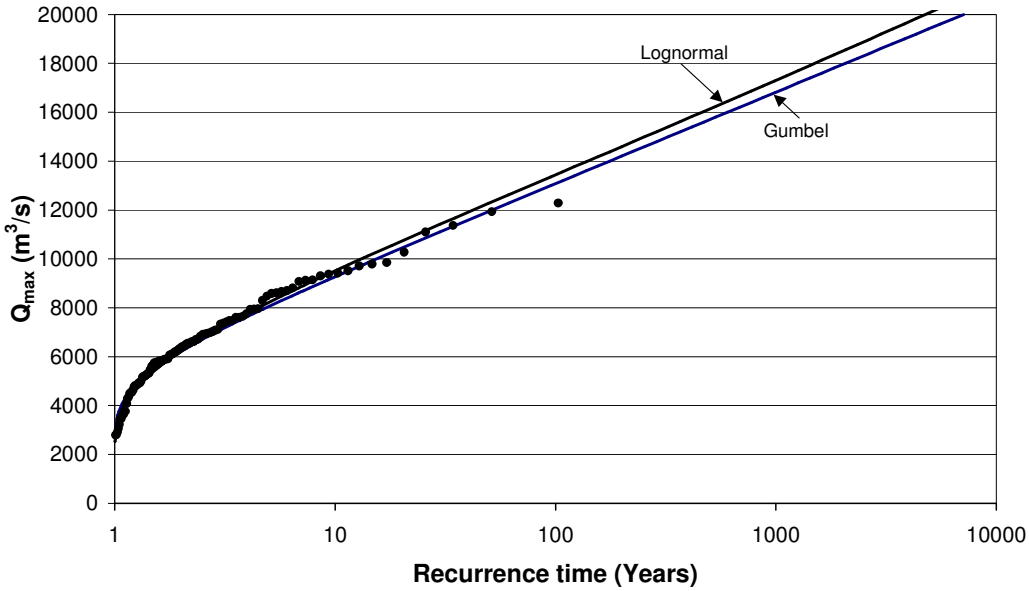



*Figure 4.8 Gumbel and lognormal distributions fitted to the same maximum values.*

### 4.2.8 Minimum values

Up to now we have only been concerned with maximum values. However in the analysis of, for instance, low flows we may be concerned with modelling the probability distribution of yearly minimum discharge. One simple way of dealing with minimum values is to transform them into maximum values, e.g. by considering $-y$ or $1/y$. Also, the GEV type III or Weibull distribution is suitable for modelling minimum values that are bounded below by zero.

## 4.3 Some useful statistical tests

Estimation of $T$-year events from records of maximum values is based on the assumption that the maximum values are independent, are homogenous (there is no change in mean and variance over time) and have a certain probability distribution (e.g. Gumbel). Here we will provide some useful tests for checking maximum value data on independence, trends and assumed distribution. For each of the assumptions there are numerous tests available. We do not presume to be complete here, but only present one test per assumption.

### 4.3.1 Testing for independence

Von Neuman's $Q$-statistic can be used to test whether a series of data can be considered as realisations of independent random variables:

$$Q = \frac{\sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2} \tag{4.30}$$

where $Y_i$ is the maximum value of year $i$ and $Y_{i+1}$ is the maximum value of year $i$+1 etc., and $n$ the length of the series. It can be shown that if the correlation coefficient between maximum values of two subsequent years is positive ($\rho > 0$, which is usually the case for natural processes), that $Q$ will be smaller if $\rho$ is larger. If we accept that if the maximum values are dependent that their correlation will be positive, we can use $Q$ as a test statistic for the following two hypothesis: $H_0$: the maximum values are independent; $H_1$: they are dependent. In that case we have a lower critical area. Table 4.2 shows lower critical values for $Q$. Under the assumption that $y_i$ are independent and from the same distribution we have that $E[Q]$=2. If $Q$ is smaller than a critical value for a certain confidence level $\alpha$ (here either 0.001, 0.01 or 0.05) we can say with an accuracy of $\alpha$ that the data are dependent. For instance, taking the maximum values of the last 20 years (1984-2003) at Lobith we obtain for $n$=20 that $Q$ = 2.16. From Table 4.2 we see that Q is not in the critical area, so there is no reason to reject the hypothesis that the data are independent. Note that for large $n$ we have that

$$Q' = \frac{Q - 2}{2\sqrt{(n-2)/(n^2-1)}} \tag{4.31}$$

is approximately standard Gaussian distributed. In case of the Rhine data set we have for $n$=103 that $Q$=2.071 and therefore $Q'$ = 0.366 which is far from the critical 0.05 value of –1.96: the zero-hypothesis of independent maximum values cannot be rejected.

*Table 4.2 Lower critical values for von Neuman's test of independence.*

| | $\alpha$ | | | | $\alpha$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | 0.001 | 0.01 | 0.05 | $n$ | 0.001 | 0.01 | 0.05 |
| 4 | 0.5898 | 0.6256 | 0.7805 | 36 | 1.0416 | 1.2581 | 1.4656 |
| 5 | 0.4161 | 0.5379 | 0.8204 | 37 | 1.0529 | 1.2673 | 1.4726 |
| | | | | 38 | 1.0639 | 1.2763 | 1.4793 |
| 6 | 0.3634 | 0.5615 | 0.8902 | 39 | 1.0746 | 1.2850 | 1.4858 |
| 7 | 0.3695 | 0.6140 | 0.9359 | 40 | 1.0850 | 1.2934 | 1.4921 |
| 8 | 0.4036 | 0.6628 | 0.9825 | | | | |
| 9 | 0.4420 | 0.7088 | 1.0244 | 41 | 1.0950 | 1.3017 | 1.4982 |
| 10 | 0.4816 | 0.7518 | 1.0623 | 42 | 1.1048 | 1.3096 | 1.5041 |
| | | | | 43 | 1.1142 | 1.3172 | 1.5098 |
| 11 | 0.5197 | 0.7915 | 1.0965 | 44 | 1.1233 | 1.3246 | 1.5154 |
| 12 | 0.5557 | 0.8280 | 1.1276 | 45 | 1.1320 | 1.3317 | 1.5206 |
| 13 | 0.5898 | 0.8618 | 1.1558 | | | | |
| 14 | 0.6223 | 0.8931 | 1.1816 | 46 | 1.1404 | 1.3387 | 1.5257 |
| 15 | 0.6532 | 0.9221 | 1.2053 | 47 | 1.1484 | 1.3453 | 1.5305 |
| | | | | 48 | 1.1561 | 1.3515 | 1.5351 |
| 16 | 0.6826 | 0.9491 | 1.2272 | 49 | 1.1635 | 1.3573 | 1.5395 |
| 17 | 0.7104 | 0.9743 | 1.2473 | 50 | 1.1705 | 1.3629 | 1.5437 |
| 18 | 0.7368 | 0.9979 | 1.2660 | | | | |
| 19 | 0.7617 | 1.0199 | 1.2834 | 51 | 1.1774 | 1.3683 | 1.5477 |
| 20 | 0.7852 | 1.0406 | 1.2996 | 52 | 1.1843 | 1.3738 | 1.5518 |
| | | | | 53 | 1.1910 | 1.3792 | 1.5557 |
| 21 | 0.8073 | 1.0601 | 1.3148 | 54 | 1.1976 | 1.3846 | 1.5596 |
| 22 | 0.8283 | 1.0785 | 1.3290 | 55 | 1.2041 | 1.3899 | 1.5634 |
| 23 | 0.8481 | 1.0958 | 1.3425 | | | | |
| 24 | 0.8668 | 1.1122 | 1.3552 | 56 | 1.2104 | 1.3949 | 1.5670 |
| 25 | 0.8846 | 1.1278 | 1.3671 | 57 | 1.2166 | 1.3999 | 1.5707 |
| | | | | 58 | 1.2227 | 1.4048 | 1.5743 |
| 26 | 0.9017 | 1.1426 | 1.3785 | 59 | 1.2288 | 1.4096 | 1.5779 |
| 27 | 0.9182 | 1.1567 | 1.3892 | 60 | 1.2349 | 1.4144 | 1.5814 |
| 28 | 0.9341 | 1.1702 | 1.3994 | | | | |
| 29 | 0.9496 | 1.1830 | 1.4091 | | | | |
| 30 | 0.9645 | 1.1951 | 1.4183 | | | | |
| 31 | 0.9789 | 1.2067 | 1.4270 | | | | |
| 32 | 0.9925 | 1.2177 | 1.4354 | | | | |
| 33 | 1.0055 | 1.2283 | 1.4434 | | | | |
| 34 | 1.0180 | 1.2386 | 1.4511 | | | | |
| 35 | 1.0300 | 1.2485 | 1.4585 | | | | |

### 4.3.2 Testing for trends

There are many tests on trends, some of them specifically focussed on linear trends or step trends or based on special assumptions about the distribution of the data sequence. Here a nonparametric test statistic is presented (the Mann-Kendall test for trends) that makes no assumptions about the type of trend (linear or non-linear) or the distribution of the data. The trend must however be monotoneous and not periodic. Consider the series of annual maximum values $Y_i$ $i$=1,..,$n$. Each value $Y_i$ $i$=2,..,$n$ is compared with all the subsequent values $Y_j$ $j$=1,..,$i$-1 to compute the Mann-Kendall test statistic:

$$T = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathrm{sgn}(Y_i - Y_j) \tag{4.32}$$

with

$$\mathrm{sgn}(Y_i - Y_j) = \begin{cases} 1 \text{ if } Y_i > Y_j \\ 0 \text{ if } Y_i = Y_j \\ -1 \text{ if } Y_i < Y_j \end{cases} \tag{4.33}$$

For large $n$ (n>40) and no serial dependence of the data the statistic $T$ is asymptotically Gaussian distributed with mean $E[T] = 0$ and variance $Var[T] = [n(n-1)(2n+5)]/18$. This means that the test $H_0$: the maximum values have no trend; $H_1$: they have a trend, using the test statistic $T' = 18T/[n(n-1)(2n+5)]$ has a two-sided critical area with critical levels given by the quantiles of the standard normal distribution: $\chi_{\alpha/2}$ and $\chi_{1-\alpha/2}$ (with $\alpha$ the significance level. Performing this test for the Rhine data set of maximum values yields: $T$=689 and $T'$=1.992. This means that the Hypothesis of no trend is rejected. The value of $T'$ suggests a positive trend at 5% accuracy.

### 4.3.3 Testing a presumed probability distribution

A very general test on a distribution is Pearson's chi-squared ($\chi^2$) test. Consider a data set with $n$ maximum values. The values are classified into $m$ non-overlapping classes to obtain a histogram. For each class $i$ we calculate the number $n_i$ of data falling in this class (note that $\sum_{i=1}^{m} n_i = n$). Using the presumed distribution function[2] $F_Y(y)$ the expected number of data falling into class $i$ can be calculated as:

$$n_i^e = n\left(F_Y(y_{\mathrm{up}}) - F_Y(y_{\mathrm{low}})\right) \tag{4.34}$$

with $y_{\mathrm{up}}$ and $y_{\mathrm{low}}$ the upper and lower boundaries of class $i$ respectively. The test statistic is given as:

$$\mathrm{X}^2 = \sum_{i=1}^{m} \frac{(n_i - n_i^e)^2}{n_i^e} \tag{4.35}$$

The statistic $\mathrm{X}^2$ has a $\chi^2_{m-1}$ distribution, i.e. a chi-squared distribution with $m$-1 degrees of freedom. Table 4.3 provides upper critical values for the $\chi^2$ distribution for various degrees of freedom and significance levels. Application to the Rhine data set gives for 20 classes of width 500 from 2500 to 13000 $m^3/d$. and the assumed Gumbel distribution with $a$=5621 and $b$= 0.000616735 leads to $\mathrm{X}^2 = 23.70$. This is outside the critical area for 19

[2] Note that if the parameters of the presumed distribution are estimated with the method of moments some bias is introduced.

degrees of freedom and an accuracy of 5%. We therefore conclude that the assumption of the maxima being Gumbel distributed cannot be rejected. Performing the same test with the lognormal distribution yields $X^2 = 14.36$ which fits even better.

*Table 4.3 Upper critical values for the $\chi^2$ distribution*

| | Probability of exceeding the critical value | | | | |
|---|---|---|---|---|---|
| $\nu$ | 0.10 | 0.05 | 0.025 | 0.01 | 0.001 |
| 1 | 2.706 | 3.841 | 5.024 | 6.635 | 10.828 |
| 2 | 4.605 | 5.991 | 7.378 | 9.210 | 13.816 |
| 3 | 6.251 | 7.815 | 9.348 | 11.345 | 16.266 |
| 4 | 7.779 | 9.488 | 11.143 | 13.277 | 18.467 |
| 5 | 9.236 | 11.070 | 12.833 | 15.086 | 20.515 |
| 6 | 10.645 | 12.592 | 14.449 | 16.812 | 22.458 |
| 7 | 12.017 | 14.067 | 16.013 | 18.475 | 24.322 |
| 8 | 13.362 | 15.507 | 17.535 | 20.090 | 26.125 |
| 9 | 14.684 | 16.919 | 19.023 | 21.666 | 27.877 |
| 10 | 15.987 | 18.307 | 20.483 | 23.209 | 29.588 |
| 11 | 17.275 | 19.675 | 21.920 | 24.725 | 31.264 |
| 12 | 18.549 | 21.026 | 23.337 | 26.217 | 32.910 |
| 13 | 19.812 | 22.362 | 24.736 | 27.688 | 34.528 |
| 14 | 21.064 | 23.685 | 26.119 | 29.141 | 36.123 |
| 15 | 22.307 | 24.996 | 27.488 | 30.578 | 37.697 |
| 16 | 23.542 | 26.296 | 28.845 | 32.000 | 39.252 |
| 17 | 24.769 | 27.587 | 30.191 | 33.409 | 40.790 |
| 18 | 25.989 | 28.869 | 31.526 | 34.805 | 42.312 |
| 19 | 27.204 | 30.144 | 32.852 | 36.191 | 43.820 |
| 20 | 28.412 | 31.410 | 34.170 | 37.566 | 45.315 |
| 21 | 29.615 | 32.671 | 35.479 | 38.932 | 46.797 |
| 22 | 30.813 | 33.924 | 36.781 | 40.289 | 48.268 |
| 23 | 32.007 | 35.172 | 38.076 | 41.638 | 49.728 |
| 24 | 33.196 | 36.415 | 39.364 | 42.980 | 51.179 |
| 25 | 34.382 | 37.652 | 40.646 | 44.314 | 52.620 |
| 26 | 35.563 | 38.885 | 41.923 | 45.642 | 54.052 |
| 27 | 36.741 | 40.113 | 43.195 | 46.963 | 55.476 |
| 28 | 37.916 | 41.337 | 44.461 | 48.278 | 56.892 |
| 29 | 39.087 | 42.557 | 45.722 | 49.588 | 58.301 |
| 30 | 40.256 | 43.773 | 46.979 | 50.892 | 59.703 |
| 31 | 41.422 | 44.985 | 48.232 | 52.191 | 61.098 |
| 32 | 42.585 | 46.194 | 49.480 | 53.486 | 62.487 |
| 33 | 43.745 | 47.400 | 50.725 | 54.776 | 63.870 |
| 34 | 44.903 | 48.602 | 51.966 | 56.061 | 65.247 |
| 35 | 46.059 | 49.802 | 53.203 | 57.342 | 66.619 |
| 36 | 47.212 | 50.998 | 54.437 | 58.619 | 67.985 |
| 37 | 48.363 | 52.192 | 55.668 | 59.893 | 69.347 |
| 38 | 49.513 | 53.384 | 56.896 | 61.162 | 70.703 |
| 39 | 50.660 | 54.572 | 58.120 | 62.428 | 72.055 |
| 40 | 51.805 | 55.758 | 59.342 | 63.691 | 73.402 |
| 41 | 52.949 | 56.942 | 60.561 | 64.950 | 74.745 |
| 42 | 54.090 | 58.124 | 61.777 | 66.206 | 76.084 |
| 43 | 55.230 | 59.304 | 62.990 | 67.459 | 77.419 |
| 44 | 56.369 | 60.481 | 64.201 | 68.710 | 78.750 |
| 45 | 57.505 | 61.656 | 65.410 | 69.957 | 80.077 |
| 46 | 58.641 | 62.830 | 66.617 | 71.201 | 81.400 |
| 47 | 59.774 | 64.001 | 67.821 | 72.443 | 82.720 |
| 48 | 60.907 | 65.171 | 69.023 | 73.683 | 84.037 |
| 49 | 62.038 | 66.339 | 70.222 | 74.919 | 85.351 |
| 50 | 63.167 | 67.505 | 71.420 | 76.154 | 86.661 |

| | Probability of exceeding the critical value | | | | |
|---|---|---|---|---|---|
| $\nu$ | 0.10 | 0.05 | 0.025 | 0.01 | 0.001 |
| 51 | 64.295 | 68.669 | 72.616 | 77.386 | 87.968 |
| 52 | 65.422 | 69.832 | 73.810 | 78.616 | 89.272 |
| 53 | 66.548 | 70.993 | 75.002 | 79.843 | 90.573 |
| 54 | 67.673 | 72.153 | 76.192 | 81.069 | 91.872 |
| 55 | 68.796 | 73.311 | 77.380 | 82.292 | 93.168 |
| 56 | 69.919 | 74.468 | 78.567 | 83.513 | 94.461 |
| 57 | 71.040 | 75.624 | 79.752 | 84.733 | 95.751 |
| 58 | 72.160 | 76.778 | 80.936 | 85.950 | 97.039 |
| 59 | 73.279 | 77.931 | 82.117 | 87.166 | 98.324 |
| 60 | 74.397 | 79.082 | 83.298 | 88.379 | 99.607 |
| 61 | 75.514 | 80.232 | 84.476 | 89.591 | 100.888 |
| 62 | 76.630 | 81.381 | 85.654 | 90.802 | 102.166 |
| 63 | 77.745 | 82.529 | 86.830 | 92.010 | 103.442 |
| 64 | 78.860 | 83.675 | 88.004 | 93.217 | 104.716 |
| 65 | 79.973 | 84.821 | 89.177 | 94.422 | 105.988 |
| 66 | 81.085 | 85.965 | 90.349 | 95.626 | 107.258 |
| 67 | 82.197 | 87.108 | 91.519 | 96.828 | 108.526 |
| 68 | 83.308 | 88.250 | 92.689 | 98.028 | 109.791 |
| 69 | 84.418 | 89.391 | 93.856 | 99.228 | 111.055 |
| 70 | 85.527 | 90.531 | 95.023 | 100.425 | 112.317 |
| 71 | 86.635 | 91.670 | 96.189 | 101.621 | 113.577 |
| 72 | 87.743 | 92.808 | 97.353 | 102.816 | 114.835 |
| 73 | 88.850 | 93.945 | 98.516 | 104.010 | 116.092 |
| 74 | 89.956 | 95.081 | 99.678 | 105.202 | 117.346 |
| 75 | 91.061 | 96.217 | 100.839 | 106.393 | 118.599 |
| 76 | 92.166 | 97.351 | 101.999 | 107.583 | 119.850 |
| 77 | 93.270 | 98.484 | 103.158 | 108.771 | 121.100 |
| 78 | 94.374 | 99.617 | 104.316 | 109.958 | 122.348 |
| 79 | 95.476 | 100.749 | 105.473 | 111.144 | 123.594 |
| 80 | 96.578 | 101.879 | 106.629 | 112.329 | 124.839 |
| 81 | 97.680 | 103.010 | 107.783 | 113.512 | 126.083 |
| 82 | 98.780 | 104.139 | 108.937 | 114.695 | 127.324 |
| 83 | 99.880 | 105.267 | 110.090 | 115.876 | 128.565 |
| 84 | 100.980 | 106.395 | 111.242 | 117.057 | 129.804 |
| 85 | 102.079 | 107.522 | 112.393 | 118.236 | 131.041 |
| 86 | 103.177 | 108.648 | 113.544 | 119.414 | 132.277 |
| 87 | 104.275 | 109.773 | 114.693 | 120.591 | 133.512 |
| 88 | 105.372 | 110.898 | 115.841 | 121.767 | 134.746 |
| 89 | 106.469 | 112.022 | 116.989 | 122.942 | 135.978 |
| 90 | 107.565 | 113.145 | 118.136 | 124.116 | 137.208 |
| 91 | 108.661 | 114.268 | 119.282 | 125.289 | 138.438 |
| 92 | 109.756 | 115.390 | 120.427 | 126.462 | 139.666 |
| 93 | 110.850 | 116.511 | 121.571 | 127.633 | 140.893 |
| 94 | 111.944 | 117.632 | 122.715 | 128.803 | 142.119 |
| 95 | 113.038 | 118.752 | 123.858 | 129.973 | 143.344 |
| 96 | 114.131 | 119.871 | 125.000 | 131.141 | 144.567 |
| 97 | 115.223 | 120.990 | 126.141 | 132.309 | 145.789 |
| 98 | 116.315 | 122.108 | 127.282 | 133.476 | 147.010 |
| 99 | 117.407 | 123.225 | 128.422 | 134.642 | 148.230 |
| 100 | 118.498 | 124.342 | 129.561 | 135.807 | 149.449 |

## 4.4 Exercises

4.1 Consider the following yearly maximum daily streamflow data of the Meuse river at the Eysden station at the Dutch-Belgian border (daily average discharge in $m^3/s$)

| 1950 | 1951 | 1952 | 1953 | 1954 | 1955 | 1956 | 1957 | 1958 | 1959 | 1960 | 1961 | 1962 | 1963 | 1964 | 1965 | 1966 | 1967 | 1968 | 1969 | 1970 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1266 | 1492 | 1862 | 861 | 715 | 1367 | 1837 | 1429 | 1429 | 1261 | 1607 | 2132 | 1652 | 1537 | 1155 | 1899 | 1956 | 1596 | 1380 | 745 | 2181 |

| 1971 | 1972 | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 | 1979 | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 955 | 1007 | 824 | 1271 | 1044 | 597 | 1216 | 1061 | 1450 | 2016 | 1270 | 1341 | 1075 | 2482 | 874 | 1689 | 1554 | 1831 | 1149 | 1444 | 1791 |

| 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|---|---|---|
| 1207 | 3050 | 1578 | 2817 | 792 | 1143 | 1698 | 2076 | 1204 | 1835 |

a. Plot the data on Gumbel-probability paper and determine the parameters of the Gumbel distribution. Estimate the 1000-year flood?
b. Determine the parameters of the Gumbel distribution by linear regression of $Q_{max}$ against $-\ln[-\ln\{i/(n+1)\}]$ with $i$ the rank of the $i$th smallest maximum. Estimate the 1000-year flood and its 95% confidence limits.
c. Determine the parameters of the Gumbel distribution with the method of moments. Estimate the 1000-year flood and its 95% confidence limits.
d. Estimate the 1000-year flood assuming a lognormal distribution of the maximum values.
e. Test whether these maximum values can be modelled as independent stochastic variables.
f. Test whether the data can be considered as outcomes of a Gumbel distribution.
g. Test whether the data can be considered as outcomes of a lognormal distribution.
h. What is the probability that the 1000-year flood occurs at least once within the next 40 years?
i. What is the probability that the 1000-year flood occurs twice in the next 100 years?

# 5. Random functions

In chapter 3 random variables were treated. In this chapter these concepts are extended to random functions. Only the basic properties of random functions are treated. Elaborate treatment of random functions can be found in standard textbooks on the subject such as Papoulis (1991) and Vanmarcke (1983). A very basic and excellent introduction to random functions can be found in Isaaks and Srivastava (1989).

## 5.1 Definitions

Figure 5.1 shows the schematically the concept of a random function (RF). Consider some property z that varies in space (e.g. hydraulic conductivity), time (e.g. surface water levels) or space and time (e.g. groundwater depth). Except at locations or times where the value of property z is observed, we do not know the exact values of $z$. In order to express our uncertainty about $z$ we adopt the following concept. Instead of using a single function to describe the variation of $z$ in space and/or time, a family of functions is chosen. Each of these functions is assumed to have an equal probability of representing the true but unknown variation in space and/or time. The family of equally probably functions is called the *ensemble*, or alternatively a *random function*. As with a random variable, a random function is usually denoted with a capital.

If $Z$ is a function of space it is also referred to as a *random space function* (RSF) or *random field*:

$$Z(\mathbf{x}), \mathbf{x} = (x, y) \in \Re^2 \text{ or } \mathbf{x} = (x, y, z) \in \Re^3$$

If $Z$ is a function of time it is referred to as a *random time function* (RTF), *random process* or *stochastic process*:

$$Z(t), t \in \Re$$

If $Z$ is a function of space and time it is referred to as a *random space-time function* (RSTF) or *space-time random field*:

$$Z(\mathbf{x}, t), \mathbf{x} \in \Re^2 / \Re^3, t \in \Re$$

In this chapter we will treat most of the theory of random functions using the temporal framework $Z(t)$**.** The spatial framework $Z(\mathbf{x})$ is used for concepts that are a) only defined in space; b) can be better explained in space c) in case certain definitions in the spatial framework are different from the temporal framework.

In Figure 5.1 four functions (ensemble members) are shown. One particular function or ensemble member out of the many is called a *realisation* and is denoted with the lower

case $z(t)$. Depending on the type of random function (see hereafter) the number of possible realisations making up a random function can be either finite or infinite.

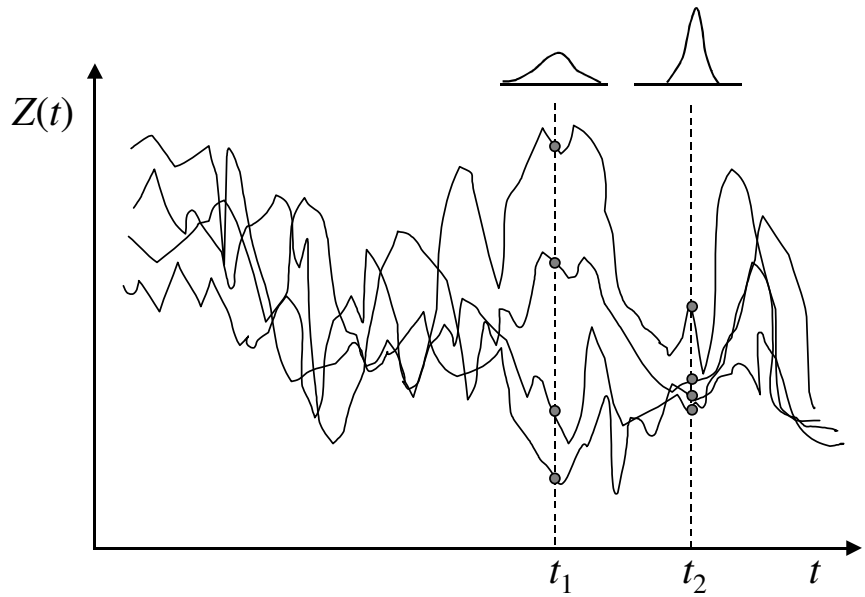



*Figure 5.1 Schematic representation of a random function*

Another way of looking at a random function is as a collection of random variables (one at every location in space or point in time) that are all mutually statistically dependent. If we return to the example of Figure 5.1: at every point in time $t$ a random variable $Z(t)$ is defined. At each $t$ the random variable $Z(t)$ is described with a probability density function (pdf) $f_Z(z;t)$ which not only depends on the value of $z$, but also on the value of $t$. This pdf could be estimated by sampling all realisations at a certain point (say $t_1$) in time or location and calculating the histogram of the samples. Figure 5.1 shows that the variation among the realisations is larger at point $t_1$ than at point $t_2$, leading to a probability density function at $t_1$ with a larger variance than at $t_2$. So, we are less certain about the unknown value at $t_1$ than we are at $t_2$.

At each point in time (or location) we can calculate the time- (or location-) dependent mean and variance:

$$\mu_Z(t) = \mathrm{E}[Z(t)] = \int_{-\infty}^{\infty} z\, f_Z(z;t)dz \tag{5.1}$$

and the variance is defined as

$$\sigma_Z^2(t) = \mathrm{E}[\{Z(t) - \mu_Z(t)\}^2] = \int_{-\infty}^{\infty} \{z - \mu_Z(t)\}^2 f_Z(z;t)dz. \tag{5.2}$$

Also, the random variables are usually correlated in time (or in space for spatial random functions): $\mathrm{COV}[Z(t_1), Z(t_2)] \neq 0$. The covariance is generally smaller when random variables are considered at locations further apart. The covariance is defined as:

$$\mathrm{COV}[Z(t_1), Z(t_2)] = \mathrm{E}[\{Z(t_1) - \mu_Z(t_1)\}\{Z(t_2) - \mu_Z(t_2)\}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{z_1 - \mu_Z(t_1)\}\{z_2 - \mu_Z(t_2)\} f(z_1, z_2; t_1, t_2) dz_1 dz_2 \tag{5.3}$$

For $t_1 = t_2$ the covariance equals the variance (5.2).

For a set of $N$ discrete random variables the joint or multivariate probability distribution $\Pr[d_1,d_2,\ldots,d_N]$ describes the probability that $N$ random variables have a certain value, i.e. that $D_1 = d_1$ and $D_2 = d_2$ and ….and $D_N = d_N$. Similarly, for $N$ continuous random variables the multivariate probability density $f(z_1,z_2,\ldots. z_N)$ is a measure of $N$ random variables $Z_i$, $i$=1,..,$N$ variables having a certain value. The analogue for a random function is the probability measure that a realisation of $Z(t)$ at a set of $N$ locations $t_i$, $i$ = 1,..,$N$ has the values between $z_1 < z(t_1) \leq z_1 + dz_1$, $z_2 < z(t_2) \leq z_2 + dz_2$,.., $z_N < z(t_N) \leq z_N + dz_N$ respectively. The associated multivariate probability density function (pdf) is denoted as $f(z_1,z_2,\ldots. z_N ; t_1,t_2,\ldots. t_N)$ and defined as:

$$f(z_1,z_2,\ldots. z_N ; t_1,t_2,\ldots. t_N) = \lim_{\partial z_1,\ldots,\partial z_N \to 0} \frac{\Pr(z_1 < z(t_1) \leq z_1 + \partial z_1, z_2 < z(t_2) \leq z_2 + \partial z_2, \ldots, z_N < z(t_N) \leq z_N + \partial z_N}{\partial z_1 \partial z_2 \cdots \partial z_N} \tag{5.4}$$

Because it relates to random variables at different points or locations, the multivariate pdf of a random function is sometime referred to as *multipoint* pdf. Theoretically, a random function is fully characterised (we know all there is to know about it) if the multivariate probability distribution for any set of points is known.

## 5.2 Types of random functions

Random functions can be divided into types based on whether their functional values are continuous (e.g. hydraulic conductivity in space or discharge in time) or discrete (e.g. the number of floods in a given period). Another distinction is based on the way the domain of the random function is defined (see Figure 5.2). For instance, a continuous valued random function $Z$ can be:

a) defined at all locations in time, space or space time: $Z(t)$, $Z(\mathbf{x})$ or $Z(\mathbf{x},t)$;
b) defined at discrete points in time, space or space time, where $Z(k\Delta t)$, $k$=1,..,$K$ is referred to as a *random time series* and $Z(i\Delta x, j\Delta y)$, $i$=1,..,$I$; $j$=1,..,$J$ as a *lattice process*;
c) defined at random times or random locations in space and time: $Z(T)$, $Z(\mathbf{X})$ or $Z(\mathbf{X},T)$. Such a process is called a *compound point process*. The occurrence of the points at

random co-ordinates in space and time is called a *point process*. If the occurrence of such a point is associated with the occurrence of a random variable (e.g. the occurrence of a thunder storm cell at a certain location in space with a random intensity of rainfall) it is called a compound point process.

Naturally, one can make the same distinction for discrete-valued random functions, e.g. $D(t)$, $D(k\Delta t)$ or $D(T)$; $D(\mathbf{x})$, $D(i\Delta x, j\Delta y)$ or $D(\mathbf{X})$ etc.

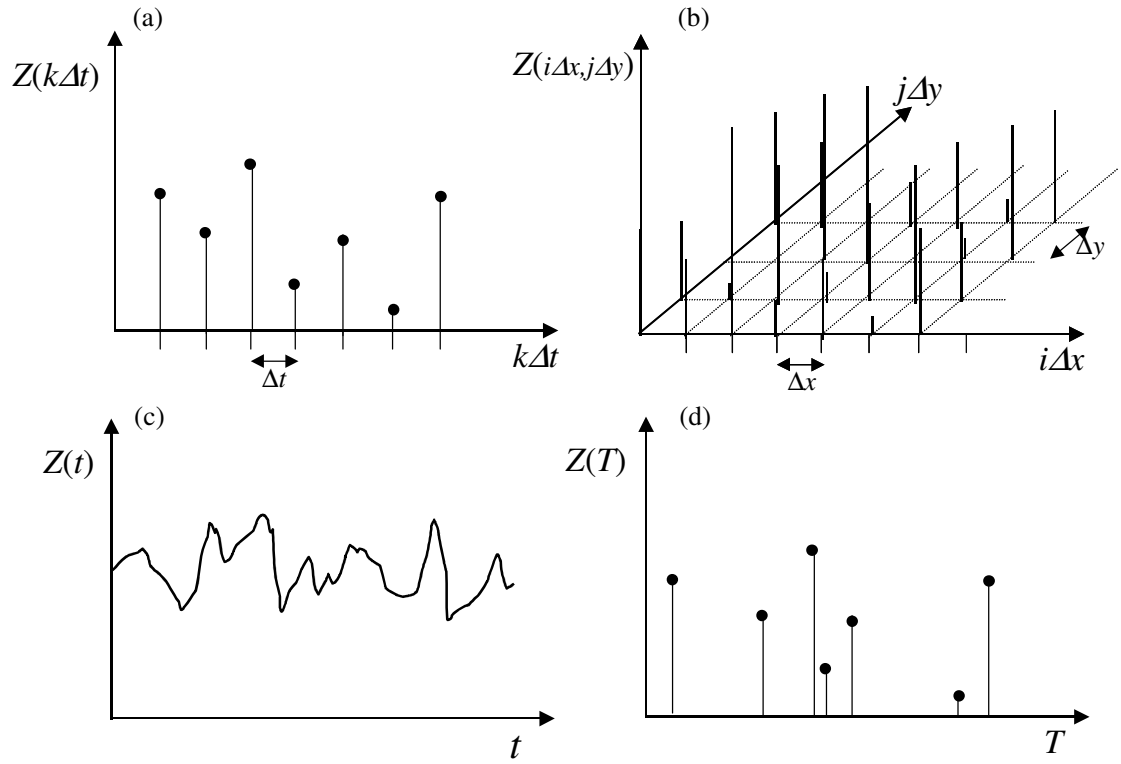



*Figure 5.2 Examples of realisations of different types of random functions based on the way they are defined on the functional domain; a) random time series; b) lattice process; c) continuous-time random function; d) compound point process.*

## 5.3 Stationary random functions

### 5.3.1 Strict stationary random functions

A special kind of random function is a called a strict *stationary* random function. A random function is called *strict stationary* if its multivariate pdf is invariant under translation. So for any set of $N$ locations and for any translation $t$' we have that

$$f(z_1, z_2, \ldots, z_N; t_1, t_2, \ldots. t_N) = f(z_1, z_2, \ldots., z_N ; t_1 + t', t_2 + t', \ldots, t_N + t') \quad \forall t_i, t' \tag{5.5}$$

Thus, we can have any configuration of points on the time axis and move this configuration (the whole configuration, not one point at the time) of points forward and backwards in time and have the same multivariate pdf. For the spatial domain we have to stress that strict stationarity means an invariant pdf under translation only, not rotation. So for a strict stationary random function in two dimensions, the two sets of locations in the left figure of 5.3 have the same multivariate pdf, but those in the right figure not necessarily so. A random function whose multivariate pdf is invariant under rotation is called a *statistically isotropic* random function.

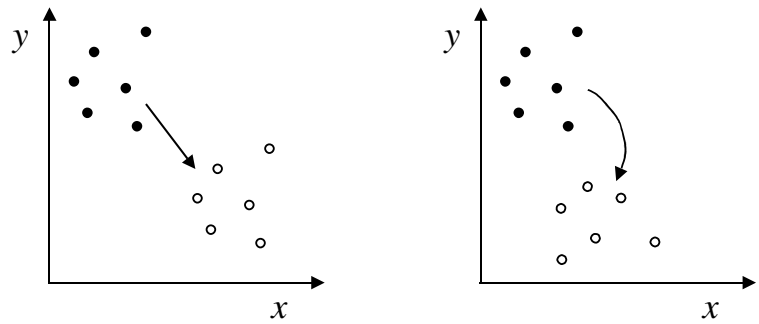



Figure 5.3. Translation of a configuration of points without rotation (left figure) and with rotation (right figure)

### 5.3.2 Ergodic random functions

One could ask why the property of stationarity is so important. The reason lies in estimating the statistical properties of a random function such as the mean, the variance and the covariance. In case of a random variable, such as the outcome of throwing dice, we can do a series of random experiments (actually throwing the dice) and estimate the mean and variance from the results of these experiments. This is not the case with random functions. To estimate the statistical properties of a random function, we should be able to draw a large number of realisations. However, in practice, we only have one realisation of the random function, namely reality itself. So we must be able to estimate all the relevant statistics of the random function from a single realisation. It turns out that this is only possible for stationary random functions. The reason is that strict stationarity actually says that all statistical properties are the same, no matter where you are. For instance, suppose we want to estimate the mean $\mu_Z(t_1)$ at a certain point $t_1$. The normal procedure would be to take the average of many realisations at point $t_1$, which is impossible because we only have one realisation (reality). However, if the random function is stationary the pdf $f_Z(z;t)$ at any location is the same and therefore also the mean. This also means that *within any single* realisation we have at every location a sample from the *same* pdf $f_Z(z;t)$. So, the mean can also be estimated if we take a sufficient number of samples from a single realisation, such as our reality. This is illustrated further in Figure 5.4. This property of a random function, i.e. being able to estimate statistical properties of a random function from a large number of samples of a

single realisation is called *ergodicity*. Apart from the random function being strict stationary, there is another condition necessary for ergodicity to apply. The samples from the single realisation should be taken from a large enough period of time or, in the spatial case, large enough area.

A more formal definition of a random function that is ergodic in its mean is:

$$\lim_{T\to\infty}\frac{1}{T}\int_{T} z(t)dt = \int_{-\infty}^{\infty} z\, f_Z(z;t)dz = \mu_Z \tag{5.6}$$

So the integral over the probability distribution (the ensemble) can be replaced by a temporal (or spatial or spatio-temporal) integral of a very large interval $T$ (or area or volume). Similarly, a random function is said to be covariance-ergodic if:

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T^2}\int_{T} z(t)z(t+\tau)dt &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{z_1-\mu_Z\}\{z_2-\mu_Z\}f(z_1,z_2;t,t+\tau)dz_1dz_2 \\ &= \mathrm{COV}[Z(t),Z(t+\tau)] \qquad \forall\tau\end{aligned} \tag{5.7}$$

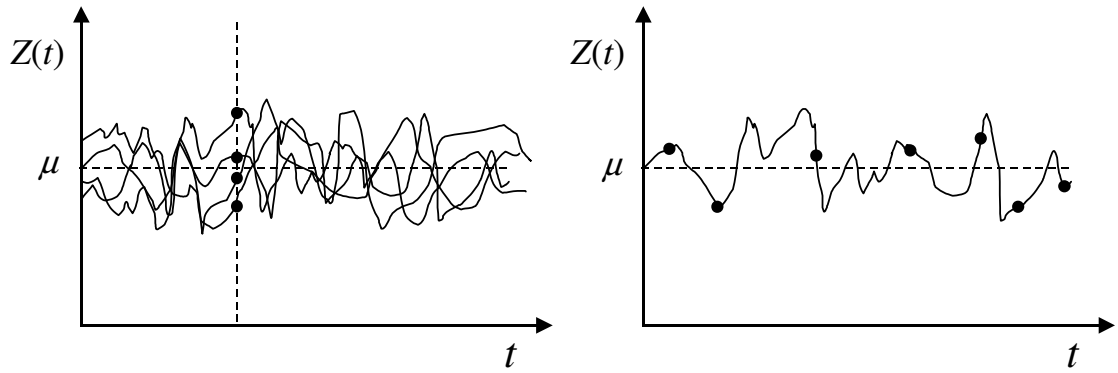



*Figure 5.4. Ergodic random functions: the average of (observations from) a single realisation is the same as the average of many realisations at a given location*

### 5.3.3 Second order stationary random functions

A weaker form of stationarity is second order stationarity. Here we require that the bivariate (or two-point) probability distribution is invariant under translation:

$$f(z_1,z_2;\ t_1,t_2) = f(z_1,z_2;\ t_1+t',t_2+t') \quad \forall t_1,t_2,\tau \tag{5.8}$$

Often, the number of observations available is only sufficient to estimate the mean, variance and covariances of the random function. So in practice, we require only

ergodicity and therefore only stationarity for the mean, variance and covariances. Hence, an even milder form of stationarity is usually assumed which is called *wide sense stationarity*. For wide sense stationary random functions (also called homogenous random functions) the mean and variance do not depend on $t$ (or $\mathbf{x}$) and the covariance depends only on the separation distance between two points in time (or space):

$$\begin{aligned} &\mu_Z(t) = \mu_Z \text{ is constant} \\ &\sigma_Z^2(t) = \sigma_Z^2 \text{ is constant and finite} \\ &\text{COV}[Z(t_1), Z(t_2)] = C_Z(t_2 - t_1) = C_Z(\tau) \end{aligned} \tag{5.9}$$

The graph describing the covariance as a function of the distance $\tau=t_2-t_1$ (also called *lag*) is called the covariance function. In case of wide sense stationarity, the covariance function for $t_2=t_1$ is equal to the variance and decreases to zero when the distances $t_2-t_1$ becomes larger. This means that random variables at sufficiently large distances are not correlated.

For random space functions weak sense stationarity means that

$$\text{COV}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = C_Z(\mathbf{x}_2 - \mathbf{x}_1) = C_Z(\mathbf{h}) \tag{5.10}$$

where $\mathbf{h} = \mathbf{x}_2 - \mathbf{x}_1$ is the difference vector between two locations (see Figure 5.5). If the random function is also isotropic we have that

$$\text{COV}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = C_Z(|\mathbf{x}_2 - \mathbf{x}_1|) = C_Z(|\mathbf{h}|) = C_Z(h) \tag{5.11}$$

where lag $h = |\mathbf{h}|$ is the norm (length) of the difference vector. If the covariance function is divided by the variance we obtain the correlation function: $\rho_Z(\tau) = C_Z(\tau)/\sigma_Z^2$ and $\rho_Z(\mathbf{h}) = C_Z(\mathbf{h})/\sigma_Z^2$.

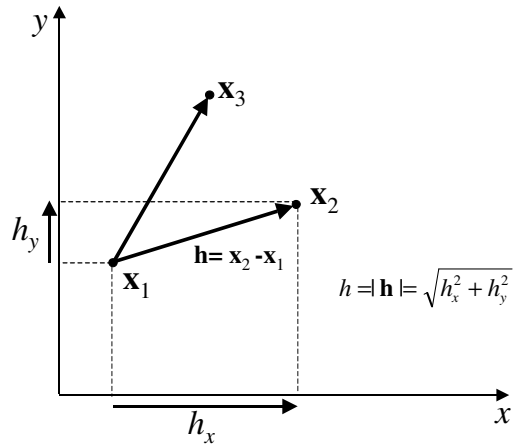



*Figure 5.5 In case of spatial random functions the covariance depends on the lag vector* $\mathbf{h}=\mathbf{x}_2-\mathbf{x}_1$ *with length h. In case of an isotropic RSF the covariance between* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *is the same as between* $\mathbf{x}_1$ *and* $\mathbf{x}_3$*, which implies that the covariance only depends on the length of the vector h.*

For data that are regularly positioned in time (or space) the covariance function can be estimated as:

$$\hat{C}_Z(k\Delta t) = \frac{1}{n-k}\sum_{i=1}^{n-k}(z_i - \hat{\mu}_Z)(z_{i+k} - \hat{\mu}_Z) \tag{5.12}$$

with $k\Delta t$ the lag (in units of time or space) and $\Delta t$ the time/distance between observations. This estimator is often used in time series analysis (see chapter 6). In case of irregularly positioned data in space the covariance function can be estimated as:

$$\hat{C}_Z(\mathbf{h}) = \frac{1}{n(\mathbf{h})}\sum_{i=1}^{n(\mathbf{h})}[z(\mathbf{x}) - \hat{\mu}_Z][z(\mathbf{x}_i + \mathbf{h} \pm \Delta\mathbf{h}) - \hat{\mu}_Z] \tag{5.13}$$

where $\mathbf{h}$ is the lag (which is a vector in space), $\Delta\mathbf{h}$ is a lag-tolerance which is needed to group a number of data-pairs together to get stable estimates for a given lag and $n(\mathbf{h})$ are the number of data-pairs that are a distance (and direction) $\mathbf{h} \pm \Delta\mathbf{h}$ apart.

In Figure 5.6 two different hydrological time series are shown. Figure 5.6a shows a time series of maximum discharge for the Rhine river at the Lobith Gauging station. The values are uncorrelated in time. Also given in Figure 5.6b is the theoretical correlation function that fit these data. The underlying stochastic process that belongs to this theoretical correlation function is called *white noise*, which is a special process consisting of uncorrelated Gaussian deviates at every two times no matter how small the lage $\tau$ between these times. Figure 5.6c shows a realisation of this continuous process. If the process is sampled at the same discrete times as the maximum discharge series (once a year) we obtain a series that looks similar as that shown in Figure 5.6a. Figure 5.6d shows a time series of groundwater depth data observed once a day for the years 1994-1995 at a town called De Bilt in The Netherlands. The correlation function estimated from these data is shown in Figure 5.6e. Also shown is a fitted correlation function belonging to the continuous process shown in Figure 6.6f. Again, when samples at the same discrete times are taken as that of the groundwater head series we obtain a discrete process with similar statistical properties as the original series. This shows that a discrete time series can be modelled with a discrete random function, as will be shown in chapter 6, but also as a discrete sample of a continuous random function.

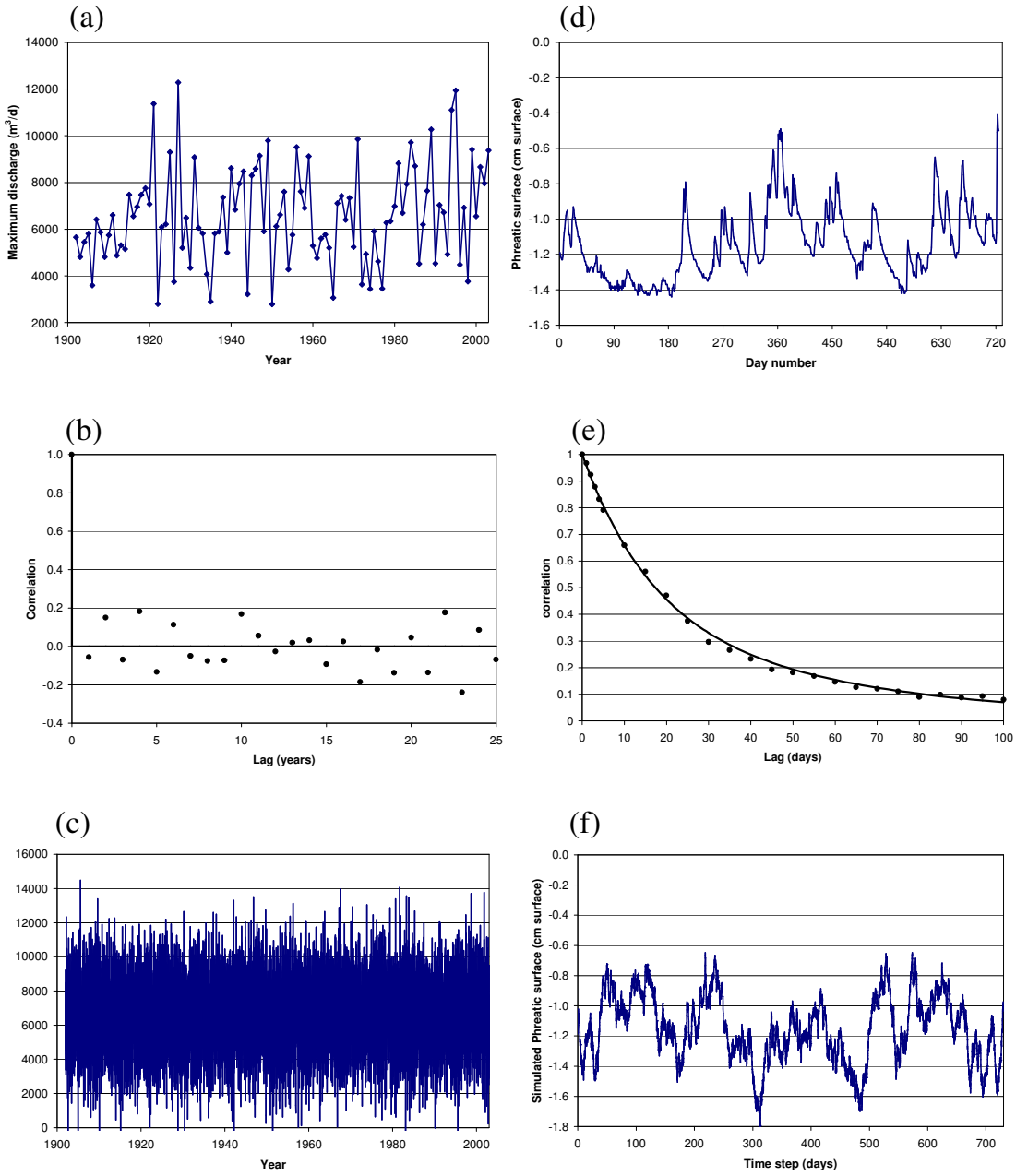



*Figure 5.6 a) Time series of yearly maximum values of daily discharge for the Rhine river at Lobith; b) estimated correlation function and fitting correlation functionof a continuous random function: white noise; c) realisation of white noise; d) time series of water table depths at De Bilt; e) estimated correlation function from groundwater level time series and fitted correlation function of a continuous RF; e) realisation of this Random Function..*

There are numerous models for modelling the covariance function of wide sense stationary processes. Table 5.1 shows four of them for isotropic random functions. The parameter *a* is called the correlation scale or integral scale of the process, and is a measure for the length over which the random variables at two locations of the random function (RF) are still correlated. In case of anisotropic random functions, for instance in

three spatial dimensions with integral scales $a_x$, $a_y$, $a_z$, the same model can be used as those shown in Table 5.1 by replacing $h/a$ with the following transformation:

$$\frac{h}{a} \Rightarrow \sqrt{\left(\frac{h_x}{a_x}\right)^2 + \left(\frac{h_y}{a_y}\right)^2 + \left(\frac{h_z}{a_z}\right)^2} \tag{5.14}$$

This form of anisotropy, where only the degree of correlation various with direction but not the variance of the process is called *geometric anisotropy*.

*Table 5.1 A number of possible covariance models for wide sense stationary random functions*

| | |
|---|---|
| Exponential covariance | $C_z(h) = \sigma_Z^2 e^{-(h/a)} \quad h \geq 0,\ a > 0$ |
| Gaussian covariance | $C_z(h) = \sigma_Z^2 e^{-(h^2/a^2)} \quad h \geq 0,\ a > 0$ |
| Spherical covariance | $C_z(h) = \begin{cases} \sigma_Z^2 \left[1 - \frac{3}{2}\left(\frac{h}{a}\right) + \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{if } h < a \\ 0 & \text{if } h \geq a \end{cases}$ $h \geq 0,\ a > 0$ |
| Hole effect (wave) model | $C_z(h) = \sigma_Z^2 \frac{b \sin(h/a)}{h} \quad h \geq 0, a > 0$ |
| White noise model* | $\rho_z(h) = \begin{cases} 1 & h = 0 \\ 0 & h > 0 \end{cases}$ |

* The white noise process has infinite variance, so strictly speaking it is not wide sense stationary. Here, we thus only provide the correlation function that *does* exist.

### 5.3.4 Relations between various forms of stationarity

A strict sense stationary random function is also second order stationary and is also wide sense stationary, but not necessarily the other way around. However, if a random function is wide sense stationary and its multivariate pdf is a Gaussian distribution (Equation 3.87), it is also second order stationary[3] and also a strict sense stationary random function. More important, a wide sense stationary random function that is multivariate Gaussian (and thus also strict stationary) is completely characterised by only a few statistics: a constant mean $\mu_Z(t) = \mu_Z$ and a covariance function $C_Z(t_2\text{-}t_1)$ that is only dependent on the separation distance. So to recapitulate (arrow means "implies")

In general:

| Type of stationarity: | Strict sense ⟶ | Second order ⟶ | Wide sense |
|---|---|---|---|
| Property: | Multivariate pdf translation invariant | Bivariate pdf translation invariant | Mean and variance translation invariant |

[3] Often in the literature the term "second order stationary" is used when in fact one means "wide sense stationary".

If the multivariate pdf is Gaussian:

| Type of stationarity: | Wide sense ⟷ | Second order ⟷ | Strict sense |
|---|---|---|---|
| Property: | Mean and variance translation invariant | Bivariate pdf translation invariant | Multivariate pdf translation invariant |

### 5.3.5 Intrinsic random functions

An even milder form of stationary random functions are *intrinsic* random functions. For an intrinsic random function we require (we show the spatial form here):

$$E[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)] = 0 \quad \forall \mathbf{x}_1, \mathbf{x}_2 \tag{5.15}$$

$$E[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)]^2 = 2\gamma_Z(\mathbf{x}_2 - \mathbf{x}_1) \quad \forall \mathbf{x}_1, \mathbf{x}_2 \tag{5.16}$$

So the mean is constant and the expected quadratic difference is only a function of the lag-vector $\mathbf{h} = \mathbf{x}_2 - \mathbf{x}_1$. The function $\gamma(\mathbf{x}_2 - \mathbf{x}_1)$ is called the *semivariogram* and is defined as:

$$\gamma_Z(\mathbf{x}_1, \mathbf{x}_2) = \gamma_Z(\mathbf{x}_2 - \mathbf{x}_1) = \tfrac{1}{2}\mathrm{E}[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)]^2 \tag{5.17}$$

The semivariogram can be estimated from observations as (similarly in the temporal domain):

$$\hat{\gamma}_Z(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h} \pm \Delta\mathbf{h})\}^2 \tag{5.18}$$

Table 5.2 shows examples of continuous semivariogram models that can be fitted to estimated semivariograms.

*Table 5.2 A number of possible semivariance models for intrinsic random functions*

| | |
|---|---|
| Exponential model | $\gamma_z(h) = c[1 - e^{-(h/a)}] \quad h \geq 0;\ a, c > 0$ |
| Gaussian model | $\gamma_z(h) = c[1 - e^{-(h^2/a^2)}] \quad h \geq 0;\ a > 0$ |
| Spherical model | $\gamma_z(h) = \begin{cases} c\left[\frac{3}{2}\left(\frac{h}{a}\right) + \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{if } h < a \\ c & \text{if } h \geq a \end{cases}$ $h \geq 0;\ a, c > 0$ |
| Hole effect (wave) model | $\gamma_z(h) = c\left[1 - \frac{b\sin(h/a)}{h}\right] \quad h \geq 0; a, c > 0$ |
| Pure nugget model | $\gamma_z(h) = \begin{cases} 0 & h = 0 \\ c & h > 0 \end{cases} \quad c > 0$ |
| Power model | $\gamma_Z(h) = ah^b$ $h \geq 0;\ a > 0;\ 0 \leq b \leq 2$ |

The semivariogram and the covariance function of a wide sense stationary random function are related as follows:

$$\gamma_Z(\mathbf{x}_2 - \mathbf{x}_1) = \sigma_Z^2 - C_Z(\mathbf{x}_2 - \mathbf{x}_1) \tag{5.19}$$

This means that the semivariogram and the covariance function are mirror images with $c = \sigma_Z^2$ as can be seen in Figure 5.7. This also means that where the covariance function becomes zero for large enough separation distances, the semivariogram will reach a plateau (called the *sill* of the semivariogram) that is equal to the variance. The distance at which this occurs (called the *range* of the semivariogram) is the distance beyond which values on the random function are no longer correlated. The first five models of table 5.2 are semivariogram models that imply wide sense stationary functions whith $c = \sigma_Z^2$. For the sixth model, the power model, this is not the case. Here, the variance does not have to be finite, while the semivariance keeps on growing with increasing lag. This shows that if a random function is wide sense stationary, it is also intrinsic. However, an intrinsic random function does not have to be wide sense stationary, i.e. if the semivariogram does not reach a sill.

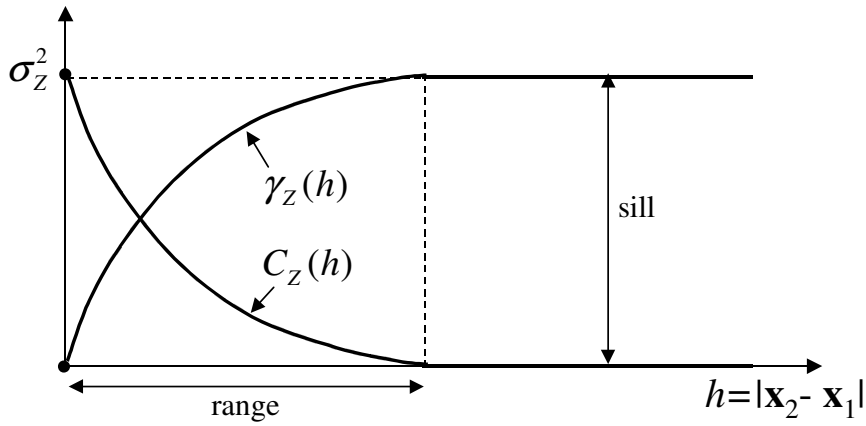



*Figure 5.7. Covariance function and semivariogram for a wide sense stationary random function*

### 5.3.6 Integral scale and scale of fluctuation

The *integral scale* or correlation scale is a measure of the degree of correlation for stationary random processes and is defined as the area under the correlation function.

$$I_{Z(t)} = \int_0^\infty \rho(\tau) d\tau \tag{5.20}$$

For the exponential, Gaussian and spherical correlation functions the integral scales are equal to *a,* $(a\sqrt{\pi})/2$ and (3/8)*a* respectively. Given that the correlation functions of wide sense stationary processes are even functions, i.e. $\rho(\tau) = \rho(-\tau)$, another measure of correlation is the *scale of fluctuation* defined as:

$$\theta = \int_{-\infty}^{\infty} \rho(\tau) d\tau = 2I_{Z(t)} \tag{5.21}$$

For a 2D and 3D random space function the integral scales are defined as:

$$I_{Z(\mathbf{x})} = \left[ \frac{4}{\pi} \int_0^{\infty} \int_0^{\infty} \rho(h_1, h_2) dh_1 dh_2 \right]^{1/2} ; I_{Z(\mathbf{x})} = \left[ \frac{6}{\pi} \int_0^{\infty} \int_0^{\infty} \rho(h_1, h_2, h_3) dh_1 dh_2 dh_3 \right]^{1/3} \tag{5.22}$$

## 5.4 Conditional random functions

In this section we investigate what happens if observations are done on a random function. Suppose that we have a stationary random function in time that is observed at a number of locations. Suppose for the moment that these observations are without error. Figure 5.8 shows a number of realisations of a continuous time random function that is observed at four locations without error. It can be seen that the realisations are free to vary and differ between the observation points but are constraint, i.e. *conditioned*, by these points. This can be seen when comparing the pdfs at two locations $t_1$ and $t_2$. It can be seen that uncertainty is larger further from an observation ($t_1$) than close to an observation ($t_2$). This is intuitively correct because an observation is able to reduce uncertainty for a limited interval proportional to the integral scale of the random function. At a distance larger than the range, the random values are no longer correlated with the random variable at the observation location and the uncertainty is as large (the variance of the pdf is a large) as that of the random function without observations.

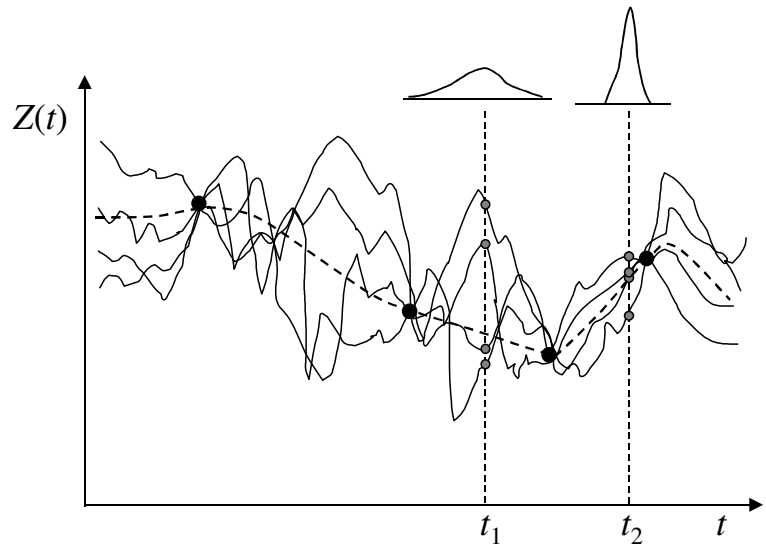



*Figure 5.8. Realisations of a random function that is conditional to a number of observations; dashed line is the conditional mean.*

The random function that is observed at a number of locations and/or times is called a conditional random function and the probability distributions at locations $t_1$ and $t_2$ conditional probability density functions (cpdfs): $f_Z(z;t_1 \mid y_1,..,y_m)$, $f_Z(z;t_2 \mid y_1,..,y_m)$. Where $y_1,\ldots, y_m$ are the observations. The complete conditional random function is defined by the conditional multivariate pdf: $f(z_1,z_2,\ldots. z_N ; t_1,t_2,\ldots. t_N \mid y_1,..,y_m)$. The conditional multivariate pdf can in theory be derived from the (unconditional) multivariate pdf using Bayes' rule. However, this is usually very cumbersome. An alternative way of obtaining all the required statistics of the conditional random function is called stochastic simulation. In chapters 7 and 8 some methods are presented for simulating realisations of both unconditional and conditional random functions.

The conditional distribution of $Z(s_1,t)$, or its mean value (see dashed line) and variance, can also be obtained directly through geostatistical prediction or kriging (chapter 7) and state-space prediction methods such as the Kalman filter (chapter 9). These methods use the observations and the statistics (e.g. semivariogram or covariance function) of the random function (statistics estimated from the observations) to directly estimate the conditional distribution or its mean and variance.

## 5.5 Spectral representation of random functions

The correlation function of the time series of water table depth at De Bilt in Figure 5.6 has only been analysed for two years. Had we analysed a longer period we would have seen a correaltion function with periodic behaviour, such as the Hole-effect model in Tables 5.1 and 5.2. Figure 5.9 shows the correlation function of the daily observations of discharge of the Rhine river at Lobith. A clear periodic behaviour is observed as well. The periodic behaviour which is also apparent in the time series (see Figure 4.1) is caused by the fact that evaporation which is driven by radiation and temperature has a clear seasonal character at higher and lower latitudes and temperate climates. In arctic climates the temperature cycle and associated snow accumulation and melt cause seasonality, while in the sub-tropics and semi-arid climates the occurrence of rainfall is strongly seasonal.

In conclusion, most hydrological time series show a seasonal variation. This means that to analyse these models with stationary random functions requires that this seasonality is removed (see for instance chapter 6). The occurrence of seasonality has also inspired the use of spectral methods in stochastic modelling, although it must be stressed that spectral methods are also very suitable for analysing stationary random functions.

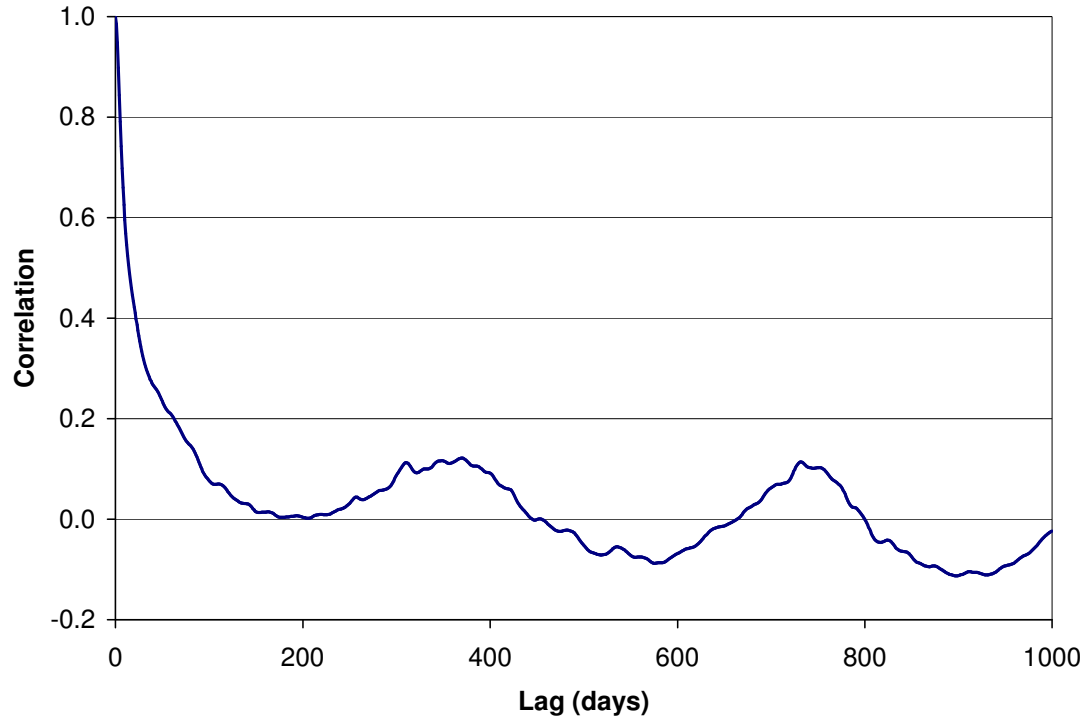



*Figure 5.9 Correlation function of daily averaged discharge of the river Rhine at Lobith.*

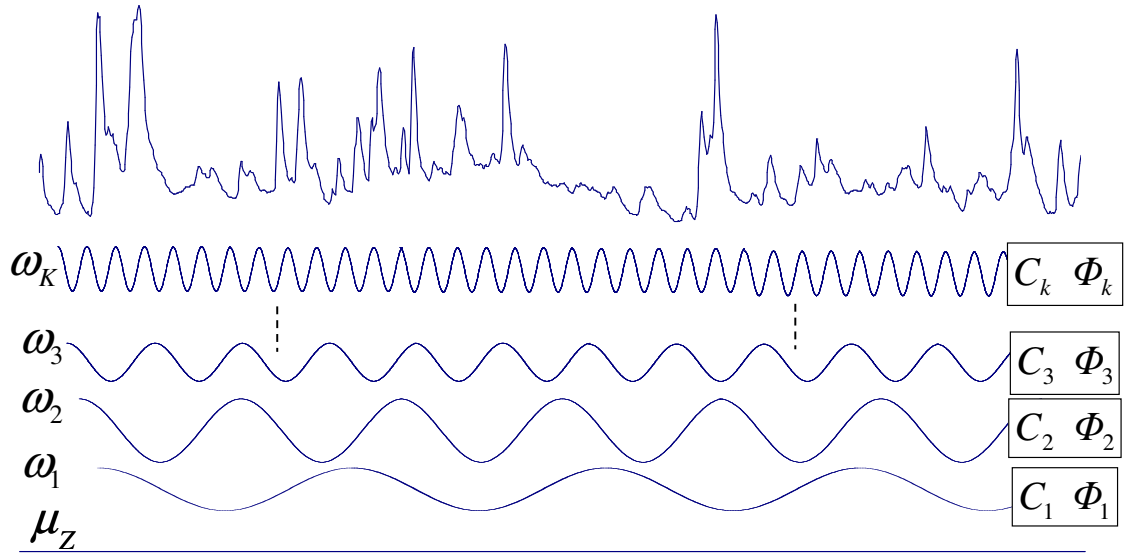
### 5.5.1 Spectral density function

We will therefore start by a spectral representation of a stationary random function $Z(t)$. Such a presentation means that the random function is expressed as a sum of its mean $\mu_Z$ and $2K$ sinusoids with increasing frequencies, where each frequency has a random amplitude $C_k$ and random phase angle $\Phi_k$:

$$Z(t) = \mu_Z + \sum_{k=-K}^{K} C_k \cos(\omega_k t + \Phi_k) \tag{5.23}$$

with $C_k = C_{-k}$, $\Phi_k = \Phi_{-k}$ and $\omega_k = \pm\frac{1}{2}(2k-1)\Delta\omega$

Figure 5.10 shows schematically how a stationary random function is decomposed into random harmonics.




$$Z(t) = \mu_Z \pm C_1 \cos(\omega t + \Phi_1) \pm C_2 \cos(2\omega t + \Phi_2) \pm ... \pm C_K \cos(K\omega t + \Phi_K)$$

*Figure 5.10 Schematic of the spectral representation of a stationary random function as a decomposition of the random signal into harmonics of increasing frequency with random amplitude $C_k$ and random phase angle $\Phi_k$:*

Based on this representation it can be shown (see Vanmarcke pp. 84-86) that the following relations hold:

$$C_Z(\tau) = \int_{-\infty}^{\infty} S_Z(\omega) \cos(\omega\tau) d\omega \tag{5.24}$$

$$S_Z(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C_Z(\tau) \cos(\omega\tau) d\tau \tag{5.25}$$

These relations are known as the Winer-Khinchine relations. The function $S_Z(\omega)$ is known as the *spectral density* function of the random process and Equations (5.24) thus show that the covariance function is a Fourier transform of the spectrum and vice versa: they form a Fourier pair.

The physical meaning of the spectral density function can best be understood by setting the lag $\tau$ equal to zero in (5.24). We then obtain:

$$C_Z(0) = \sigma_Z^2 = \int_{-\infty}^{\infty} S_Z(\omega) d\omega \tag{5.26}$$

It can be seen that the variance of the random function is equal to the integral over the spectral density. This means that the variance is a weighted sum of variance components,

where each component consists of a random harmonic function of a given frequency. The spectral density then represents the weight of each of the attributing random harmonics, i.e. the relative importance of each random harmonic in explaining the total variance of the random function. It is easy to see the analogy with the electromagnetic spectrum where the total energy of electromagnetic radiation (which is analogous to the variance of our random signal) is found by the area under the spectrum and can be attributed to relative contributions from different wavelengths.

In table 5.3 expressions are given of spectral density functions belonging to some of the covariance functions given in Table 5.1. Figure 5.11 (From Gelhar, 1993) shows typical realisations of the random functions involved, their correlation function and the associated spectrum. What can be seen from this is that the spectrum of white noise is a horizontal line, implying an infinite variance according to Equation (5.26). This shows that white noise is a mathematical construct, and not a feasible physical process: the area under the spectrum is a measure for the total energy of a process. This area is infinitely large, such that all the energy in the universe would not be sufficient to generate such a process. In practice one often talks about wide band processes, where the spectrum has a wide band of frequencies, but encloses a finite area.

*Table 5.3 A number of possible covariance function and associated spectral density functions* ($\tau=|t_2-t_1|$)

| | | |
|---|---|---|
| Exponential | $C_z(\tau) = \sigma_Z^2 e^{-(\|t\|/a)}$ $a>0$ | $S_Z(\omega) = \frac{a\sigma_Z^2}{\pi(1+a^2\omega^2)}$ |
| Random harmonic: $Z(t) = a\cos(\omega_0 t + \phi)$ where $a, \omega_0$ are deterministic constants and $\phi$ is random | $C_Z(\tau) = \frac{a^2}{2}\cos(\omega_0 \tau)$ $\tau \geq 0,\ a, \omega_0 > 0$ | $S_Z(\omega) = \frac{a^2}{4}\delta(\omega - \omega_0)$ |
| Hole effect (wave) model | $C_z(\tau) = \sigma_Z^2(1-\|t\|/a)e^{-(\|\tau\|/a)}$ $a>0$ | $S_Z(\omega) = \frac{a^3\sigma_Z^2\omega^2}{\pi(1+a^2\omega^2)^2}$ |
| White noise model | $\rho_z(\tau) = \begin{cases} 1 & \tau = 0 \\ 0 & \tau > 0 \end{cases}$ | $S_Z(\omega) = c$ |

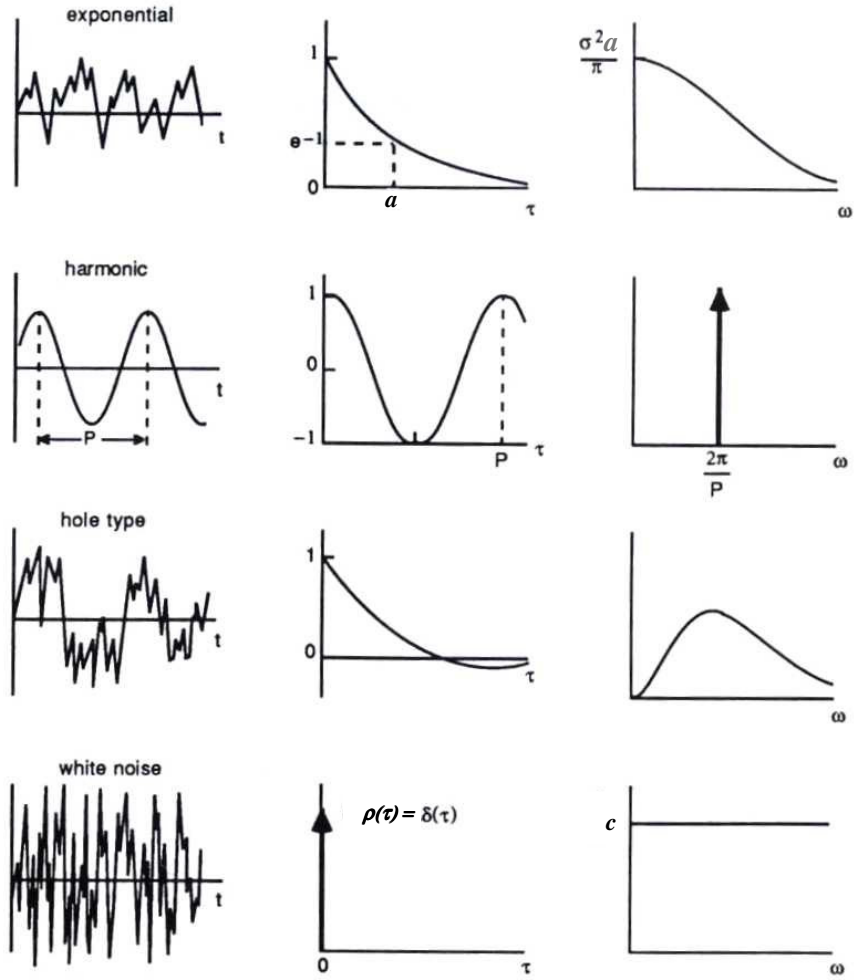



*Figure 5.11 Schematic examples of covariance function-spectral density pairs (adapted from Gelhar, 1993).*

The spectral density function is an *even function* (as is the covariance function): $S_Z(\omega) = S_Z(-\omega)$. This motivates the introduction of the *one-sided* spectral density function $G_Z(\omega) = 2S_Z(\omega), \quad \omega \geq 0$. The Wiener-Khinchine relations relations then become:

$$C_Z(\tau) = \int_0^\infty G_Z(\omega)\cos(\omega\tau)d\omega \tag{5.27}$$

$$G_Z(\omega) = \frac{2}{\pi}\int_0^\infty C_Z(\tau)\cos(\omega\tau)d\omega \tag{5.28}$$

Sometimes it is convenient to work with the normalised spectral density functions, by dividing the spectra by the variance: $s_Z(\omega) = S_Z(\omega)/\sigma_Z^2$ and $g_Z(\omega) = G_Z(\omega)/\sigma_Z^2$. For instance, from (5.25) we can see that there is a relation between the normalised spectral

density function $s_Z(\omega)$ and the scale of fluctuation. Setting $\omega = 0$ in Equation (5.25) we obtain:

$$s_Z(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\rho_Z(\tau)d\omega = \frac{\theta}{2\pi} \tag{5.29}$$

### 5.5.2 Formal (complex) spectral representation

Often a more formal definition of the spectral density is used in the literature based on the formulation in terms of complex calculus. Here the random function $Z(t)$ is defined as the real part of a *complex random function* $Z^*(t)$:

$$Z(t) = \mathrm{Re}\{Z^*(t)\} = \mathrm{Re}\left\{\mu_Z + \sum_{k=-K}^{K} X_k e^{i\omega_k t}\right\} \underset{K\to\infty}{\rightarrow} \mathrm{Re}\left\{\mu_Z + \int_{-\infty}^{\infty} e^{i\omega t} dX(\omega)\right\} \tag{5.30}$$

with $\omega_k = \frac{1}{2}(2k-1)\Delta\omega$ the frequency, and $X_k$ a complex random number representing the amplitude. This equation entails that the complex random process is decomposed into a large number of complex harmonic functions $e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$ with random complex amplitude, Given this representation it is possible derive the Wiener-Khinchine equations as (Vanmarcke, 1983, p. 88):

$$C_Z(\tau) = \int_{-\infty}^{\infty} S_Z(\omega) e^{i\omega\tau} d\omega \tag{5.31}$$

$$S_Z(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} C_Z(\tau)^{i\omega\tau} d\tau \tag{5.32}$$

It can be shown (Vanmarcke, 1983, p.94) that Equations (5.31) and (5.32) are mathematically equivalent to Equations (5.24) and (5.25) respectively.

### 5.5.3 Estimating the spectral density function

For wide sense stationary random functions the spectral density can be estimated from estimated covariance function as:

$$S_Z(\omega_i) = \frac{1}{2\pi}\left(\lambda_0 C_Z(0) + 2\sum_{k=1}^{M}\lambda_k \hat{C}_Z(k\Delta\tau)\cos(\omega_i k)\right) \tag{5.33}$$

with $\omega_i = i\pi/M, |i| = 1,..,M$. The weights $\lambda_k$ are necessary to smooth the covariances before performing the transformation. This way a smoothed spectral density function is

obtained displaying only the relevant features. There are numerous types of smoothing weights. Two frequently used expressions are the *Tukey window*

$$\lambda_k = \frac{1}{2}\left(1+\cos\frac{k\pi}{M}\right), \quad k = 0,1,....,M \tag{5.34}$$

and the *Parzen window*

$$\lambda_k = \begin{cases} 1-6\left(\frac{k}{M}\right)^2+6\left(\frac{k}{M}\right)^3 & 0 \le k \le M/2 \\ \left(2-\frac{k}{M}\right)^3 & M/2 \le k \le M \end{cases} \tag{5.35}$$

The highest frequency that is analysed is equal to $f_{\max} = \omega_{\max}/2\pi = 0.5$. This is the highest frequency that can be estimated from a time series, i.e. half of the frequency of the observations. This frequency is called the Nyquist frequency. So if hydraulic head is observed once per day than the highest frequency that can be detected is one cycle per two days. The smallest frequency (largest wavelength) that can be analysed depends on the discretisation $M$: $f_{\min} = \pi/(2\pi M) = (1/2M)$, where $M$ is also the cutoff level (maximum lag considered) of the covariance function. The width of the smoothing windows is adjusted accordingly.

As an example the spectrum of the daily discharge data of the Rhine river at Lobith (Figure 4.1 and Figure 5.9 for the correlation function) was estimated using a Parzen window with $M$=9000. Figure 5.12 shows the normalised spectral density function so obtained. Clearly, small frequencies dominate with a small peak between 4 and 5 years. Most prominent of course, as expected, there is a peak at a frequency of once a year, which exemplifies the strong seasonality in the time series.

### 5.3.4 Spectral representations of random space functions

If we extend the previous to two dimensions the stationary random function $Z(x_1,x_2)$ can be expressed in terms of random harmonics as:

$$Z(x_1,x_2) = \mu_Z + \sum_{k_1=-K_1}^{k_1=K_1}\sum_{k_2=-K_2}^{k_1=K_2} C_{12}\cos(\omega_{k_1}x_1 + \omega_{k_2}x_2 + \Phi_{12}) \tag{5.36}$$

with $= C_{12}$ and $\Phi_{12}$ random amplitude and phase angle belonging to frequency $\omega_{k_i}, i = 1,2$:

$$\omega_{k_i} = \pm\frac{(2k_i\Delta\omega_i - 1)}{2} \tag{5.37}$$

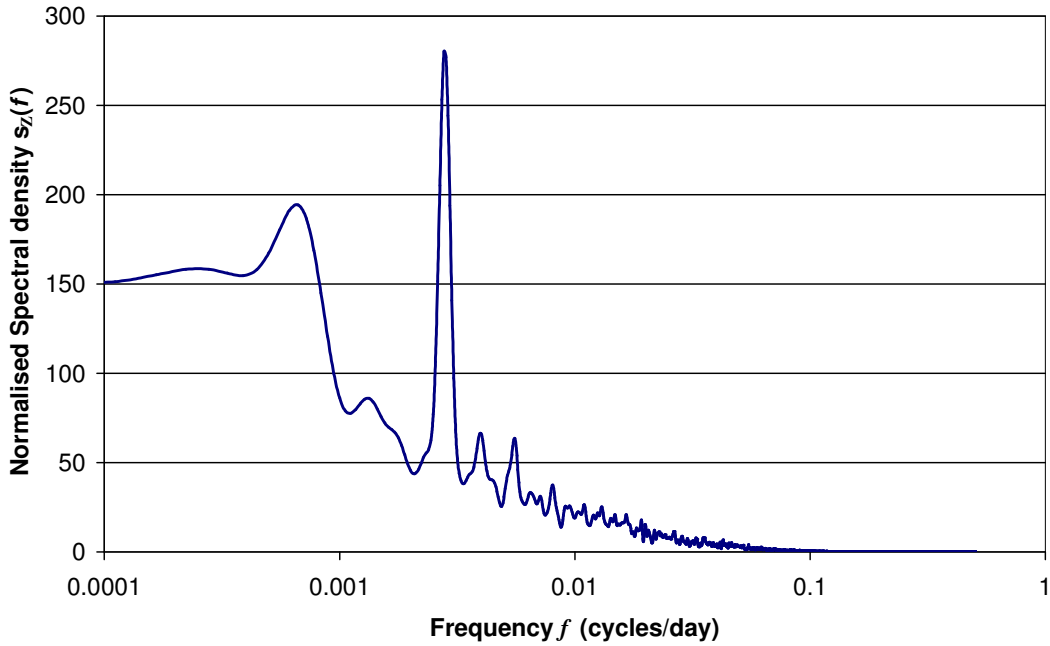



*Figure 5.11 Normalised spectral density function of daily averaged discharge of the river Rhine at Lobith.*

The Wiener-Khinchine equations then become

$$C_Z(h_1,h_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_Z(\omega_1,\omega_2)\cos(\omega_1 h_1 + \omega_1 h_1)d\omega_1 d\omega_2 \tag{5.38}$$

$$S_Z(\omega_1,\omega_2) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C_Z(h_1,h_2)\cos(\omega_1 h_1 + \omega_1 h_1)dh_1 dh_2 \tag{5.39}$$

The variance is given by the volume under the 2D spectral density function.

$$\sigma_Z^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_Z(\omega_1,\omega_2)d\omega_1 d\omega_2 \tag{5.40}$$

If we use a vector-notation we have: $\boldsymbol{\omega} = (\omega_1,\omega_2)^{\mathrm{T}}, \mathbf{h} = (h_1,h_2)^{\mathrm{T}}$ and $\omega_1 h_1 + \omega_1 h_1 = \boldsymbol{\omega}\cdot\mathbf{h}$. A short hand way of writing (5.38) and (5.39) results:

$$C_Z(\mathbf{h}) = \int_{-\infty}^{\infty} S_Z(\boldsymbol{\omega})\cos(\boldsymbol{\omega}\cdot\mathbf{h})d\boldsymbol{\omega} \tag{5.41}$$

$$S_Z(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} C_Z(\mathbf{h})\cos(\boldsymbol{\omega}\cdot\mathbf{h})d\mathbf{h} \tag{5.42}$$

These equations are valid for higher dimensional random functions, where $1/(2\pi)^D$ ($D$ = dimension process) replaces $1/(2\pi)^2$ in (5.42). The more formal definition using complex calculus then gives:

$$Z(\mathbf{x}) = \mathrm{Re}\left\{\mu_Z + \int_{-\infty}^{\infty} e^{i\boldsymbol{\omega}\cdot\mathbf{x}} dX(\boldsymbol{\omega})\right\} \tag{5.43}$$

The Wiener Khinchine equations become

$$C_Z(\mathbf{h}) = \int_{-\infty}^{\infty} S_Z(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}\cdot\mathbf{h}} d\boldsymbol{\omega} \tag{5.44}$$

$$S_Z(\boldsymbol{\omega}) = \frac{1}{(2\pi)^D} \int_{-\infty}^{\infty} C_Z(\mathbf{h}) \exp^{-i\boldsymbol{\omega}\cdot\mathbf{h}} d\mathbf{h} \tag{5.45}$$

## 5.6 Local averaging of stationary random functions

Consider a stationary random function $Z(t)$ and consider the random function $Z_T(t)$ that is obtained by local moving averaging (see Figure 5.12):

$$Z_T(t) = \frac{1}{T} \int_{t-T/2}^{t+T/2} Z(\tau) d\tau \tag{5.46}$$

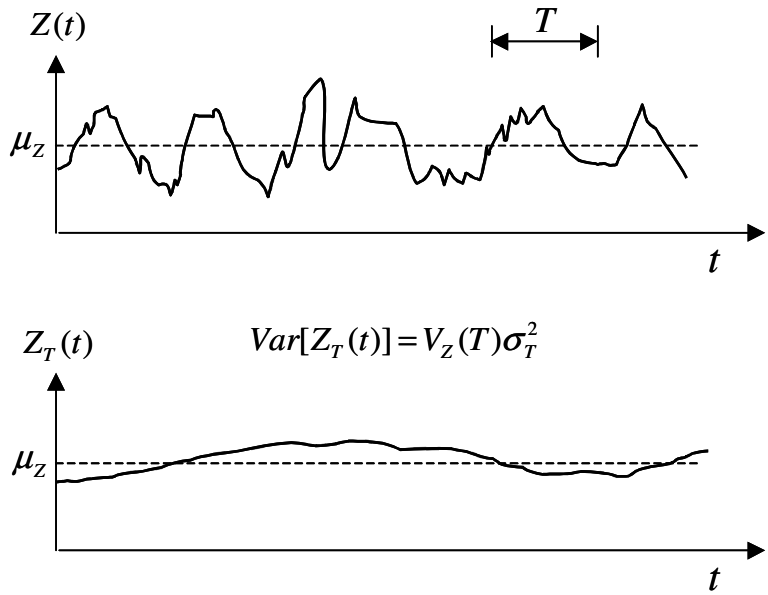



*Figure 5.12 Local (moving) averaging of a stationary random function*

Local averaging of a stationary random process will not affect the mean, but it does reduce the variance. The variance of the averaged process can be calculated as (without loss of generality we can set the mean to zero here):

$$\begin{aligned} Var[Z_T(t)] &= E[\frac{1}{T}\int_{t-T/2}^{t+T/2} Z(\tau_1)d\tau_1 \frac{1}{T}\int_{t-T/2}^{t+T/2} Z(\tau_2)d\tau_2] \text{ (because } Z \text{ is stationary)} = \\ & E[\frac{1}{T}\int_0^T Z(\tau_1)d\tau_1 \frac{1}{T}\int_0^T Z(\tau_2)d\tau_2] = \\ & \frac{1}{T^2}\int_0^T\int_0^T E[Z(\tau_1)Z(\tau_2)]d\tau_1 d\tau_2 = \\ & \frac{1}{T^2}\int_0^T\int_0^T C_Z(\tau_2-\tau_1)d\tau_1 d\tau_2 \end{aligned} \tag{5.47}$$

A new function is introduced that is called the variance function:

$$V_Z(T) = \frac{Var[Z_T(t)]}{\sigma_Z^2} \tag{5.48}$$

The variance function thus describes the reduction in variance when averaging a random function as a function of the averaging interval $T$. From (5.47) we can see that the variance function is related to the correlation function as:

$$V_Z(t) = \frac{1}{T^2}\int_0^T\int_0^T \rho_Z(\tau_2-\tau_1)d\tau_1 d\tau_2 \tag{5.49}$$

Vanmarcke (1983, p 117) shows that Equation (5.49) can be simplified to:

$$V_Z(t) = \frac{1}{T}\int_0^T\left(1-\frac{\tau}{T}\right)\rho_Z(\tau)d\tau \tag{5.50}$$

In Table 5.4 a number of correlation functions and their variance functions are given. If we examine the behaviour of the variance function for large $T$ we get (see Vanmarcke, 1983):

$$\lim_{T\to\infty} V_Z(T) \to \frac{\theta}{T} \tag{5.51}$$

where $\theta$ is scale of fluctuation. In table 5.1 the scale of fluctuation is also given for the three correlation models. The scale of fluctuation was already introduced earlier as a measure of spatial correlation of the random function and can also be calculated using the correlation function (Equation 5.21) or the spectral density (Equation 5.29). Equation

5.51 thus states that for larger averaging intervals the variance reduction through averaging is inversely proportional to the scale of fluctuation: the larger the scale of fluctuation the larger $T$ should be to achieve a given variance reduction. In practice, relation (5.51) is already valid for $T > 2\theta$ (vanmarcke, 1983).

*Table 5.4 Variance functions and scale of fluctuation for three different correlation functions* ($\tau=|t_2-t_1|$)

| | | |
|---|---|---|
| Exponential (first order autoregressive process) | $\rho_z(\tau) = e^{-(\tau/a)}$<br>$\tau \geq 0,\ a>0$ | $V(T) = 2\left(\frac{a}{T}\right)^2 \left(\frac{T}{a} - 1 + e^{-T/a}\right)$<br>$\theta = 2a$ |
| Second order Autoregressive process (see chapter 6) | $\rho_z(\tau) = \left[1+\frac{\tau}{a}\right]e^{-(\tau/a)}$<br>$\tau \geq 0,\ a>0$ | $V(T) = \frac{2a}{T}\left[2 + e^{-T/a} - \frac{2a}{T}(1-e^{-T/a})\right]$<br>$\theta = 4a$ |
| Gaussian correlation Function | $\rho_z(\tau) = e^{-(\tau^2/a^2)}$<br>$\tau \geq 0,\ a>0$ | $V(T) = \left(\frac{a}{T}\right)^2\left[\sqrt{\pi}\,Erf\left(\frac{T}{a}\right) - 1 + e^{-(T/a)^2}\right]$<br>$\theta = a\sqrt{\pi}$ |

The covariance of the averaged process is given by:

$$C_Z(T,\tau) = \frac{1}{T^2}\int_0^T \int_\tau^{\tau+T} C_Z(t_1,t_2)\,dt_1 dt_2 \tag{5.52}$$

Generally it is not easy to obtain closed form expressions for (5.52). However as shown in chapter 7, it is relatively easy to obtain values for this function through numerical integration.

We end this section by giving the equations for the 2D-spatial case, where it is straightforward to generalise these results to higher dimensions. The local average process is for an area $A=L_1L_2$ defined as:

$$Z_T(x_1,x_2) = \frac{1}{L_1L_2}\int_{x_1-L_1/2}^{x_1+L_1/2}\int_{x_2-L_2/2}^{x_2+L_2/2} Z(u_1,u_2)\,du_1 du_2 \tag{5.53}$$

The variance function is given by:

$$V_Z(L_1,L_2) = \frac{1}{L_1L_2}\int_0^{L_1}\int_0^{L_2}\left(1-\frac{h_1}{L_1}\right)\left(1-\frac{h_2}{L_2}\right)\rho_Z(h_1,h_2)\,dh_1 dh_2 \tag{5.54}$$

The limit of the variance function defines the spatial "scale of fluctuation" or *characteristic area* $\alpha$:

$$\lim_{L_1,L_2\to\infty} V_Z(L_1,L_2) \to \frac{\alpha}{L_1 L_2} \tag{5.55}$$

where $\alpha$ can calculated from the correlation function as follows:

$$\alpha = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_Z(u_1,u_2)du_1du_2 \tag{5.56}$$

The characteristic area $\alpha$ can also be obtained through the spectral representation by setting $\omega_1 = \omega_2 = 0$ in the Wiener-Khinchine relation (5.39)

$$s_Z(0,0) = S_Z(0,0)/\sigma_Z^2 = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_Z(h_1,h_2)dh_1dh_2 \tag{5.57}$$

Combining equations (5.56) and (5.57) then leads to:

$$\alpha = 4\pi^2 s_Z(0,0) \tag{5.58}$$

In Table 5.5 the various way of obtaining the scale of fluctuation and the characteristic area summarized:

*Table 5.5 Three ways of obtaining the scale of fluctuation (time) and characteristic area (2D space) (after Vanmarcke, 1983)*

| Scale of fluctuation $\theta$ | Characteristic area $\alpha$ |
|---|---|
| $\lim_{T\to\infty} TV_Z(T)$ | $\lim_{L_1,L_2\to\infty} L_1L_2V_Z(L_1,L_2)$ |
| $\int_{-\infty}^{\infty}\rho_Z(\tau)d\tau$ | $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_Z(u_1,u_2)du_1du_2$ |
| $2\pi s_Z(0)$ | $4\pi^2 s_Z(0,0)$ |

Finally, the covariance of the averaged random function in two dimensions is given by:

$$C_Z(L_1,L_2;h_1,h_2) = \frac{1}{(L_1L_2)^2}\int_0^{L_1}\int_0^{L_2}\int_{h_1}^{L_1+h_1}\int_{h_2}^{L_2+h_2} C_Z(x_1,y_1,x_2,y_2)dx_1dy_1dx_2dy_2 \tag{5.59}$$

The covariance of the spatially averaged random function is frequently used in geostatistical mapping, as explained in chapter 7, where its values are approximated with numerical integration. To limit the notational burden, Equation (5.59) is usually written in vector notation with $\mathbf{x}_1 = (x_1,y_1)^{\mathrm{T}}, \mathbf{x}_2 = (x_2,y_2)^{\mathrm{T}}, \mathbf{h} = (h_1,h_2)^{\mathrm{T}}$ and $A$=$L_1L_2$:

$$C_Z(A;\mathbf{h}) = \frac{1}{A^2}\int_A\int_{A_{+\mathbf{h}}} C_Z(\mathbf{x}_1,\mathbf{x}_2)d\mathbf{x}_1d\mathbf{x}_2 \tag{5.60}$$

## 5.7 Exercises

1. Give examples of hydrological variables that can be modelled with a continuous-valued and a discrete-valued 1) random series, 2) lattice process, 3) continuous-time process, 4) continuous-space process, 5) continuous space-time process, 6) time compound point process, 7) space-compound point process. Note that 14 combinations are asked for.

2. The covariance of a random function is described by: $C(h) = 10\exp(-h/30)$, and a constant mean. The pdf at a given location is the Gamma distribution. Is this process: a) wide sense stationary; b) second order stationary; d) strict stationary?

3. The covariance of a random function is described by: $C(h) = 10\exp(-h/30)$, and a constant mean. The pdf at a given location is the Gaussian distribution. Is this process: a) wide sense stationary; b) second order stationary; c) strict stationary?

4. The covariance of a random function is described by: $C(h) = 10\exp(-h/30)$, and a constant mean. The multivariate pdf of any set of locations is the Gaussian distribution. Is this process: a) wide sense stationary; c) second order stationary; d) strict stationary?

5. Consider the following isotropic covariance function of a random function $Z(\mathbf{x})$:

$$C_z(h) = \begin{cases} 15\left\{1 - \left[\frac{3}{2}\left(\frac{h}{20}\right) - \frac{1}{2}\left(\frac{h}{20}\right)^3\right]\right\} & \text{if } h < 20 \\ 0 & \text{if } h \geq 20 \end{cases}$$

$$lag\ h \geq 0$$

The random function has a constant mean.

a. What type of stationarity can be assumed here?
b. What is the variance of the random function?
c. Calculate the values of the correlation function for lags $h$ = 2, 5, 10, 15.
d. Calculate the integral scale of the random function.

6. Consider a stationary random function $Z(t)$ whose spectral density is given by the following equation:

$$S_Z(\omega) = 20e^{-\omega/10}$$

What is the variance of the random function?

7.* Show that the integral scale of the exponential correlation function of Table 5.1 is equal to parameter $a$ and of the spherical correlation function is equal to (3/8)$a$.

8.* Consider a random function $Z(t)$ with a scale of fluctuation $\theta$=50 days, an exponential covariance function and $\sigma_Z^2 = 20$. Plot the relation between the variance of the averaged process $Z_T(t)$ with $T$ increasing from 1 to 100 days.

# 6. Time series analysis

## 6.1 Introduction

Many dynamic variables in hydrology are observed at more or less regular time intervals. Examples are rainfall, surface water stage and groundwater levels. Successive observations from a particular monitoring station observed at *regular intervals* are called *a time series*. In the context of stochastic hydrology we should look at a time series as a realization of a random function. In the terminology of Chapter 5 a time series can either be viewed as real-valued discrete-time random function (Figure 5.2a) or a real-valued continuous-time random function that has been observed at discrete times (Figure 5.6). Irrespective of this view of reality, separately from the theory of random functions, hydrologists have been using techniques (mostly coming from econometrics) specially designed to analyze and model hydrological time series. The most common of these techniques, collectively know as "time series analysis", will be treated in this chapter.

The main reasons for analyzing hydrological time series are:

1. Characterization. This includes not only the analysis of properties like average values and probability of exceeding threshold values, but also characteristics such as seasonal behavior and trend.
2. Prediction and forecasting. The aim of prediction and forecasting is to estimate the value of the time series at non-observed points in time. This can be a prediction at a time in future (forecasting), or a prediction at a non-observed point in time in the past, for example to fill in gaps in the observed series due to missing values.
3. Identify and quantify input-response relations. Many hydrological variables are the result of a number of natural and man-induced influences. To quantify the effect of an individual influence and to evaluate water management measures, the observed series is split into components which can be attributed to the most important influences.

The focus of this chapter is on time series models, expressing the value of the time series as a function of its past behavior and time series of influences factors (e.g. input variables that influence the hydrological variable that is analyzed). In this chapter we restrict ourselves to linear time series models as described by Box and Jenkins (1976). This means that the value of the variable under consideration is a linear function of its past and of the relevant influence factors. Extensive discussions on time series analysis can, amongst others, be found in books of Box and Jenkins (1976), Priestly (1989) and Hipel and McLeod (1996).

## 6.2 Definitions

Similar to the analysis and modeling of spatial random functions by geostatistics (Chapter 7), the time series analysis literature uses its own terminology which may be slightly different from the more formal definitions used in chapters 3 and 5. Consequently, some definitions of properties of random time series will be repeated first. Also the symbols used may be slightly different than used in the previous chapters, although we try to keep the notation as close a possible to that used elsewhere in the book. For instance, contrary to convention, stochastic variables representing noise are represented with lower case letters, while deterministic input variables are denoted as capitals.

### 6.2.1 Discrete stationary time series

As stated before, most hydrological variables, like river stages, are continuous is time. However, if we consider the variable $Z(t)$ at regular intervals in time $\Delta t$, we can define a discrete time series (see figure 6.1).

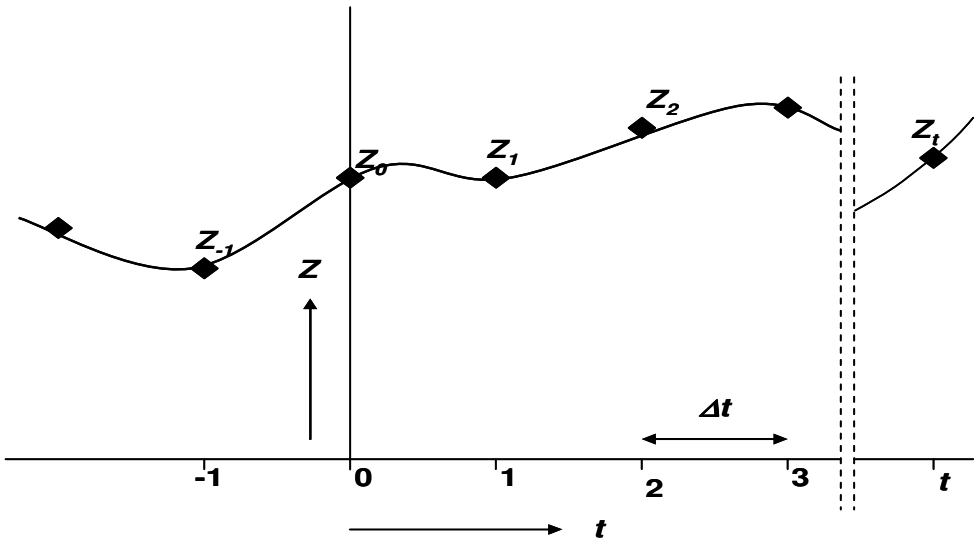



*Figure 6.1 Schematic of a continuous random function observed at discrete times* $Z(t)$

The values of the continuous time series at the regular times intervals are:

$$Z_k = Z(k\Delta t) \quad k = -\infty,...,-1,0,1,.....,\infty . \tag{6.1}$$

The series $Z_k$ is called a *discrete time series*. In the time series literature often the subscript $t$ is used instead of the subscript $k$.

$$Z_t \quad t = -\infty,...,-1,0,1,.....,\infty . \tag{6.2}$$

In the remainder of this chapter we will use the subscript *t*. Note that the subscript *t* is a rank number rather than a value of the running time.

### 6.2.2 Moments and Expectation

A single time series is considered to be a stochastic process that can be characterized by its (central) statistical moments. In particular the first and second order moments are relevant: the mean value, the variance and the autocorrelation function. For a statistical stationary process the mean value and the variance are:

$$\mu_z = E[Z_t] \tag{6.3}$$

$$\sigma_z^2 = VAR[Z_t] = E[\{Z_t - \mu_z\}\{Z_t - \mu_z\}] \tag{6.4}$$

The autocovariance is a measure of the relationship of the process at two points in time. For two points in time *k* time steps apart, the autocovariance is defined by:

$$COV[Z_t, Z_{t+k}] = E[\{Z_t - \mu_z\}\{Z_{t+k} - \mu_z\}] \quad k = -\infty,...,-1,0,1,...,\infty \tag{6.5}$$

Often *k* is called the time lag.

$$\text{i. note that for } k = 0 \quad COV[Z_t, Z_{t+k}] = \sigma_z^2 \tag{6.6a}$$

$$\text{ii.note that } COV[Z_t, Z_{t+k}] = COV[Z_{t+k}, Z_t] = COV[Z_t, Z_{t-k}] \tag{6.6b}$$

In time series analysis we often use the *autocorrelation function (ACF*), defined by:

$$\rho_{zz,k} = \frac{COV[Z_t, Z_{t+k}]}{COV[Z_t, Z_t]} = \frac{E[\{Z_t - \mu_z\}\{Z_{t+k} - \mu_z\}]}{\sigma_z^2} \quad k = -\infty,...,-1,0,1,...,\infty \tag{6.7}$$

It can be proven that the value of the ACF is always between 1 and -1. A value of 1 or -1 means a perfect correlation, while a value 0 indicates the absence of correlation. From the definition it follows that the ACF is maximum for *k*=0 ($\rho_{zz,0} = 1$). Just like for the autocovariance it follows from the definition that:

$$\rho_{zz,k} = \rho_{zz,-k} \quad k = -\infty,...,-1,0,1,...,\infty \tag{6.8}$$

The graphical representation of the ACF is called the autocorrellogram (see figure 6.2). Because the ACF is symmetrical around *k*=0, only the right (positive) side is shown.

The dynamic behavior of a time series is characterized by its variance and ACF. This is visualized in figure 6.2 for zero mean time series.

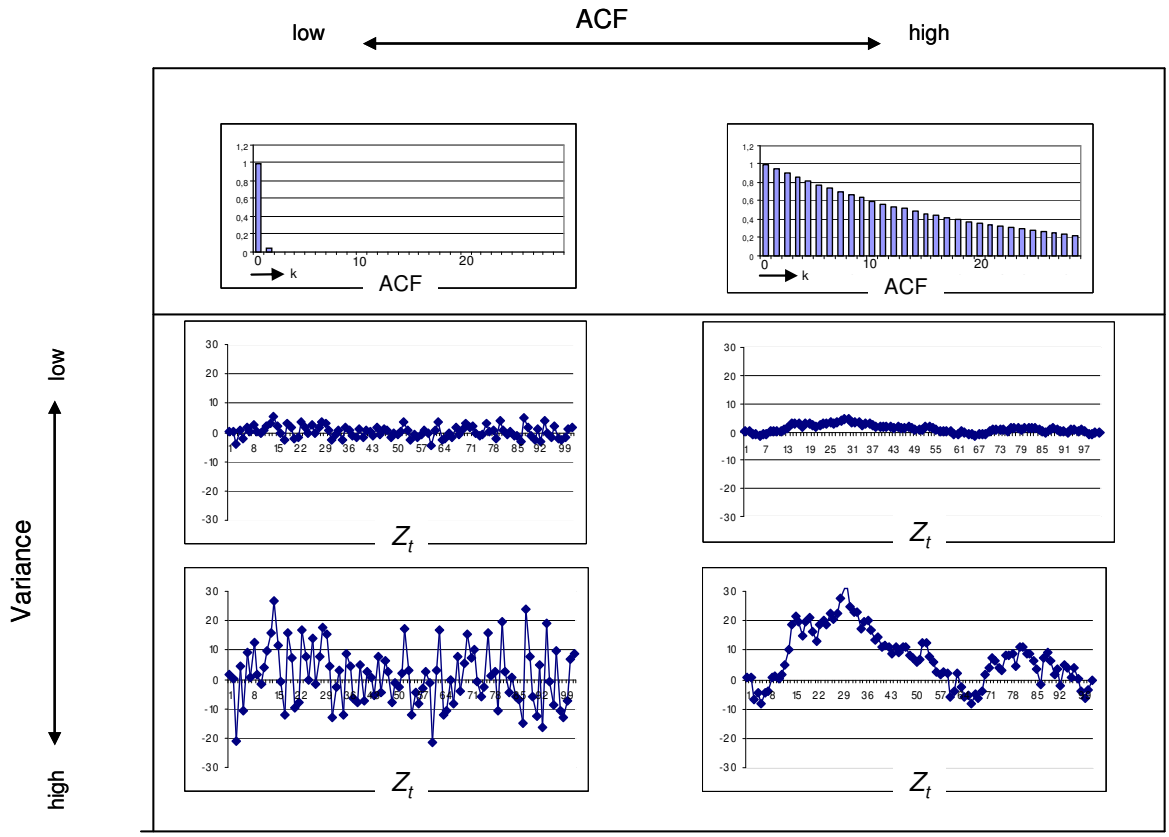



*Figure 6.2 Different characteristics of zero mean time series due to high and low values for the variance and the ACF.*

Analogous to the autocovariance and the ACF, that expresses the relation of a time series with itself, the relation between two different time series is expressed by the cross covariance and the *crosscorrelation* function (CCF). The cross covariance and CCF for the time series $Z_t$ and $X_t$ is defined as:

$$COV[X_t, Z_{t+k}] = E[\{X_t - \mu_x\}\{Z_{t+k} - \mu_z\}] \tag{6.9}$$

$$\rho_{XZ,k} = \frac{E[\{X_t - \mu_x\}\{Z_{t+k} - \mu_z\}]}{\sigma_x \sigma_z} \tag{6.10}$$

Analogous to the ACF it can be proven that the value of the CCF is always between 1 and -1. However the CCF is not symmetrical around $k=0$.

$$\rho_{XZ,k} \neq \rho_{XZ,-k} \tag{6.11}$$

From the definition it follows that

$$\rho_{XZ,k} = \rho_{ZX,-k} \tag{6.12}$$

In the real world we don't know the process $Z_t$ exactly and we have to estimate all information from observations. In absence of observation errors, the observation $z_t$ equals the value of the process $Z_t$ . Suppose we have observations of the process from time step 0 until time step $t$. Than the conditional expectation of the process at time step $\tau$ is denoted as:

$$E[Z_\tau \mid z_0,..., z_t] = \hat{Z}_{\tau|t} \tag{6.13}$$

We will use two operators: the backshift operator $B$ and the difference operator $\nabla$ .
The backshift operator is defined as:

$$B\,Z_t = Z_{t-1} \tag{6.14}$$

And therefore

$$B^n\,Z_t = B^{n-1}\,Z_{t-1} = Z_{t-n} \tag{6.15}$$

The difference operator is defined as:

$$\nabla\,Z_t = Z_t - Z_{t-1} \tag{6.16}$$

And

$$\nabla^n\,Z_t = \nabla^{n-1}\,(Z_t - Z_{t-1}) \tag{6.17}$$

### 6.2.3 Discrete white noise process

An important class of time series is the discrete white noise process $a_t$. This is a zero mean time series with a Gaussian probability distribution and no correlation in time.

$$\begin{aligned} &E[a_t] = 0 \\ &E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & if \quad k = 0 \\ 0 & if \quad k \neq 0 \end{cases} \end{aligned} \tag{6.18}$$

Because of the absence of correlation in time, the discrete white noise process at time step $t$ does not contain any information about the process at other time steps.

### 6.2.4 Rules of calculus

Calculation rules with expectations are summarized as:

$$\begin{aligned} &E[c]=c \\ &E[c\,Z]=c\,E[Z] \\ &E[Z+X]=E[Z]+E[X] \\ &E[ZX]=E[Z]E[X]+COV[Z,X] \end{aligned} \tag{6.19}$$

Where $X$ and $Z$ are discrete time series and c is a constant.

## 6.3 Principle of linear univariate time series models

The general concept of (linear) time series models is to capture as much information as possible in the model. This information is characterized by the mean value, the variance and the ACF. We consider the time series $Z_t$ as a linear function of a white noise process $a_t$ (see figure 6.3).

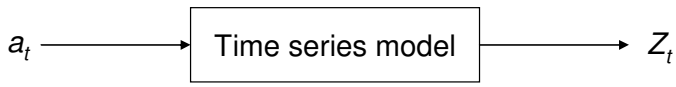



*Figure 6.3 Schematic representation of a time series model.*

Because the ACF($k$) of the white noise process equals zero for any $k \neq 0$, all information of the autocorrelation in $Z_t$ is captured in the time series model.

In the following we will describe the different types of random processes that underly the most commonly used time series models. For each process, we start with introducing the recurrent equation defining the process, followed by the relation between the process parameters and statistical properties of the process, showing how such processes can be predicted if observations are taken and ending with a numerical example.

## 6.4 Autoregressive processes

### 6.4.1 AR(1) process

*Definition*

The first class of linear time series processes discussed in this chapter is formed by the autoregressive (AR) processes. The most simple *autoregressive* process is the AR(1)

process (AutoRegressive process of order 1). A zero mean AR(1) process ($Z_t$ ) is defined as:

$$Z_t = \phi_1 Z_{t-1} + a_t \tag{6.20}$$

*Parameter Determination*

As stated before, the white noise process is a zero mean, uncorrelated process. Therefore the AR(1) process contain two unknowns:

- the first order auto regressive parameter $\phi_1$ and
- the variance of the white noise process $\sigma_a^2$

These unknowns have to be determined from the characteristics of the time series $Z_t$, in particular the variance and the ACF.

By multiplying both sides of equation 6.20 with $Z_{t-1}$ and taking the expectation we obtain

$$E[Z_{t-1} Z_t] = E[Z_{t-1} \phi_1 Z_{t-1} + Z_{t-1} a_t] = E[\phi_1 Z_{t-1}^2] + E[Z_{t-1} a_t] \tag{6.21}$$

The process $Z_t$ is independent of future values of the white noise process. Therefore, the value of the process $Z_{t-1}$ is independent of the white noise at time step $t$ and the second term at the right hand side of 6.21 equals zero. Dividing 6.21 by the variance $\sigma_Z^2$ yields:

$$\frac{E[Z_{t-1} Z_t]}{\sigma_Z^2} = \phi_1 \frac{E[Z_{t-1}^2]}{\sigma_Z^2} \tag{6.22}$$

And because $E[Z] = 0$ and $E[Z_{t-1}^2] = \sigma_Z^2$ the first order auto regressive parameter is:

$$\phi_1 = \rho_{ZZ,1} \tag{6.23}$$

The variance of the white noise process is determined by taking the expectations of the square of both sides of equation 6.20.

$$E[Z_t^2] = E[\{\phi_1 Z_{t-1} + a_t\}^2] = E[\phi_1^2 Z_{t-1}^2 + a_t^2 + 2\phi_1 Z_{t-1} a_t] = \phi_1^2 E[Z_{t-1}^2] + E[a_t^2] + 0 \tag{6.24}$$

From 6.24 it follows that:

$$\sigma_Z^2 = \phi_1^2 \sigma_Z^2 + \sigma_a^2 \quad \rightarrow \quad \sigma_Z^2 = \frac{\sigma_a^2}{1 - \phi_1^2} \tag{6.25}$$

*Properties of an AR(1) process*

Stationarity: Because the time series $Z_t$ is a temporally correlated process, the past values of the process contain information about the future behavior. In hydrology (and in many

other fields of application as well) if we go further into the future mostly the influence of past values eventually disappears. In other words, the process has a limited memory. To ensure this, the time series process should be *stationary*. In order for an AR(1) process to be stationary the absolute value of the model parameter should be smaller than 1.

$$|\phi_1| < 1 \tag{6.26}$$

Note that if the condition 6.26 is not fulfilled, from 6.25 it follows that the variance of the process $Z_t$ does not exist. In this case, the process $Z_t$ is said to be non-stationary.

ACF of an AR(1) process: By repetitive use of equation 6.20, $Z_t$ can be expressed as:

$$Z_t = \phi_1 Z_{t-1} + a_t = \phi_1(\phi_1 Z_{t-2} + a_{t-1}) + a_t = \phi_1^k Z_{t-k} + \sum_{i=1}^{k} \phi^{i-1} a_{t-i+1} \tag{6.27}$$

Multiplying both sides by $Z_{t-k}$ and taking the expectations yields:

$$E[Z_t Z_{t-k}] = E[\phi_1^k Z_{t-k}^2] + E[\sum_{i=1}^{k} \phi^{i-1} a_{t-i+1} Z_{t-k}] \tag{6.28}$$

Since $Z_{t-k}$ is independent of future values of the white noise process only the first term at the right hand side is non-zero. Dividing both sides by $E[Z_{t-k}^2] = \sigma_Z^2$ yields for the ACF:

$$\rho_{ZZ,k} = \frac{E[Z_t Z_{t-k}]}{E[Z_{t-k}^2]} = \frac{E[\phi_1^k Z_{t-k}^2]}{E[Z_{t-k}^2]} = \phi_1^k \tag{6.29}$$

From 6.29 it can be seen that the ACF of a stationary AR(1) process is an exponential function.

*Forecast of an AR(1) process*

Suppose we have observations up to time $t$ ( $z_i \quad i = -\infty, \cdots, t$ ). From 6.20 it follows that the forecast for $t+1$ is:

$$\hat{Z}_{t+1|t} = \phi_1 \hat{Z}_{t|t} + \hat{a}_{t+1|t} \tag{6.30}$$

Since

$$\hat{Z}_{t|t} = z_t \quad \text{and} \quad \hat{a}_{t+1|t} = 0 \tag{6.31}$$

It follows that

$$\hat{Z}_{t+1|t} = \phi_1 z_t \tag{6.32}$$

In general the forecast for time $t+\ell$ equals

$$\hat{Z}_{t+\ell|t} = \phi_1 \hat{Z}_{t+\ell-1|t} + \hat{a}_{t+\ell|t} = \phi_1^{\ell} \hat{Z}_{t|t} + \sum_{i=1}^{l} \phi_l^{i-1} \hat{a}_{t+\ell-i+1|t} = \phi_1^{\ell} z_t \tag{6.33}$$

The forecast error for time $t+\ell$ is:

$$e_{t+\ell|t} = Z_{t+\ell} - \hat{Z}_{t+\ell|t} = \phi_1^{\ell} z_t + \sum_{i=1}^{l} \phi_1^{i-1} a_{t+\ell-i+1} - \phi_1^{\ell} z_t = \sum_{i=1}^{l} \phi_1^{i-1} a_{t+\ell-i+1} \tag{6.34}$$

Taking the expectations of the square of both sides the variance of the forecast error for time $t+\ell$ is:

$$\begin{aligned} E[\{e_{t+\ell|t}\}^2] &= E\left[\left\{\sum_{i=1}^{l} \phi_1^{i-1} a_{t+\ell-i+1}\right\}^2\right] = \sum_{i=1}^{l} \phi_1^{2(i-1)} E[a_{t+\ell-i+1}^2] \\ \sigma_{e_{t+\ell|t}}^2 &= \sigma_a^2 \sum_{i=1}^{\ell} \phi_1^{2(i-1)} \end{aligned} \tag{6.35}$$

Note that the error variance for a forecast far in the future approaches the variance of the process $Z_t$.

$$\ell \to \infty; \quad \sum_{i=1}^{\ell} \phi_1^{2(i-1)} \to \frac{1}{1-\phi_1^2} \quad \text{and thus} \quad \sigma_{e_{t+\ell|t}}^2 \approx \frac{\sigma_a^2}{1-\phi_1^2} = \sigma_z^2 \tag{6.36}$$

*Example of an AR(1) process*

Let $Z_t$ be a zero mean AR(1) process, with $\phi_1 = 0.9$ and $\sigma_a^2 = 1$:

$$Z_t = 0.9 Z_{t-1} + a_t \tag{6.37}$$

From (6.29) it follows that the ACF(k) of the process $Z_t$ equals:

$$\rho_{zz,k} = 0.9^k \tag{6.38}$$

The autocorrellogram of the process $Z_t$ is given in figure 6.4.

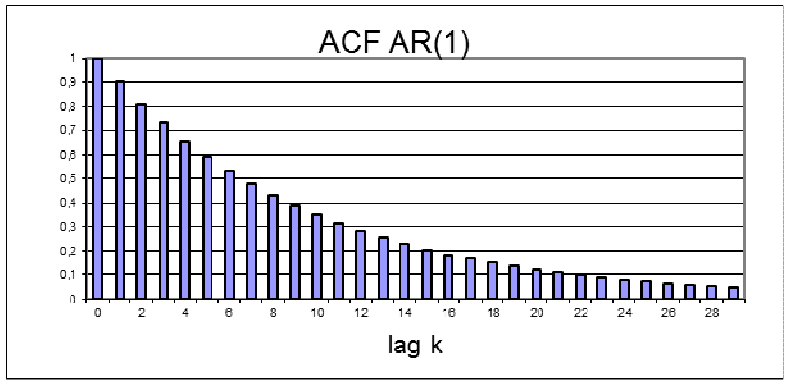



*Figure 6.4 Autocorrellogram of the AR(1) process*

The forecast of the AR(1) process and the forecast error variance are given by 6.33 and 6.35:

$$\hat{Z}_{t+\ell|t} = 0.9^{\ell} \quad \text{and} \quad \sigma^2_{e_{t+\ell|t}} = 1\sum_{i=1}^{\ell} 0.9^{2(i-1)} \tag{6.39}$$

The forecast and the corresponding 95% confidence interval ($\pm 1.96\,\sigma_{\hat{z}_{t+\ell|t}}$) are given in figure 6.5.

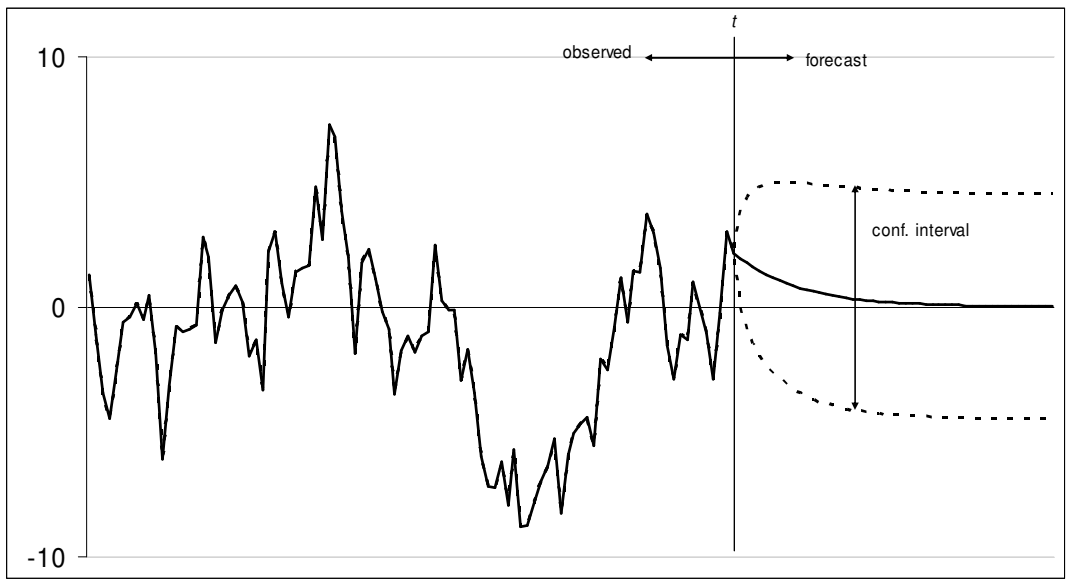



*Figure. 6.5 Forecast of an AR(1) process with corresponding 95% confidence interval.*

As can be seen from figure 6.5, the forecast of the AR(1) process gradually approaches zero for a forecast further into the future. The decay curve of the forecast reflects the memory of the AR(1) process. Consequently, the confidence interval gradually approaches its maximum value $\pm 1.96\,\sigma_Z$ for forecasts further in the future.

### 6.4.2 AR(p) process

*Definition*

The general form of a zero mean AutoRegressive process of order AR($p$) is defined as:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + a_t \tag{6.40}$$

The parameters $\phi_i$ ($i$=1,…,$p$) are called the auto regressive parameters of the order $i$.

*Parameter determination*

Similar to the parameter determination of an AR(1) process, the parameters of the AR(p) process can be expressed in terms of the variance and auto correlation of the process $Z_t$. Both sides of (6.40) are multiplied by $Z_{t-i}$ ($i$=1,…,$p$). This yields the set of equations:

$$\begin{aligned}
Z_{t-1}Z_t &= \phi_1 Z_{t-1}Z_{t-1} + \phi_2 Z_{t-1}Z_{t-2} + \cdots + \phi_p Z_{t-1}Z_{t-p} + Z_{t-1}a_t \\
Z_{t-2}Z_t &= \phi_1 Z_{t-2}Z_{t-1} + \phi_2 Z_{t-2}Z_{t-2} + \cdots + \phi_p Z_{t-2}Z_{t-p} + Z_{t-2}a_t \\
&\vdots \\
Z_{t-p}Z_t &= \phi_1 Z_{t-p}Z_{t-1} + \phi_2 Z_{t-p}Z_{t-2} + \cdots + \phi_p Z_{t-p}Z_{t-p} + Z_{t-p}a_t
\end{aligned} \tag{6.41}$$

Taking expectations and dividing both sides by the variance $\sigma_Z^2$ yields:

$$\begin{aligned}
\rho_{ZZ,1} &= \phi_1 + \phi_2 \rho_{ZZ,1} + \cdots + \phi_p \rho_{ZZ,p-1} \\
\rho_{ZZ,2} &= \phi_1 \rho_{ZZ,1} + \phi_2 + \cdots + \phi_p \rho_{ZZ,p-1} \\
&\vdots \\
\rho_{ZZ,p} &= \phi_1 \rho_{ZZ,p-1} + \phi_2 \rho_{ZZ,p-2} + \cdots + \phi_p
\end{aligned} \tag{6.42}$$

Writing this set of equations in matrix form yields:

$$\begin{bmatrix} 1 & \rho_{ZZ,1} & \cdots & \rho_{ZZ,p-1} \\ \rho_{ZZ,1} & 1 & \cdots & \rho_{ZZ,p-2} \\ \vdots & & \ddots & \vdots \\ \rho_{ZZ,p-1} & \rho_{ZZ,p-2} & \cdots & 1 \end{bmatrix} \bullet \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_{ZZ,1} \\ \rho_{ZZ,2} \\ \vdots \\ \rho_{ZZ,p} \end{bmatrix} \tag{6.43}$$

The AR-parameters can de solved from the set of equations (6.43) which is known as the *Yule-Walker Equations*. For example the parameters of an auto regressive model of order 2 are:

$$\phi_1 = \frac{\rho_{ZZ,1}(1-\rho_{ZZ,2})}{1-\rho_{ZZ,1}^2}$$
$$\phi_2 = \frac{\rho_{ZZ,2}-\rho_{ZZ,1}^2}{1-\rho_{ZZ,1}^2} \qquad (6.44)$$

Note that unlike the parameter in the AR(1) process, for the AR(2) process $\phi_1 \neq \rho_1$ .

*Properties of a AR(p) process*

Using the backshift operator $B$ (defined with $Z_{t-1} = BZ_t$), an AR(p) process can be written as:

$$Z_t = \phi_1 BZ_t + \phi_2 B^2 Z_t + \cdots + \phi_p B^p Z_t + a_t \qquad (6.45)$$

Defining:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p \qquad (6.46)$$

The general form of an AR(p) process is:

$$Z_t = \frac{1}{\Phi(B)} a_t$$

Stationarity Analogous to the AR(1) process, the values of the parameters are limited in order for the AR(p) process to be stationary. Without proof it is stated here that an AR(p) model is stationary if all (complex) roots of the equation:

$$\Phi(B) = 0 \qquad (6.47)$$

lie outside the unit circle.

For example, for an AR(2) process the roots of the function

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 = 0 \qquad (6.49)$$

must lie outside the unit circle. This implies that the parameters $\phi_1$ and $\phi_2$ must lie in the region

$$\phi_1 + \phi_2 < 1$$
$$\phi_2 - \phi_1 < 1 \qquad (6.50)$$
$$-1 < \phi_2 < 1$$

ACF of an AR(p) process. Multiplication of equation (6.40) by $Z_{t-k}$ , taking the expectations and dividing both sides by the variance $\sigma_Z^2$ yields:

$$\rho_{ZZ,k} = \phi_1 \rho_{ZZ,k-1} + \phi_2 \rho_{ZZ,k-2} + \cdots \phi_p \rho_{ZZ,k-p} \tag{6.51}$$

As an example, the autocorrellogram for the AR(2) process $Z_t = 0.35Z_{t-1} + 0.45Z_{t-2} + a_t$ is given in figure 6.6.

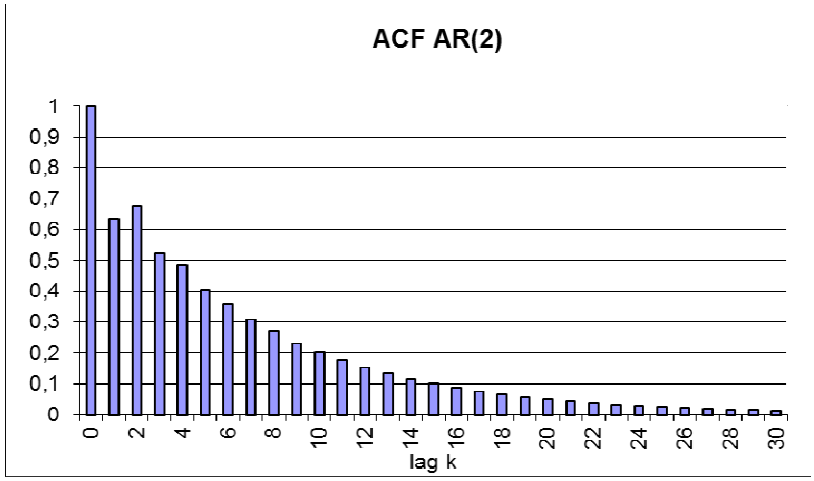



*Figure 6.6. Auto correllogram of an AR(2) process*

The acf(0) is 1 by definition. From the acf(1) and acf(2) in figure 6.6, the effect of the second order auto regressive parameter can be seen. Similar to the AR(1) model, the auto correlation for larger time lags decays gradually. This gradual decay is a general characteristic of AR-models.

*Forecast of an AR(p) process*

Forecasting an AR(p) process is similar to that of an AR(1) process. The AR(p) process at some point in time is dependent on the *p* time steps before. Therefore, the forecast of an AR(p) process does not show a single exponential pattern. In the first *p* time steps of the forecast the pattern of the last observations is reflected, but eventually the forecast will decay exponentially to zero.

As an example the AR(2) $Z_t = 0.35Z_{t-1} + 0.45Z_{t-2} + a_t$ process is given in figure 6.7. As show, the confidence interval tends to a constant value when the forecast approaches zero. Like in the case of an AR(1) process it can be proven that:

$$\ell \to \infty \quad \sigma^2_{e_{t+\ell|t}} \approx \sigma_Z^2 \tag{6.52}$$

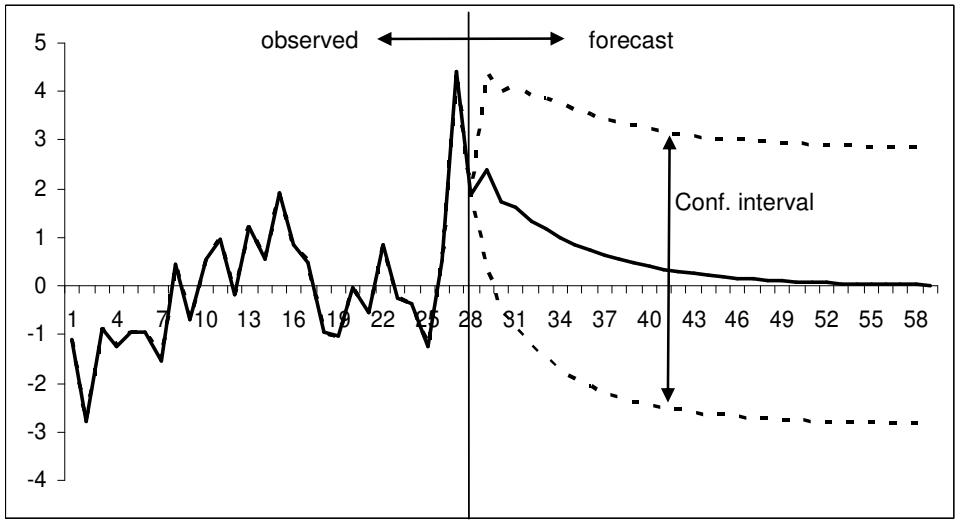



*Figure 6.7 Forecast of an AR(2) process with corresponding 95% confidence interval.*

## 6.5 Moving average processes

### 6.5.1 MA(1) process

*Definition*

The second class of time series processes consists of the Moving Average (MA) processes. The most simple in this class is the MA(1) process (Moving Average process of order 1). The zero mean MA(1) process is defined as:

$$Z_t = a_t - \theta_1 a_{t-1} \tag{6.53}$$

*Parameter determination*

Like in the case of the AR-processes, from the statistical properties of the process $Z_t$, we have to determine:

- the moving average parameter $\theta_1$ and
- the variance of the white noise process $\sigma_a^2$

Taking the expectation of the square of both sides of (6.53) yields:

$$E[Z_t Z_t] = E[(a_t - \theta_1 a_{t-1})(a_t - \theta_1 a_{t-1})] \quad \rightarrow \quad \sigma_Z^2 = (1+\theta_1^2)\sigma_a^2 \tag{6.54}$$

Multiplication of 6.53 by $Z_{t-1}$, taking the expectation and dividing both sides by $\sigma_Z^2$ yields:

$$E[Z_t Z_{t-1}] = E[(a_t - \theta_1\, a_{t-1})Z_{t-1}] = E[(a_t - \theta_1\, a_{t-1})(a_{t-1} - \theta_1\, a_{t-2})]$$
$$= E[a_t a_{t-1}] - \theta_1 E[a_{t-1}a_{t-1}] - \theta_1 E[a_t a_{t-2}] + \theta_1^2 E[a_{t-1}a_{t-2}] = 0 - \theta_1\sigma_a^2 - 0 + 0$$
$$E[Z_t Z_{t-1}] = -\theta_1\sigma_a^2 \tag{6.55}$$
$$\rho_{zz,1} = \frac{-\theta_1\sigma_a^2}{\sigma_Z^2}$$

Combining (6.54) and (6.55) we obtain the equation:

$$\rho_{zz,1} = \frac{-\theta_1\,\sigma_a^2}{(1+\theta_1^2)\sigma_a^2} = \frac{-\theta_1}{(1+\theta_1^2)} \tag{6.56}$$

Mathematically, equation (6.56) has two solutions for $\theta_1$ and the question arises which one of the two solutions is the proper one to describe the MA(1) process. The selection of the proper value of $\theta_1$ is based on the invertibility criterion. This criterion means that we like to be able to reconstruct the white noise series $a_t$ uniquely from the realization of the process $Z_t$. By repetitive use of the definition of the MA(1) process (6.53) it follows that:

$$\begin{aligned}
Z_t &= a_t - \theta_1\, a_{t-1} \\
Z_t &= a_t - \theta_1(Z_{t-1} + \theta_1 a_{t-2}) = a_t - \theta_1 Z_{t-1} - \theta_1^2 a_{t-2} \\
&\vdots \\
Z_t &= a_t - \theta_1 Z_{t-1} - \theta_1^2 Z_{t-2} - \cdots - \theta_1^k Z_{t-k} - \theta_1^{k+1} a_{t-k-1} = a_t - \sum_{i=1}^{k} \theta_1^i Z_{t-i} - \theta_1^{k+1} a_{t-k-1}
\end{aligned} \tag{6.57}$$

And therefore:

$$a_t = Z_t + \sum_{i=1}^{k} \theta_1^i Z_{t-i} + \theta_1^{k+1} a_{t-k-1} \tag{6.58}$$

If we require the MA- process $Z_t$ to be invertible the term $\sum_{i=1}^{k} \theta_1^i Z_{t-i}$ should approach zero for $k \to \infty$. This requirement leads to the condition:

$$|\theta_1| < 1 \tag{6.59}$$

It can be proven that (6.56) always has one root whose absolute value is larger than 1 and one root whose absolute value is smaller than 1. Therefore with the condition (6.59) the MA parameter $\theta_1$ can be determined uniquely. The variance of the white noise process follows from (6.54):

$$\sigma_a^2 = \frac{\sigma_Z^2}{1+\theta_1^2}. \tag{6.60}$$

*Properties of a MA(1) process*

Stationarity. From the definition (6.53) it follows that the MA(1) process is always stationary. Note that the AR(p) process is always invertible. This can be seen by subtracting two successive values of the process $Z_t$.

ACF of the MA(1) process. The lag $k$ auto correlation coefficient of the MA(1) process $Z_t$ is defined by:

$$\begin{aligned} \frac{E[Z_t Z_{t-k}]}{\sigma_Z^2} &= \frac{E[(a_t - \theta_1 a_{t-1})(a_{t-k} - \theta_1 a_{t-k-1})}{\sigma_Z^2} \\ \rho_{ZZ,k} &= \frac{E[a_t a_{t-k}] - \theta_1 E[a_{t-1} a_{t-k}] - \theta_1 E[a_t a_{t-k-1}] + \theta_1^2 E[a_{t-1} a_{t-k-1}]}{(1+\theta_1^2)\sigma_a^2} \end{aligned} \tag{6.61}$$

Because $E[a_t a_{t-k}] = 0$ for $k \neq 0$ it follows from (6.61) that:

$$\begin{aligned} \rho_{ZZ,0} &= 1 \\ \rho_{ZZ,1} &= \frac{-\theta_1}{1+\theta_1^2} \\ \rho_{ZZ,k} &= 0 \qquad |k| > 1 \end{aligned} \tag{6.62}$$

*Forecast of the MA(1) process*

From the definition of the MA(1) process (6.53) it follows that the forecast for time step $t$+1 given observations up to time $t$, is:

$$\hat{Z}_{t+1|t} = \hat{a}_{t+1|t} - \theta_1 \hat{a}_{t|t} \tag{6.63}$$

Because the conditional expectation of the white noise process at time step $t$+1 given the observations until time step $t$ is zero, the forecast is:

$$\hat{Z}_{t+1|t} = -\theta_1 \hat{a}_{t|t} \tag{6.64}$$

For $k \to \infty$ then $\theta_1^{k+1} a_{t-k-1} \to 0$. Therefore, with (6.58), the conditional expectation of the white noise process at time step $t$, given the observations until time step $t$ can be derived as:

$$\hat{a}_{t|t} = E[a_t \mid Z_0, \cdots, Z_t] = E[(Z_t + \sum_{i=1}^{k} \theta_1^i Z_{t-i}) \mid Z_0, \cdots, Z_t] = Z_t + \sum_{i=1}^{k} \theta_1^i Z_{t-i} \approx a_t \tag{6.65}$$

Therefore the one time step ahead forecast is:

$$\hat{Z}_{t+1|t} = -\theta_1 a_t \tag{6.66}$$

The conditional expectation for the white noise process at time step $t+\ell$ ; $\ell>1$ given the observations up to time step *t* is zero. Therefore the forecast of the MA(1) process is given by:

$$\hat{Z}_{t+\ell|t}=\hat{a}_{t+\ell|t}-\theta_1\hat{a}_{t+\ell-1|t}=\begin{cases}-\theta_1 a_t & \text{for} \quad \ell=1\\ 0 & \text{for} \quad \ell\neq 1\end{cases} \tag{6.67}$$

The error in the forecast is:

$$\begin{aligned} e_{t+\ell|t}&=Z_{t+\ell}-\hat{Z}_{t+\ell|t}=(a_{t+\ell}-\theta_1 a_{t+\ell-1})-\hat{Z}_{t+\ell|t}\\ e_{t+\ell|t}&=\begin{cases}a_{t+1} & \ell=1\\ a_{t+\ell}-\theta_1 a_{t+\ell-1} & \ell>1\end{cases}\end{aligned} \tag{6.68}$$

And the error variance of the forecast of the MA(1) process is:

$$\sigma^2_{e_{t+\ell|t}}=\begin{cases}\sigma_a^2 & \ell=1\\ (1+\theta_1)\sigma_a^2=\sigma_Z^2 & \ell>1\end{cases} \tag{6.69}$$

From (6.68) and (6.69) it can be seen that the memory of the MA(1) process is only 1 time step. Forecasts for time steps larger than 1 are all zero. The variance of the forecast error is equal to the variance of the process $Z_t$.

*Example of a MA(1) process*

Let $Z_t$ be a zero mean MA(1) process with $\theta_1=-0.9$ and $\sigma_a^2=1$.

$$Z_t=a_t+0.9a_{t-1} \tag{6.70}$$

From (6.62) is follows that the ACF equals:

$$\begin{aligned}&\rho_{zz,0}=1\\ &\rho_{zz,1}=0.497\\ &\rho_{zz,k}=0 \qquad |k|>1\end{aligned} \tag{6.71}$$

The auto correllogram is given in figure 6.8.

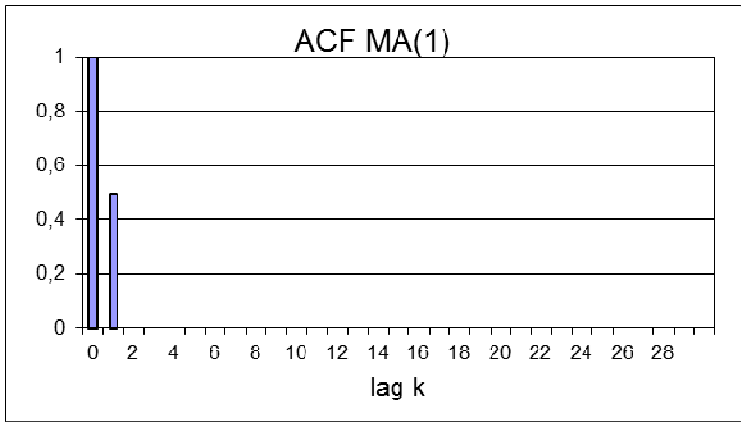



*Figure 6.8. ACF of the MA(1) process with* $\theta_1 = -0.9$.

The forecast of the MA(1) process and the corresponding error variance is calculated using (6.67) and (6.69). In figure 6.9 the forecast of the MA(1) process with $\theta_1 = -0.9$ is plotted with the 95% confidence interval.

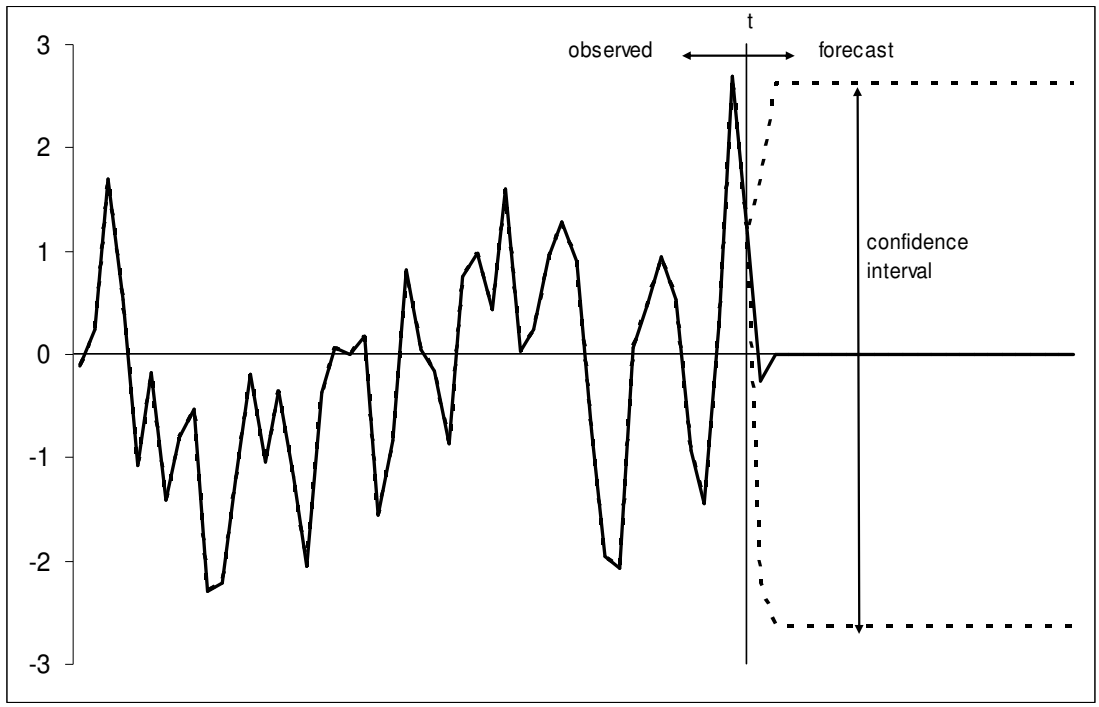



*Figure 6.9. Forecast of the MA(1) process with* $\theta_1 = -0.9$ *with the corresponding 95% confidence interval.*

Figure 6.9 shows that the memory of the MA(1) process is only one time step. Forecasts for more than one time step are all zero, which equals the unconditional expected value of the process. Consequently, the confidence interval reaches its maximum value after one time step.

### 6.5.2 MA(q) process

*Definition*

The general form of a Moving Average processes of order $q$ (MA(q) processes) is defined as:

$$Z_t = a_t - \theta_1 a_{t-1} \cdots - \theta_q a_{t-q} \tag{6.72}$$

The parameters $\theta_i$ ; $i = 1, \cdots, q$ are called the moving average parameters of order *i*. Using the backshift operator (6.72) can be written as:

$$Z_t = \Theta(B) a_t \tag{6.73}$$

Where: $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

*Parameter determination*

To determine the parameters of the MA(q) model, equation (6.72) is multiplied by $Z_{t-k.}$ Taking the expectation of both sides thereafter gives:

$$E[Z_t Z_{t-k}] = E[(a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q})(a_{t-k} - \theta_1 a_{t-k-1} - \cdots - \theta_q a_{t-k-q})] \tag{6.74}$$

For *k=0* equation (6.74) yields:

$$\begin{aligned} &E[Z_t Z_t] = \sigma_Z^2 = E[(a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q})^2] \\ &\sigma_Z^2 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_a^2 \end{aligned} \tag{6.75}$$

For *k*>0 equation (6.74) gives:

$$E[Z_t Z_{t-k}] = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)\sigma_a^2 & \text{for } k = 1,2,\ldots,q \\ 0 & \text{for } k > q \end{cases} \tag{6.76}$$

Combining (6.75) and (6.76) yields the set of equations:

$$\rho_{ZZ,k} = \begin{cases} \dfrac{(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)}{(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)} & \text{for } k = 1,2,\ldots,q \\ 0 & \text{for } k > q \end{cases} \tag{6.77}$$

The set (6.77) gives $q$ equations with $q$ unknown Moving Average parameters. Because the equations have a quadratic form, there are multiple solutions. Similar to the

determination of the parameter of the MA(1) process, the parameters of the MA($q$) process can be determined using the invertibility criterion. Without proof, it is stated here that for a MA($q$) process to be invertible, the roots of the equation

$$\Theta(B) = 0 \tag{6.78}$$

should all be outside the unit circle, and there is only one solution that obeys the invertibility criterion. For example the parameters of a MA(2) process are determined by the set of equations:

$$\begin{aligned} \rho_{ZZ,1} &= \frac{-\theta_1 + \theta_1\theta_2}{(1+\theta_1^2+\theta_2^2)} \\ \rho_{ZZ,2} &= \frac{-\theta_2}{(1+\theta_1^2+\theta_2^2)} \end{aligned} \tag{6.79}$$

The parameters can be found using the condition that the roots of the equation (6.78) should lie outside the unit circle. For the parameters of a MA(2) process this condition means:

$$1-\theta_1 B-\theta_2 B^2 = 0 \quad \rightarrow \quad \begin{cases} \theta_1+\theta_2 < 1 \\ \theta_1-\theta_2 < 1 \\ -1 < \theta_2 < 1 \end{cases} \tag{6.80}$$

The variance of the white noise process $a_t$ can be derived from (6.75).

$$\sigma_a^2 = \frac{\sigma_Z^2}{1+\theta_1^2+\cdots+\theta_q^2} \tag{6.81}$$

*Properties of a MA(q) process*

Stationarity. Like in the case of a MA(1) the stationarity of a MA(q) is always assured.

ACF of a MA(q) process. The ACF of a MA($q$) process follows from (6.77). As can be seen the MA($q$) process has a memory of $q$ time steps. After $q$ time steps, the ACF is cut off and all values >$q$ are zero. The ACF of the MA(2) process $Z_t = a_t + 0.4a_{t-1} + 0.5a_{t-2}$ is given in figure 6.10.

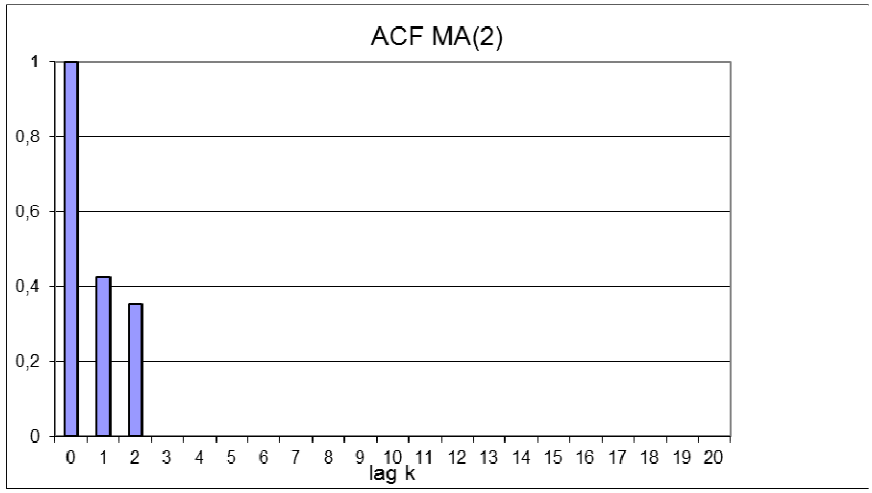



*Figure 6.10. ACF of the MA(2) process* $Z_t = a_t + 0.4a_{t-1} + 0.5a_{t-2}$.

*Forecast of an MA(q) process*

The forecast of a MA($q$) process is a straight forward extension of the forecast of a MA(1) model. Up to $q$ time steps ahead the forecast has a non-zero value. Further than $q$ time steps the forecast equals the mean value of the process, which is zero. The variance of the forecast error after $q$ time steps equals the variance of the process $Z_t$. Consequently, the 95% confidence interval has a constant value beyond the forecast of $q$ time steps. In figure 6.11 the forecast and the corresponding 95% confidence interval is given for the MA(2) process $Z_t = a_t + 0.4a_{t-1} + 0.5a_{t-2}$. As can be seen from this figure, after two time steps the forecast is zero and the confidence interval has a constant value.

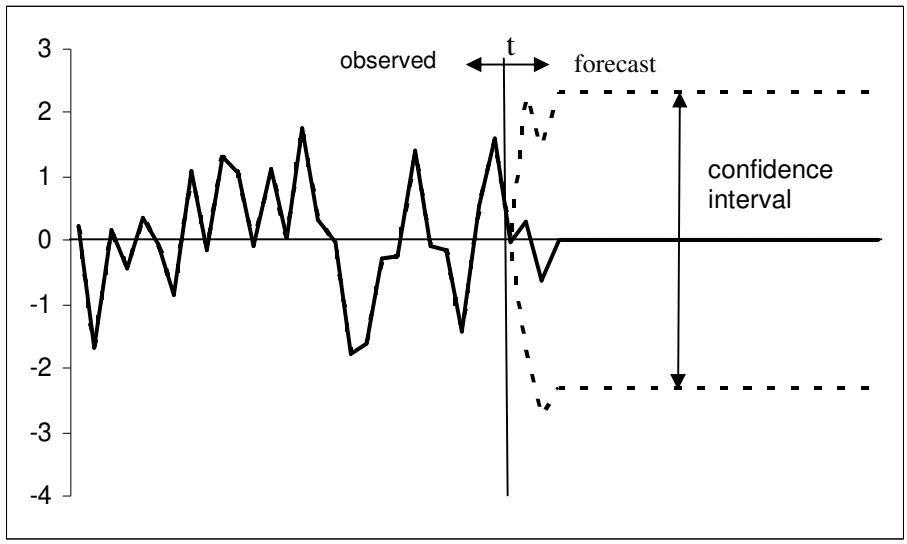



*Figure 6.11. Forecast of the MA(2) process* $Z_t = a_t + 0.4a_{t-1} + 0.5a_{t-2}$ *with the corresponding 95% confidence interval.*

## 6.6 Autoregressive Integrated Moving Average processes

### 6.6.1 Autoregressive Moving Average processes

*Definition*

We have seen that the typical characteristic of an AR($p$) process is the gradual decay of the temporal correlation. If at some point in time the value of an AR($p$) process is high, it will stay high for some time. A measured value at time step $t$ carries information for many time steps into the future. The amount of information gradually decreases when we go further into the future. In contrast, MA($q$) processes have a memory of a limited number of time steps. A measurement at time step $t$ carries information of $Z_t$ $q$ time steps into future. The temporal correlation of a MA($q$) process drops to zero after $q$ time steps. Many processes in hydrology show characteristics of both processes. At small time lags we see the random shock characteristics that look like a Moving Average process, and for larger time lags the process exhibits a gradual decay. This class of processes combining both characteristics is the Autoregressive Moving Average processes (ARMA($p,q$)), which is defined as:

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{p-1} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{q-1} \tag{6.82}$$

or using the back shift operator $B$:

$$Z_t = \frac{1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p} = \frac{\Theta(B)}{\Phi(B)} a_t \tag{6.83}$$

*Parameter determination and properties*

In an ARMA(p,q) process there are $p$+$q$+1 parameters which have to be determined. These are the $p$ auto regressive parameters ($\phi_p$), $q$ moving average parameters ($\theta_q$) and the variance of the white noise process ($\sigma_a^2$). The determination of the parameters of an ARMA process is similar to the determination of the parameters of AR and MA processes by multiplying (6.82) with respectively $Z_{t-k}$ and $a_{t-k}$ and take the expectations of both sides. Although straight forward, the equations are more complex than in the case of AR($p$) or MA($q$) processes. Therefore we will not give the general equations for parameter determination of an ARMA(p,q) process. As an illustration the determination and the correllogram are given for an ARMA(1,1) process in the next section.

In order for an ARMA($p,q$) process to be stationary and invertible the same conditions hold as in case of an AR($p$) process and a MA($q$) process. To assure stationarity all the roots of the function $\Phi(B) = 0$ should lie outside the unit circle and for invertibility the roots of the function $\Theta(B) = 0$ should lie outside the unit circle.

The ACF of an ARMA process shows elements of both the AR process and the MA process. There might be some spikes in the first $q$ time lags and for larger time lags the ACF exponentially decays to zero. As an example the auto correllogram of the ARMA(1,1) process is given below.

*Example of an ARMA(1,1) process*

Consider the ARMA(1,1) process:

$$Z_t = 0.7Z_{t-1} + a_t + 0.5a_{t-1} \tag{6.84}$$

It can be derived that the ACF of an ARMA(1,1) process is given by:

$$\begin{aligned} &\rho_{ZZ,0} = 1 \\ &\rho_{ZZ,1} = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} \\ &\rho_{ZZ,k} = \phi_1\rho_{ZZ,k-1} \quad for \quad k \geq 2 \end{aligned} \tag{6.85}$$

The auto correllogram is given in Figure 6.12.

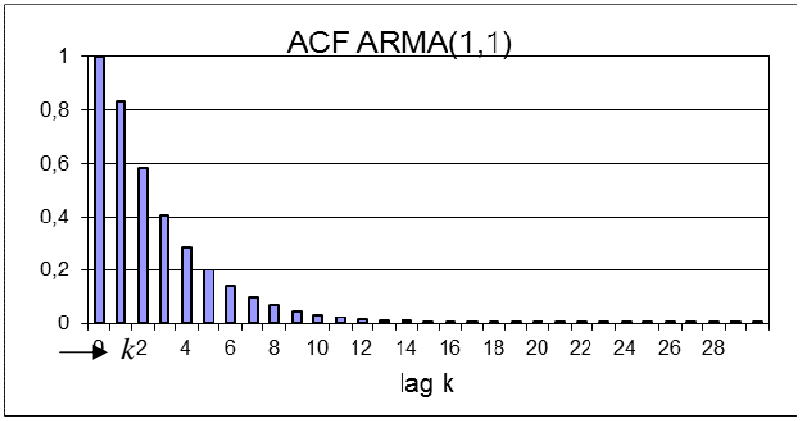



*Figure 6.12. Auto correllogram of the ARMA(1,1) process* $Z_t = 0.7Z_{t-1} + a_t + 0.5a_{t-1}$.

In Figure 6.12 it can be seen that for lags higher than 2, the ACF shows an exponential decay, like an AR process. The ACF lag 1 is higher than it would be in case of an AR(1) process. This reflects the first moving average term.
An illustration of a forecast of the ARMA(1,1) process $Z_t = 0.7Z_{t-1} + a_t + 0.5a_{t-1}$ is given in figure 6.13. The forecast shows characteristics of an AR process as well as a MA process. In the forecast one time step ahead, we see the influence of the MA term. For larger time steps the forecast decays gradually to zero, just like an AR process.

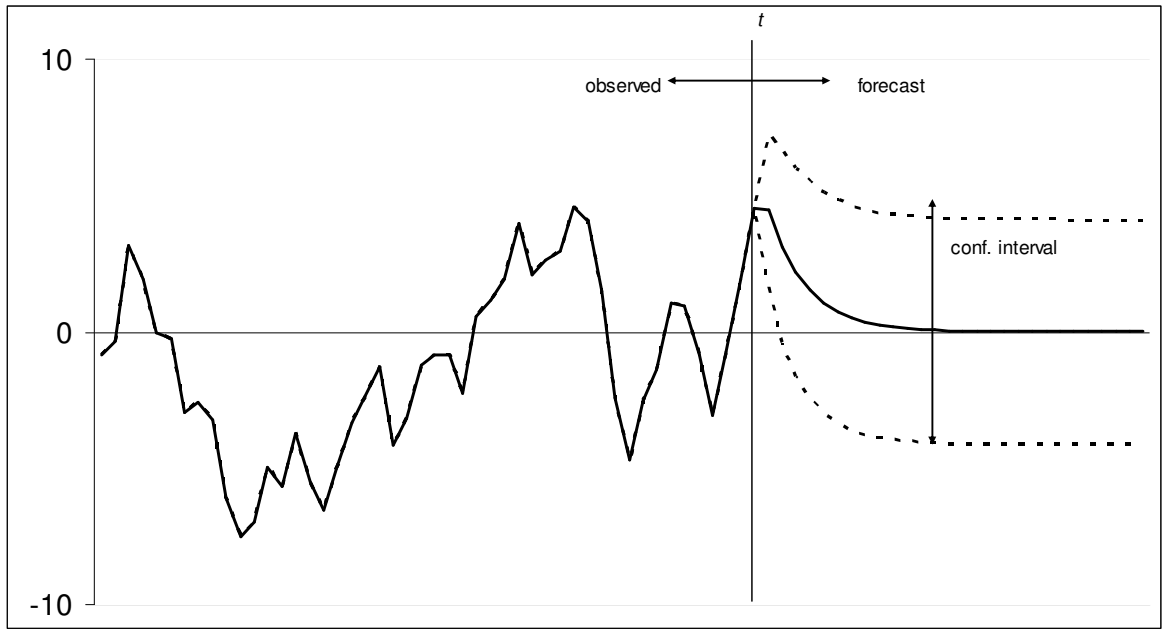



Fig. 6.13. Forecast of the ARMA(1,1) process $Z_t = 0.7Z_{t-1} + a_t + 0.5a_{t-1}$.

### 6.6.2 Non-zero mean ARMA processes and non-Gaussian time series

The ARMA(p,q) process can be extended to non-zero processes. Let the expected value of the process $Z_t$ be $\mu_Z$. The non-zero-mean ARMA process is now defined as:

$$(Z_t - \mu_Z) = \frac{\Theta(B)}{\Phi(B)} a_t \tag{6.86}$$

A non-zero mean ARMA process has the same characteristics as a zero-mean ARMA process except of a shift in the level.

We assumed the processes $Z_t$ and $a_t$ to be Gaussian processes. If the series $Z_t$ is non-Gaussian, we can try to transform the series in order to get a Gaussian process. Widely use are the so called Box-Cox transformations (Box and Cox, 1960).

### 6.6.3 Autoregressive Integrated Moving Average processes

Some processes have a time dependent expected value. For example if the time series shows a trend. In this case we speak of a non-stationary process. However, the difference between two successive values of the time series might be described by a stationary process. In general the $d^{th}$ order difference of a non-stationary time series might be described as a stationary ARMA process. To arrive at the original series, the stationary

ARMA process of the differences has to be integrated. The class of time series processes that describe this type of non-stationary behavior is formed by the Autoregressive Integrated Moving Average processes ARIMA(*p,d,q*), where *d* stands for the order of difference operations. Using the difference operator, the general form of an ARIMA(*p,d,q*) model is:

$$\nabla^d Z_t = -\frac{\Theta(B)}{\Phi(B)} a_t \tag{6.87}$$

Note that when applying a difference operation the non-zero expected value of the process disappears.

For example the ARIMA(1,1,1) model can be expanded to:

$$\begin{aligned} &\nabla^1 Z_t = Z_t - Z_{t-1} = \frac{1-\theta_1 B}{1-\phi_1 B} a_t \\ &(1-\phi_1 B)(Z_t - Z_{t-1}) = (1-\theta_1) a_t \\ &Z_t = (1+\phi_1) Z_{t-1} - \phi_1 Z_{t-2} + a_t - \theta_1 a_t \end{aligned} \tag{6.88}$$

And the ARIMA(0,2,1) process is:

$$\begin{aligned} &\nabla^2 Z_t = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2}) = (1-\theta_1) a_t \\ &Z_t = 2Z_{t-1} - Z_{t-2} + a_t - \theta_1 a_t \end{aligned} \tag{6.89}$$

The ARIMA process can be useful to describe the past behavior of non-stationary processes, but we should be careful forecasting a non-stationary time series, because the variance of the forecast error has no limited value for forecasts further in future. Consider the ARIMA(1,1,0) process:

$$Z_t - Z_{t-1} = \phi_1 (Z_{t-1} - Z_{t-2}) + a_t \tag{6.90}$$

Similar to (6.34) it follows that

$$\begin{aligned} &e_{t+\ell|t} - e_{t+\ell-1|t} = \sum_{i=1}^{l} \phi_1^{i-1} a_{t+\ell-i+1} \\ &e_{t+\ell|t} = e_{t+\ell-1|t} + \sum_{i=1}^{l} \phi_1^{i-1} a_{t+\ell-i+1} \end{aligned} \tag{6.91}$$

And the variance of the forecast error is:

$$\sigma^2_{e_{t+\ell|t}} = \sigma^2_{e_{t+\ell-1|t}} + \sigma_a^2 \sum_{i=1}^{\ell} \phi_1^{2(i-1)} \tag{6.92}$$

The variance of the forecast error grows unlimited unless the variance of the white noise equals zero. As an example the forecast of the ARIMA(1,1,0) process $Z_t - Z_{t-1} = 0.7(Z_{t-1} - Z_{t-2}) + a_t$ is shown in figure 6.14. The variance of the white noise equals 1. As can be seen in Figure 6.14 the forecast further into the future tends to a constant value but the confidence interval increases continuously.

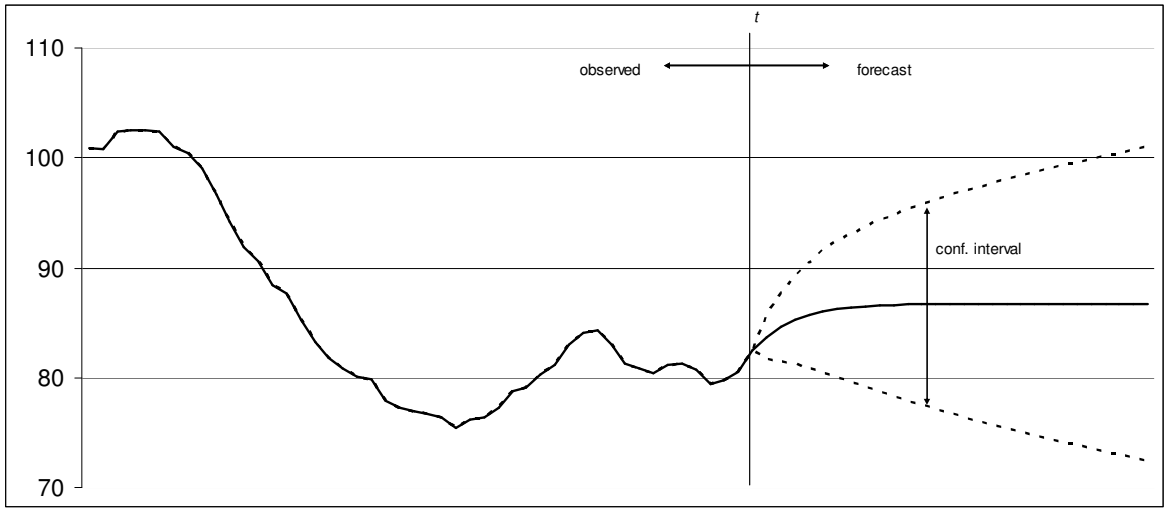



*Figure 6.14. Forecast of the ARIMA(1,1,0) process* $Z_t - Z_{t-1} = 0.7(Z_{t-1} - Z_{t-2}) + a_t$ .

### 6.6.4 Seasonal ARIMA processes

Many time series in hydrology, like air temperature, rainfall and river discharge, show a seasonal behavior. In order to describe seasonal (or periodic) behavior, the Seasonal Autoregressive Integrated Moving Average (SARIMA) process is introduced. The seasonal process has exactly the same form as the ARIMA process discussed in the previous section. However, the time steps of the terms of the seasonal process are related to the seasonal period instead of the regular time. The seasonal ARIMA process is denoted as: SARIMA(P,D,Q)$_s$ , where P is the number of seasonal Autoregressive terms, D is de number of seasonal differences, Q is the number of seasonal Moving Average terms and s is the seasonal period. The general form of a SARIMA(P,D,Q)$_s$ process is:

$$\nabla_s Z_t = \frac{\Theta_s(B^s)}{\Phi_s(B^s)} a_t \tag{6.93}$$

For example a time series with monthly values and a periodic behavior with a period of one year, might be described by a SARIMA(1,1,1)$_{12}$ process:

$$\nabla_{12} Z_t = \frac{(1-\theta_{12})B^{12}}{(1-\phi_{12})B^{12}} a_t \qquad (6.94)$$
$$Z_t = (1+\phi_{12})Z_{t-12} - \phi_{12} Z_{t-24} + a_t - \theta_{12} a_{t-12}$$

In many cases, the time series doesn't show a pure seasonal behavior, as given in (6.93), but also exhibits a dependency on values of previous time steps. To account for the seasonal behavior as well as for the behavior in regular time, both descriptions can be combined by multiplication. Such a process is denoted as: SARIMA(p,d,q)x(P,D,P)$_s$. The symbols p,d,q,P,D,Q,s have the same meaning as before. The general form of a SARIMA(p,d,q)x(P,D,Q)$_s$ process is:

$$\nabla \nabla_s = \frac{\Theta(B)}{\Phi(B)} \frac{\Theta_s(B^s)}{\Phi_s(B^s)} a_t \qquad (6.95)$$

Obviously the general form (6.95) can describe a broad class of processes, depending on the orders of the regular and seasonal differences and the auto regressive and moving average terms. As an example here the SARIMA(1,0,0)x(0,1,1)$_{12}$ process with $\phi_1 = 0.8$, $\theta_{12} = 0.6$ and $\sigma_a^2 = 1$ is given:

$$\nabla_{12} Z_t = \left( \frac{1}{1-0.8B} \right) \left( \frac{1-0.6B^{12}}{1} \right) a_t \qquad (6.96)$$
$$Z_t = 0.8 Z_{t-1} + Z_{t-12} - 0.8 Z_{t-13} + a_t - 0.6 a_{t-12}$$

A graph with observed and forecasted values of the process is given in figure 6.15. This figure shows that the time series is dominated by a periodic behavior with a period of 12 month. In addition to the periodic function, variations can be seen that are due to the regular part of the process. In the forecasts, the information of the regular part dampes out after some time steps and the forecast further into the future only reflects the seasonal behavior of the process.

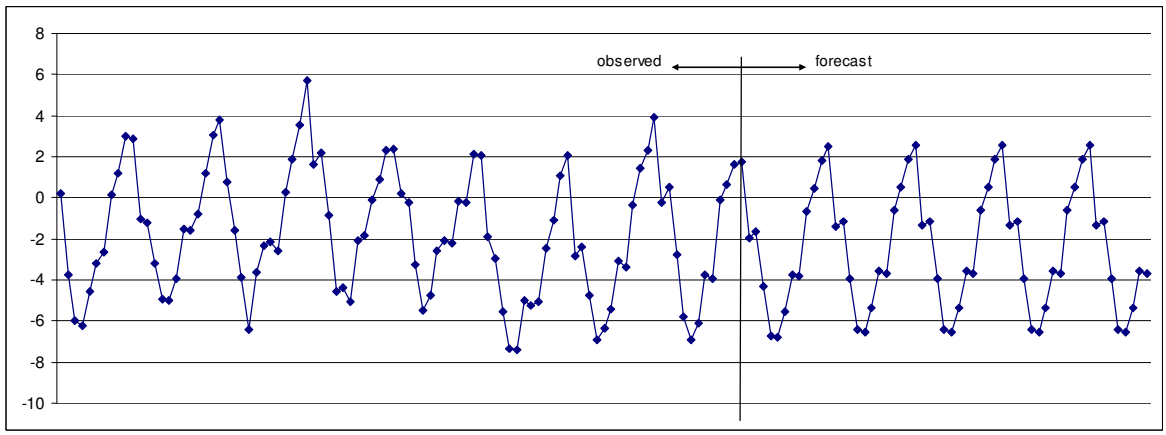



*Fig. 6.15. Forecast of the SARIMA(1,0,0)x(0,1,1)$_{12}$ process* $Z_t = 0.8 Z_{t-1} + Z_{t-12} - 0.8 Z_{t-13} + a_t - 0.6 a_{t-12}$

## 6.7 Modelling aspects

### 6.7.1 General

In practice, we have only observations of a time series over a limited period in time. The real process is unknown, and we don't know whether the series is a realization of a linear process, let alone what type of process AR(1), ARMA(1,2),..,we are dealing with. Nevertheless, for many processes in hydrology, we can fit a linear time series model, that describes the major behavior of the process. The general formulation of equation (6.95) allows taking a large number of possible linear time series models into account. The question arises how we can determine the 'best' time series model. In general, the modeling process consists of three stages.

1. In de identification stage we try to characterize the time series and we select possible time series models.
2. In the estimation stage, the parameters of the possible models are estimated.
3. In de diagnostic (or verification) stage, it is evaluated whether a time series models is valid, and which of the valid time series models is chosen to be the 'best' to describe the time series. Based on the diagnostics other model structures might be considered and stages 2 and 3 repeated.

In almost all computer programs for time series modeling, these three stages can be recognized. Here only a brief introduction to the three stages is given. An extensive discussion can for example be found in Box and Jenkins (1976).

### 6.7.2 Identification

The goal of the identification is to select the most likely form of the time-series model. To decide whether we should apply a seasonal model we examine the periodic behavior of the time series. Also we determine whether or not the time series has a constant expected value pattern. The most important tools for identification of the model structure are the plot of the time series and the auto correllogram. In order to have reliable estimates a rule of thumb is that the maximum lag of the ACF should preferable not exceed 1/3 of the observation period. Also, in general, the more observations we have, the more reliable the estimates of the ACF.

From the ACF and the graph of the time series, we have to decide whether the series is non-stationary and/or exhibits a periodic pattern. Many time series in hydrology show a seasonal pattern with a period of one year. The ACF of a periodic time series also shows a cyclic pattern. In many cases in hydrological time series the seasonal behavior asks for a seasonal difference in the time series model.

Non-stationary behavior of a time series means that the expected value is not a constant, but a function of time. This function may have different forms. A step function occurs when there is a sudden change of the regime, for example a change in the groundwater level series due to the start of pumping groundwater. Also other forms of non-stationarity are possible indicating a (linear) trend in a time series. You can model the melting of a

glacier as such a trend. Often non-stationarity is reflected as a very slow decay in the ACF. Non-stationary behavior requires differencing in the regular part of the time series model.

The following examples show typical characteristics of time series and how that behavior is reflected in the auto correllogram. The first example is a groundwater head series, with an observation frequency of 24 times/year, given in Figure 6.16. The observation well is located close to a groundwater abstraction. At some point in time (around 70 time steps) the abstracted volume of groundwater increased. This is reflected in a sharp decrease of the groundwater head. It is obvious that in this case the expected value before the increase of the abstracted groundwater differs from the expected values after the increase of the abstraction. Therefore the time series is non-stationary. The corresponding auto correllogram is given in Figure 6.17.

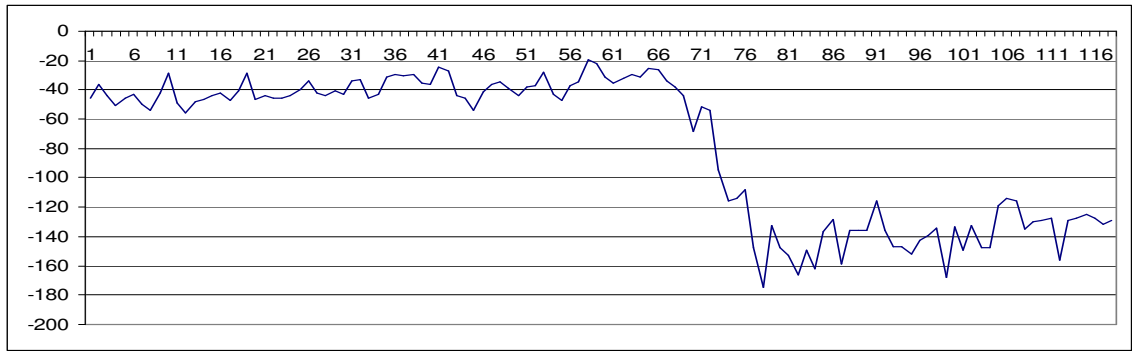



*Figure 6.16. Groundwater head series from an observation well close to a groundwater abstraction.*

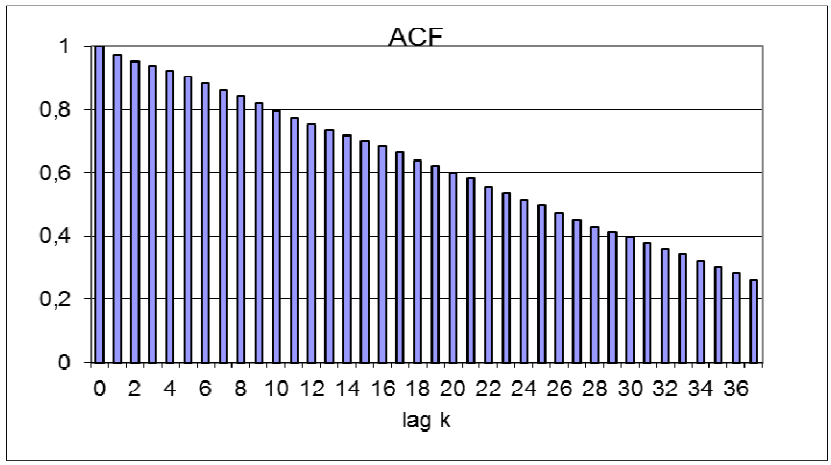



*Figure 6.17 Auto correllogram of the groundwater series given inFigure 6.16.*

The second example is the groundwater head series, with observation frequency of 24 times/year, given in Figure 6.18. This figure shows a clear seasonal behavior with a period of 24. The high values of the groundwater head occur in winter and the low values in summer. The seasonal behavior is reflected in the auto correllogram (Figure 6.19). The auto correlation shows a positive maximum at time lag 24 (=one year) and a negative maximum at time lag 12.

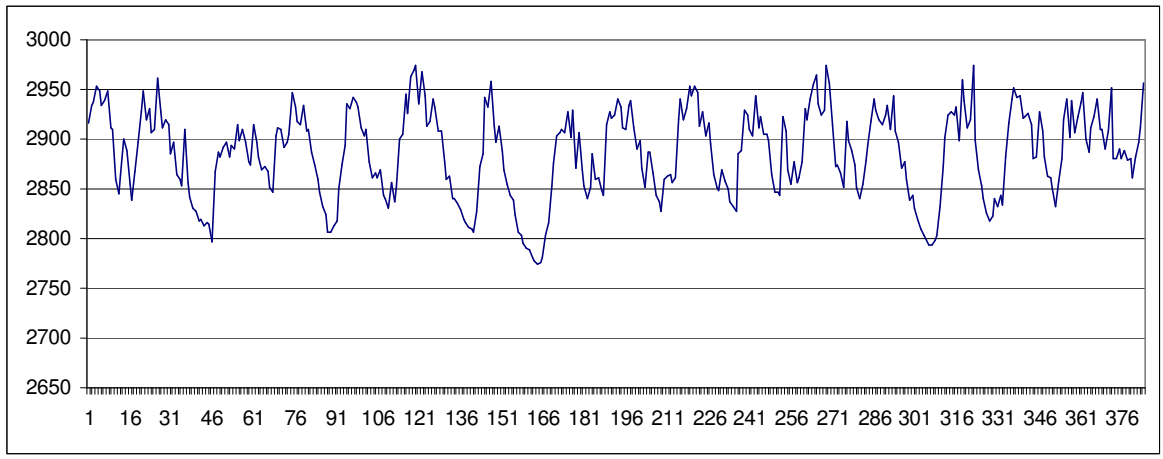



*Figure 6.18 Groundwater head series showing a seasonal pattern.*

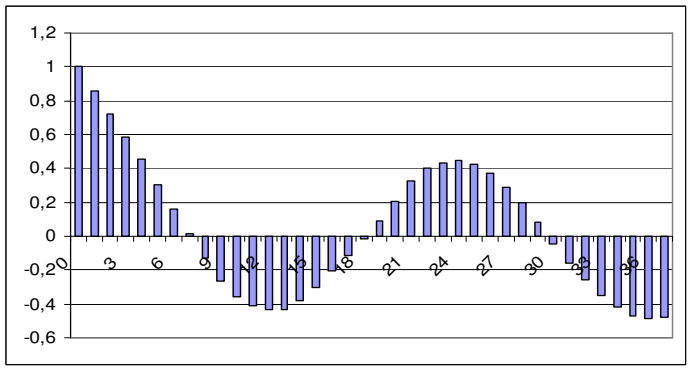



*Figure 6.19 Autocorrellogram of the groundwater series given in figure 6.18.*

The last example, given in Figure 6.20, is also a groundwater head series with observation frequency of 24 times/year. Like the previous example, the time series in figure 6.20 shows a seasonal behavior. In addition, there also is a positive trend over many years. The seasonal behavior is reflected in the auto correllogram (Figure 6.21) by the periodic pattern. The trend is reflected in the slow decay of the auto correlation, with the periodic pattern superimposed.

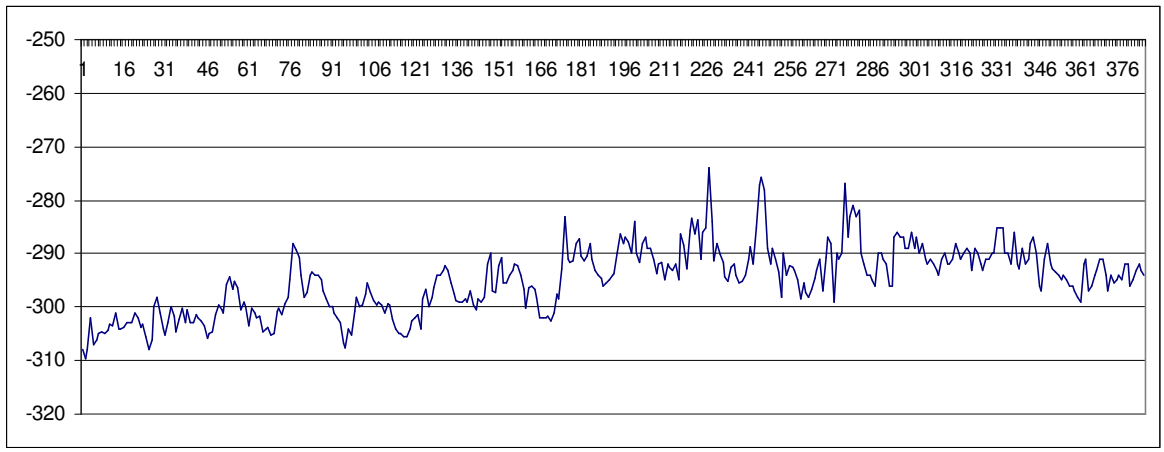



*Figure 6.20 Groundwater head series showing a seasonal pattern and a trend.*

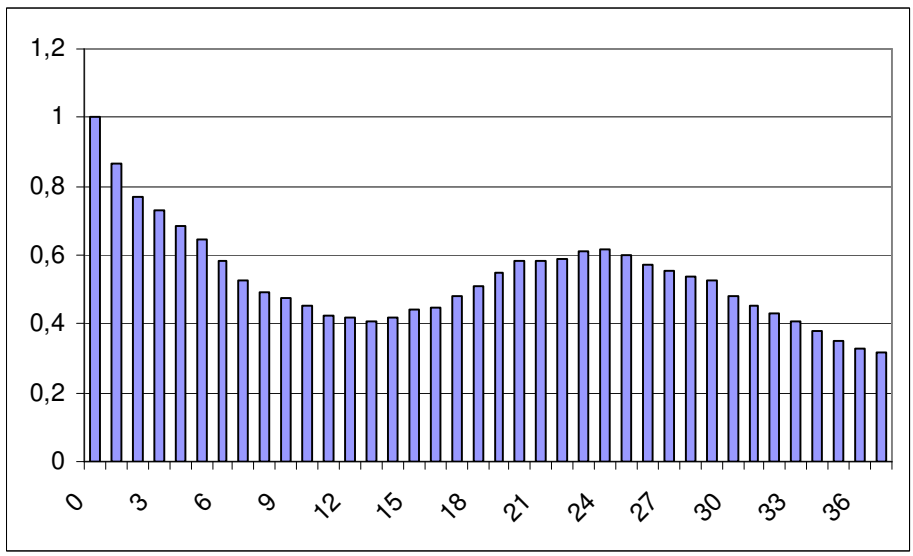



*Figure 6.21 Autocorrellogram of the groundwater series given in Figure 6.20.*

### 6.7.3 Estimation

The identification phase gives us indications whether the time series models should have a seasonal part, whether we should use a difference operator and it provides us with diagnostics of the memory of the system (in the form of the ACF). Therefore we have an idea what type of time series model will be the most appropriate. In order to estimate the model parameters we have to specify the following input to the estimation program:

- whether the model should have a seasonal component, and if so, the seasonal period (s);
- whether we should apply a difference operation in the regular model or the seasonal model, and if so, the order of difference operations (d and D);
- whether or not the series is zero mean;

- the order of AR terms in the regular model, and if applicable also in the seasonal model (p and P);
- the order of MA terms in the regular model, and if applicable also in the seasonal model (q and Q).

With these specifications the type of time series model is defined, and the parameters can be estimated. There are several computer codes available to perform the estimation. Most of them use a maximum likelihood estimation procedure. The estimation program privides of estimates of:

- the regular AR-parameters ( $\phi_i \;\; i = 1,..., p$ );
- the regular MA-parameters ( $\theta_i \;\; i = 1,..., q$ );
- for seasonal models the seasonal AR-parameters ( $\phi_i \;\; i = 1,..., P$ );
- for seasonal models the seasonal MA-parameters ( $\theta_i \;\; i = 1,..., Q$ );
- the residual variance $\sigma_a^2$;
- the expected value of the series $\mu_Z$.

In addition to the model parameters listed above, most estimation programs also provide useful model diagnostics, in particular:

- the residual series (which should be white noise) and its ACF;
- the standard error of the estimated parameters, (this enables establishing the statistical significance of each parameter);
- the correlation between the parameter estimation errors.

### 6.7.4 Diagnostics

We can estimate different time series models with different orders of the AR and MA coefficients. For example, for a time series we might have estimated an AR(1) model and an AR(2) model. Now the question arises which one should we use? Therefore we have to answer two questions:

1. Is the model valid?
2. Which of the models is the best one?

*Model validity*

As discussed in previous sections, for a time series model to be valid, it should be stationary and invertible. Mostly, stationarity and invertibility is checked in the estimation programs, as the estimation process will not converge if these conditions are not met. The residual series should be white noise. This can be tested for example with a $\chi^2$ test. Also the autocorrellogram of the residual series can be plotted and tested against the hypothesis of white noise. Finally, it is advisable to plot the residual series itself.

*Selecting the best time series*

To discriminate between several time series models, there are two effects that we should be aware of. The residual variance is the estimate of the variance of the white noise series. The smaller the residual variance, the more variation of the time series is captured in the time series model. Therefore in principle, we go for the time series model with the smallest residual variance. In general for the same time series, a model with more parameters will result in a smaller value of the residual variance. However, an increase of the number of parameters, also results in higher standard errors of the parameter estimates and higher correlations between parameter estimates. High standard errors of the parameter estimates implicate that the forecast uncertainty increases. Higher correlations between parameter estimates indicate dependence estimates and leads to redundant terms in de model. In conclusion, we strive to minimize the residual variance, but at the same time to develop models with the smallest number of parameters ('the principle of parsimony'). Here, we will not go into further detail, but in literature (for example Hipel and McLeod, 1996), several criteria are developed to find the balance between residual variance and model reliability.

## 6.8 Transfer function/noise models

### 6.8.1 Formulation of Transfer function models.

Not all time series that we measure in hydrology are best described by process driven by a white noise process only. Often the measured time series is the result of hydrological processes with known driving forces. Examples are the variation of the river stage driven by precipitation and the draw-down of groundwater head due to groundwater abstraction. An observed time series of a driving force is called the input series and the series that we like to describe as a function of the input series is called the output series. We can describe the output series by a linear transfer function (TF) of the input series. The general form of a TF($b,r,d,s$) model is:

$$\nabla^d Z_t = \frac{\Omega(B)}{\Delta(B)} X_{t-b} \tag{6.97}$$

With:

$$\Omega(B) = \omega_0 - \omega_1 B - .....\omega_s B^s$$
$$\Delta(B) = 1 - \delta_1 B - .....\delta_r B^r$$

Where: $X_t$ is the input series
$b$ the delay time between the input and the output.
$\omega_i$ is the moving average parameter of lag $i$ ($i$=0,…,$s$)

*s* is the order of the moving average part of the TF-model (Note that the symbol *s* is also being used to indicate the seasonal period in SARIMA models).

$\delta_i$ is the auto regressive parameter of lag *i* (*i*=1,…,*r*)

*r* is the order of the auto regressive part of the TF-model.

For example the TF(4,1,0,2) model is:

$$\nabla^0 Z_t = \frac{\omega_0 - \omega_1 B - \omega_2 B^2}{1 - \delta_1 B} X_{t-4} \qquad (6.98)$$
$$Z_t = \delta_1 Z_{t-1} + \omega_0 X_{t-4} - \omega_1 X_{t-5} - \omega_2 X_{t-6}$$

The general form of a linear transfer model is similar to an ARIMA model, but there are some differences.

- In stead of a white noise series, the driving force is an observed input series which can have any pattern. The input series might show auto correlation, a seasonal pattern, or a trend. This implies that the output series does not necessarily need to have a Gaussian probability distribution.
- The input series is an observed series, that often has another dimension than the output series. Therefore, also the parameters $\omega_i$ have a dimension and unlike the ARIMA model, $\omega_0 \neq 1$. For example, if the input series *X* is the precipitation in [mm/day] and the output series is the river stage in [m], than the dimension of the parameters $\omega_i$ is [m (mm/day)$^{-1}$].
- In principle, we can have a difference operation in a TF-model. This indicates a non stationary relation between input and output. In hydrology, non-stationary behavior of the output series is mostly due to a non-stationary input but the transfer function itself is stationary. In case the transfer function itself is non-stationary it is questionable whether a transfer function model is the most suitable way of describing the process. Therefore, in the reminder of this chapter we will not consider the difference operation in the TF models.

The behavior of an output series might be influenced by more that one input series. For example the groundwater head at a particular location might be influenced by precipitation, a groundwater abstraction and the surface water level. We can extent the TF-model (6.97) to accommodate more than one input series. For *m* input series the TF-model is defined as:

$$Z_t = Z_{1,t} + Z_{2,t} + .... + Z_{m,t} = \sum_{i=1}^{m} \frac{\Omega_i(B)}{\Delta_i(B)} X_{i,t-b} \qquad (6.99)$$

With:

$$\Omega_i(B) = \omega_{i,0} - \omega_{i,1} B - .....\omega_{i,s} B^s$$
$$\Delta_i(B) = 1 - \delta_{i,1} B - .....\delta_{i,r} B^r$$

Note that the orders *r* and *s* and the delay time *b* can be different for each input series.

The response series $Z_{i,t}$ is called the component of the output series due to the input $X_i$. In line with linear theory, the TF-models can be regarded impulse response functions.

### 6.8.2 Formulation of transfer function/noise models.

In hydrological practice an output series will never be exactly the response of a limited number of input series. The difference between the sum of all components in (6.99) and the output series is called the innovation series $n_t$. This innovation series, which is also called the *noise component*, can be modeled as an (S)ARIMA-model. The goal of TF/noise models is almost always to link patterns like seasonal patterns and trends in the output series to observed input series. Therefore here we restrict ourselves to noise components that can be modeled by a regular ARMA model. The formulation of a TF/noise model is given in (6.100) and depeicted schematically in Figure 6.22.

$$Z_t = Z_{1,t} + Z_{2,t} + .... + Z_{m,t} + n_t = \sum_{i=1}^{m} \frac{\Omega_i(B)}{\Delta_i(B)} X_{i,t-b} + \frac{\Theta(B)}{\Phi(B)} a_t \qquad (6.100)$$

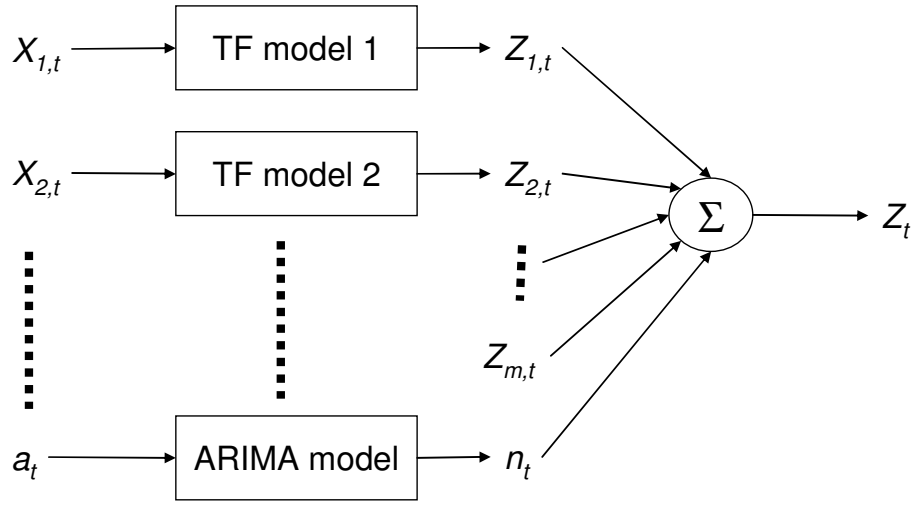



*Figure 6.22 Schematic structure of a TF/noise model*

### 6.8.3 Modeling aspects of TF/noise models.

Similar to the ARIMA models, we distinguish three phases in the modeling process: identification, estimation and verification/diagnostics.

*Identification.*

The relationship between two time series is characterized by the cross correlation function CCF (6.10). The CCF is illustrated by two examples given in the figures 6.23 and 6.24.

The first example is the cross correlation between an abstraction of groundwater and the groundwater head close to that abstraction. This situation is shown in Figure 6.23a. The observation frequency is 24 times/year. The observed time series of the groundwater abstraction and the groundwater head in the observation well is given in figure 6.23b. Figure 6.23c presents the corresponding CCF. As can be seen the groundwater head drops when the groundwater abstraction is increased. This is reflected in a large negative value of the CCF. The CCF also shows that the largest value occurs at time lag 0, so there is no time delay. The value of the CCF at time lag 0 is close to -1. This indicates that a large part of the pattern in the groundwater head series can be explained by the variations of the groundwater abstraction.

If we look at the observed series (6.23b) we might expect that the CCF shows high values only at a few small time lags, because the groundwater head is more or less the scaled mirror image of the groundwater abstraction. However, we see in the CCF a gradual decay for larger time lags, both at the positive and the negative side. This effect is caused by the auto correlation of both series, and it is not a property of the relation between the two series. The auto correlation of the series hampers a clear identification of the relationship between the series. As shown after the second example, in some cases we can remove the autocorrelation by means of ARIMA modeling.

In the second example the groundwater head variation is driven by the seasonal behavior of the precipitation and evapotranspiration (figure 6.24a). These driving forces are combined in the precipitation excess, defined by:

$$N_t = P_t - 0.8EP_t \tag{6.101}$$

Where: $N_t$ is the precipitation excess
$P_t$ is the precipitation and
$EP_t$ is the potential evapotranspiration (according to Penman)

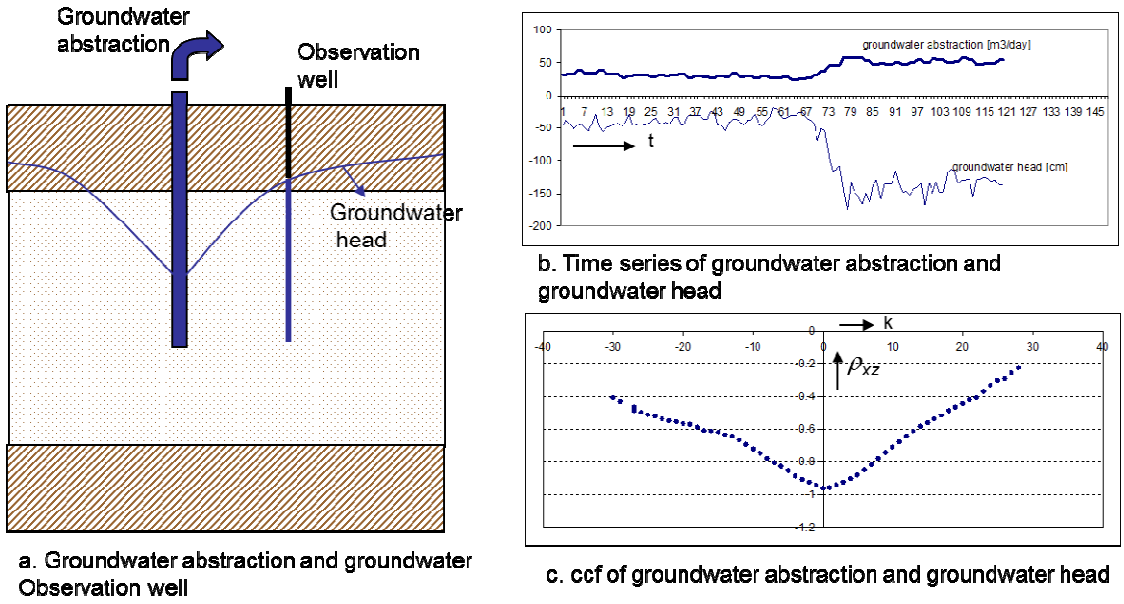



*Figure 6.23. Observed groundwater abstraction and groundwater head series and the corresponding CCF.*

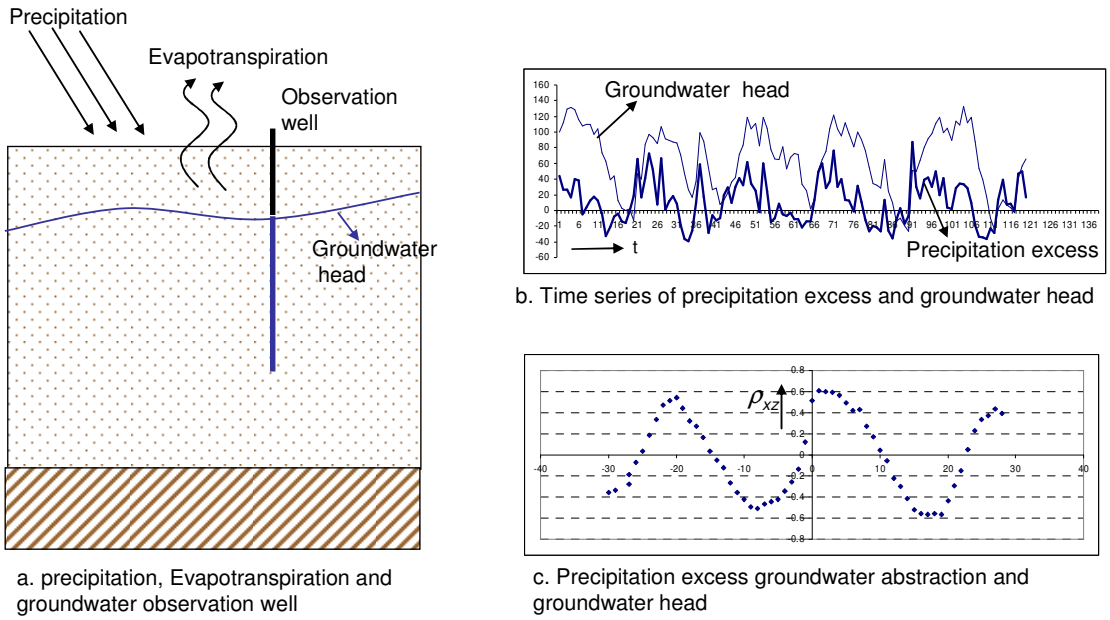



*Figure 6.24 Observed precipitation excess and groundwater head series and the corresponding CCF.*

Figure 6.24b shows that both the precipitation excess and the groundwater head have a seasonal pattern with a period of one year (the observation frequency is 24 times/year). The seasonal behavior can also be seen in the CCF (6.24c). Like in the previous example, the relation between the precipitation excess and the groundwater head is obscured by the (seasonal) auto correlation in both series. Nevertheless, figure 6.24c indicates a shift of one or two time steps, indicating that the full reaction of the groundwater head to precipitation excess is not instantaneous, but is spread over some time steps. Also, the

maximum value of the CCF is around 0.6. Obviously, the groundwater head series can not be fully explained by the precipitation excess.

As stated above, a clear view of the relation between two series is hampered by the auto correlation of both series. In case we can describe the input series by an (S)ARIMA model we can apply the following procedure, which is known as *prewhitening*. The principle of prewhitening is denoted in Figure 6.25.

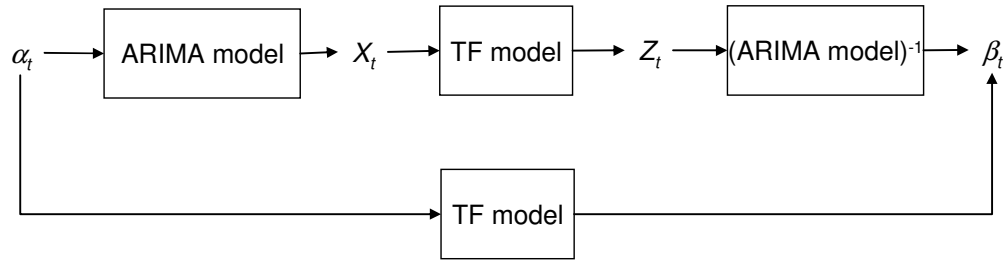



*Figure 6.25. Principle of prewhitening.*

Suppose the relation between the output series $Z_t$ and the input series $X_t$ is described by a TF model and the input series can be modeled by an ARIMA model. The residual white noise series of this ARIMA model is $\alpha_t$. Now, we also apply the ARIMA model of the input series to the output series $Z_t$. This yields the series $\beta_t$, which generally will not be a white noise series. The ARIMA model is a linear operation. Therefore, the TF model between $X_t$ and $Z_t$ is exactly the same as the TF model between $\alpha_t$ and $\beta_t$. Because $\alpha_t$ is a white noise series, the CCF between $\alpha_t$ and $\beta_t$ does not show temporal correlation due to the auto correlation, and we can identify the TF model between $\alpha_t$ and $\beta_t$ easier than between $X_t$ and $Z_t$.

To illustrate prewhitening we take the example of Figure 6.24. The SARIMA model of the input series $N_t$ is estimated as SARIMA(1,0,0)(0,1,0)$^{24}$

$$N_t = N_{t-24} - 0.33(N_{t-1} - N_{t-25}) + \alpha_t \tag{6.102}$$

Applying the SARIMA model (6.102) to the output series $Z_t$ yields the series $\beta_t$. The series $\alpha_t$ and $\beta_t$ are plotted in Figure 6.26 and the CCF $\rho_{\alpha\beta}$ is given in Figure 6.27

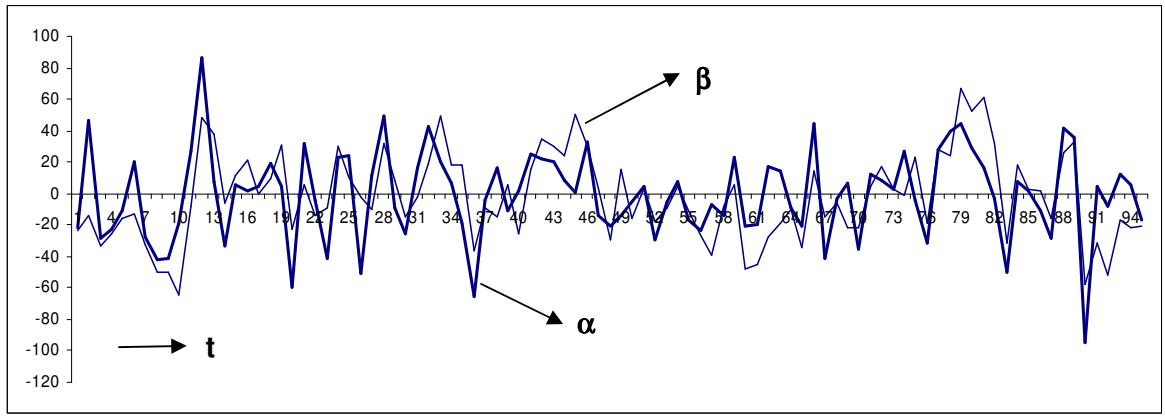



*Figure 6.26. The series $\alpha_t$ and $\beta_t$ obtained after after prewhitening $X_t$ and $Z_t$ with the SARIMA(1,0,0)(0,1,0)$^{24}$ model.*

The CCF $\rho_{\alpha\beta}$ in Figure 6.27 is much clearer than the CCF in Figure 6.25. Figure 6.27 shows a high value (0.64) at time lag 0 and a decay at the positive side. This indicates that there is no delay time (b=0) and we can expect an auto regressive part in the Transfer model.

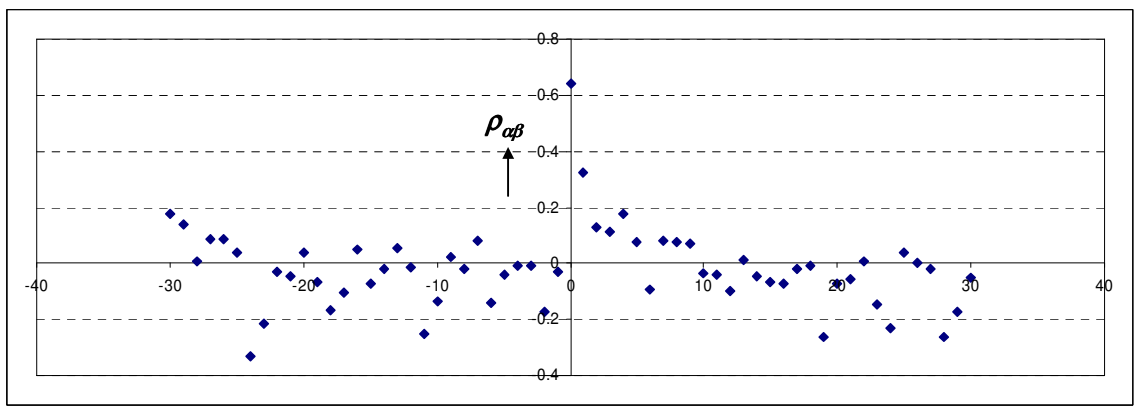



*Figure 6.27. CCF $\rho_{\alpha\beta}$ of the time series The series $\alpha_t$ and $\beta_t$ in Figure 6.26.*

*Estimation*

The estimation of a TF/noise model is similar to the estimation of an ARIMA model. We have to provide the orders(b,r,d,s) of all TF models and the orders of the noise model (see section 6.7.3). Generally, with TF/noise models we try to explain dominant patterns, like seasonal behavior and trends, by the input series. Therefore, most noise models (ARIMA model of the noise component) do not have a seasonal part or a difference operation. Often the noise model is very simple, for example an ARIMA(1,0,0) model.

The estimation program will return the same type of output as in case of ARIMA models. These are:

- the estimated values of all parameters and there statistical significances;
- the expected value and variance of the residual white noise series;
- the ACF of the residuals;
- the standard error of all estimates and the correlation between all estimation errors.

Often also time series of all individual components ($Z_{i,t}$ and $n_t$) and the residual white noise series $a_t$ can be extracted from the estimation program.

*Verification/diagnostics*

As with the ARIMA models, we have to check whether the TF/noise model is valid, and we have to select the 'best' model. The validity check is the same as in case of an ARIMA model (see 6.7.4), by checking the residual white noise series. To choose the best model, we select the model with the best balance between minimizing the residual variance and the standard error of the parameter estimation. In particular we have to pay attention to the correlation between the estimation errors of the parameters. As with ARIMA models, highly correlated estimation errors of a TF model indicate that the order of the TF model should be reduced. More serious is correlation between estimated parameters of different TF models. High correlation between parameters of two TF models implies that we can not separate the influence of one input series from the other. Consequently, we can not use the individual TF models as the input response functions.

### 6.8.4 Use of TF/noise models; an example

Consider the situation of Figure 6.23. The groundwater head time series is modeled as a TF/noise model with one input series, the abstraction $Q_t$. The TF/noise model is:

$$\begin{aligned} Z_{1,t} &= \omega_0 Q_t - \omega_1 Q_{t-1} \\ (n_t - \overline{n}) &= \phi_1 (n_{t-1} - \overline{n}) + a_t \\ Z_t &= Z_{1,t} + n_t \end{aligned} \qquad (6.103)$$

The parameter values, residual white noise variance and the corresponding standard errors are given in the table below.

| parameter | value | s.e. |
|---|---|---|
| $\omega_0$ [1000·cm/m$^3$] | -2.59126 | 0.3327 |
| $\omega_1$ [1000·cm/m$^3$] | 1.75496 | 0.3297 |
| $\phi_1$ [-] | 0.47027 | 0.08078 |
| $\overline{n}$ [cm] | 172.35 | 7.12726 |
| $\sigma_a^2$ [cm$^2$] | 198 | - |

This table shows that the standard errors for all parameters are small in comparison to the values of the corresponding parameters. The parameter $\omega_0$ is negative and the parameter $\omega_1$ is positive. This is due to the fact that a positive value of the groundwater abstraction results in a drawdown of the groundwater level (see (6.103)).

The correlation matrix between the estimation errors is:

| | $\omega_0$ | $\omega_1$ | $\phi_1$ |
|---|---|---|---|
| $\omega_0$ | 1 | 0.86 | 0.17 |
| $\omega_1$ | 0.86 | 1 | 0.16 |
| $\phi_1$ | 0.17 | 0.16 | 1 |

This table shows that the parameters of the TF model are highly correlated (0.86), with might indicate redundant information and we might consider a transfer model with only one moving average term.. However, because both parameters are always used together in the same model and the correlation between these parameters and the parameters of the noise component is small, we can successfully separate the component of the groundwater head due to the abstraction and the noise component.

*Decomposition.*

In Figure 6.28 the decomposition of the groundwater head series is graphically displayed. The groundwater head is split into a component of the groundwater abstraction and a noise component. In Figure 6.28a the observed groundwater head series (+) and the component of the groundwater abstraction (line) are plotted in the same figure. The component of the groundwater abstraction is the drawdown of the groundwater head due to the abstraction. In Figure 6.26b the noise component is given. This component represents all other variations of the groundwater head.

*Forecasting*

We can also apply the TF model to forecast the effect of an increase of the groundwater abstraction. This is done by simply providing values for the groundwater abstraction $Q_t$ to the TF model in (6.103). In particular the equilibrium drawdown ($Z_{1,\infty}$) is of interest. This is the drawdown that will occur if the groundwater abstraction is constant in time ($Q_\infty$). From (6.103) it follows that:

$$Z_{1,\infty} = (\omega_0 - \omega_1)Q_\infty = -4.34622Q_\infty \qquad (6.104)$$

With the standard error of both parameters and the corresponding correlation coefficient we can calculate the standard error $\sigma_{(\omega_0-\omega_1)} = 0.175$ and we can construct the 95% confidence interval for any volume of abstracted groundwater:

$$(-4.34622 \pm 0.175)Q_{\infty} \tag{6.105}$$

The estimated equilibrium drawdown and 95% confidence interval (assuming Gaussian error in the parameters $\omega_0, \omega_1$) is given in Figure 6.29.

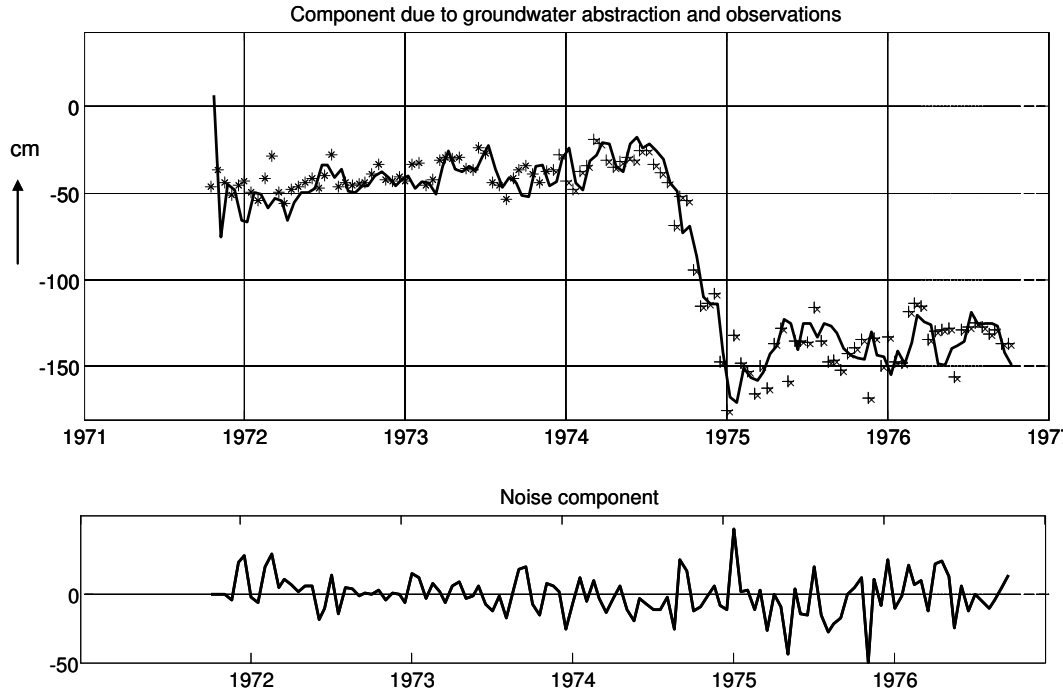



*Figure 6.28 Decomposition of a groundwater head series in a component due to groundwater abstraction and a noise component.*

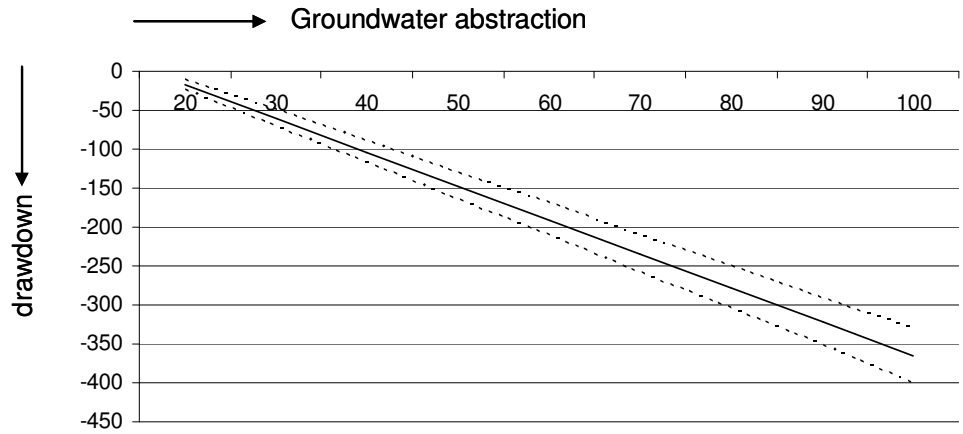



*Figure 6.29. Estimated equilibrium drawdown as function of the abstracted volume, inclusive the 95% confidence interval.*

# 7. Geostatistics

## 7.1 Introduction

Geostatistics is the part of statistics that is concerned with geo-referenced data, i.e. data that are linked to spatial coordinates. To describe the spatial variation of the property observed at data locations, the property is modelled with a spatial random function (or random field) $Z(\mathbf{x})$, $\mathbf{x}^T = (x,y)$ or $\mathbf{x}^T = (x,y,z)$. The focus of geostatistics can be further explained by Figure 7.1. Suppose that the values of some property (for instance hydraulic conductivity) have been observed at the four locations $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$ and that, respectively, the values $z_1$, $z_2$, $z_3$ and $z_4$ have been found at these locations. Geostatistics is concerned with the unknown value $z_0$ at the non-observed location $\mathbf{x}_0$. In particular, geostatistics deals with:

1. spatial interpolation and mapping: predicting the value of $Z_0$ at $\mathbf{x}_0$ as accurately as possible, using the values found at the surrounding locations (note that $Z_0$ is written here in capitals to denote that it is considered to be a random variable);
2. local uncertainty assessment: estimating the probability distribution of $Z_0$ at $\mathbf{x}_0$ given the values found at the surrounding locations, i.e. estimating the probability density function $f_z(z_0; \mathbf{x}_0 \mid z_1(\mathbf{x}_1), z_2(\mathbf{x}_2), z_3(\mathbf{x}_3))$. This probability distribution expresses the uncertainty about the actual but unknown value $z_0$ at $\mathbf{x}_0$;
3. simulation: generating realisations of the conditional RF $Z(\mathbf{x}|z(\mathbf{x}_i), i = 1,..,4)$ at many non-observed locations $\mathbf{x}_i$ simultaneously (usually on a lattice or grid) given the values found at the observed locations; e.g. hydraulic conductivity is observed at a limited number of locations but must be input to a groundwater model on a grid.

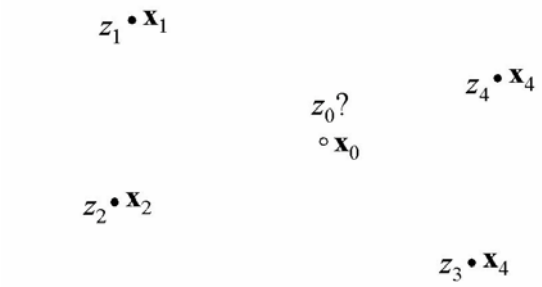



*Figure*7.1 *Focus of geostatistics*

Geostatistics was first used as a practical solution to estimating ore grades of mining blocks using observations of ore grades that were sampled preferentially. i.e. along outcrops (Krige, 1993 ). Later it was extended to a comprehensive statistical theory for geo-referenced data (Matheron, 1970) Presently, geostatistics is applied in a great number of fields such as petroleum engineering, hydrology, soil science, environmental pollution and fisheries. Standard text books have been written by David (1977) , Journel and Huijbregts (1998) , Isaaks and Srivastava (1989) and Goovaerts (1997). Some important hydrological problems that have been tackled using geostatistics are among others:

- spatial interpolation and mapping of rainfall depths and hydraulic heads;
- estimation and simulation of representative conductivities of model blocks used in groundwater models;
- simulation of subsoil properties such as rock types, texture classes and geological facies;
- uncertainty analysis of groundwater flow and -transport through heterogeneous formations (if hydraulic conductivity, dispersivity or chemical properties are spatially varying and largely unknown) (see chapter 8).

The remaining of this chapter is divided into four parts. The first part briefly revisits descriptive statistics, but now in a spatial context. The second part is concerned with spatial interpolation using a technique called kriging. The third part uses kriging for the estimation of the local conditional probability distribution. The last part deals with the simulation of realisations of spatial random functions.

## 7.2 Descriptive spatial statistics

*Declustering*

In this section we will briefly revisit the subject of descriptive statistics, but now focussed on spatial (i.e. geo-referenced) data. Looking at Figure 7.1 it can be seen that not all observation locations are evenly spread in space. Certain location appear to be clustered. This can for instance be the case because it is convenient to take a number of samples close together. Another reason could be that certain data clusters are taken purposively, e.g. to estimate the short distance variance. If the histogram or cumulative frequency distribution of the data is calculated with the purpose of estimating the true but unknown spatial frequency distribution of an area, it would not be fair to give the same weight to clustered observations as to observations that are far from the others. The latter represent a much larger area and thus deserve to be given more weight. To correct for the clustering effect *declustering methods* can be used. Here, one particular declustering method called polygon declustering is illustrated. Figure 7.2 shows schematically a spatial array of measurement locations. The objective is to estimate the spatial statistics (mean, variance, histogram) of the property (e.g. hydraulic conductivity) of the field. The idea is to draw Thiessen polygons around the observation locations first: by this procedure each location of the field is assigned to the closest observation. The relative sizes of the Thiessen polygons are used as declustering weights: $w_i = A_i/(\sum_j A_j)$. Using these weights the declustered histogram and cumulative frequency distribution can be calculated as shown in Figure 7.3, as well as the declustered moments such as the mean and variance:

$$m_z = \sum_{i=1}^{n} w_i z_i \tag{7.1}$$

$$s_z^2 = \sum_{i=1}^{n} w_i (z_i - m_z)^2 \tag{7.2}$$

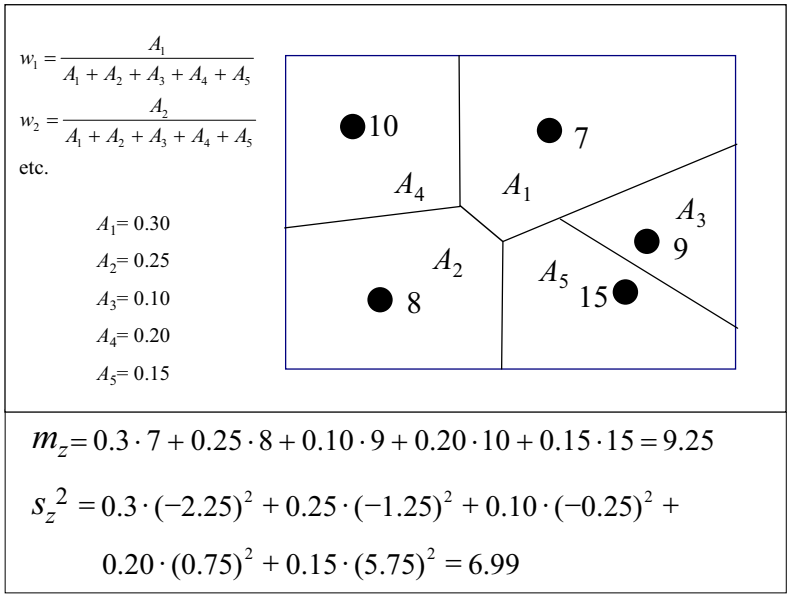



*Figure* 7.2 *Schematic example of polygon declustering.*

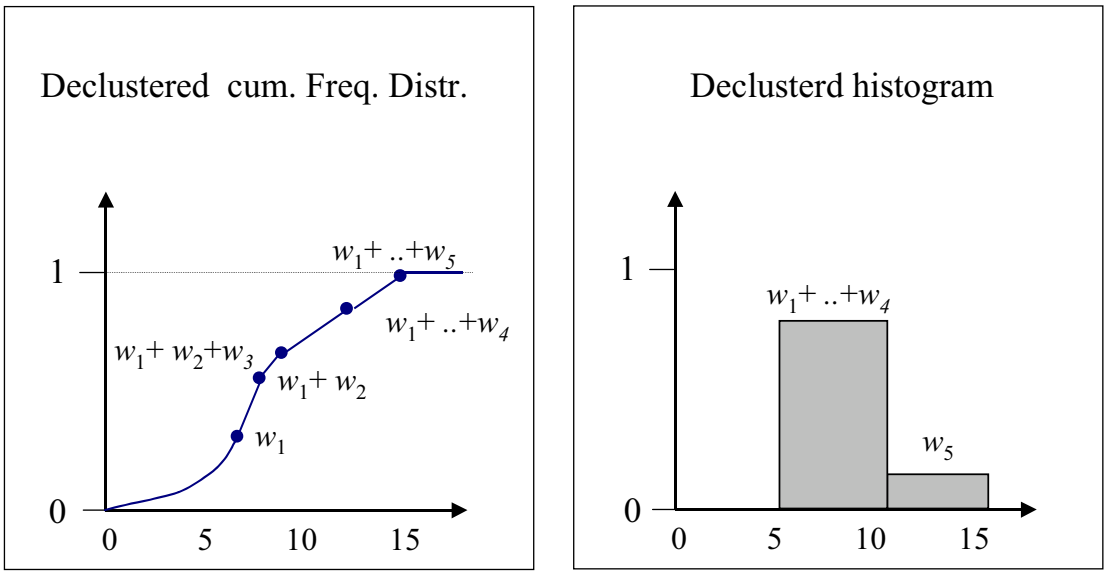



*Figure* 7.3 *Schematic example of declustered frequency distributions*

The effect of declustering can be demonstrated using the synthetic Walker lake data-set shown in Figure 2.1 (all of the larger numerical examples shown in this chapter are based on the Walker-lake data set. The geostatistical analyses and the plots are performed using the GSLIB geostatistical software of Deutsch and Journel (1998). Figure 2.1 only shows the 140 values at the sample locations. The associated histogram is shown in Figure 2.2. Because this is a synthetic data set we also have the exhaustive dat of the entire area (2500 values). Figure 7.4 shows the declusterd histogram based on the 140 data and the "true" histogram based on 2500 values. Clearly, the are very much alike, while the histogram based on the non-weighted data (Figure 2.2) is much different. The estimated mean without weighting equals 4.35 which is much too large. The reason is the existance of clusters with very high data values present in the observations (see Figure 2.1). Declustering can correct for this as can be seen from the decslustered mean in Figure 7.4 which is 2.53 and very close to te true mean of 2.58.

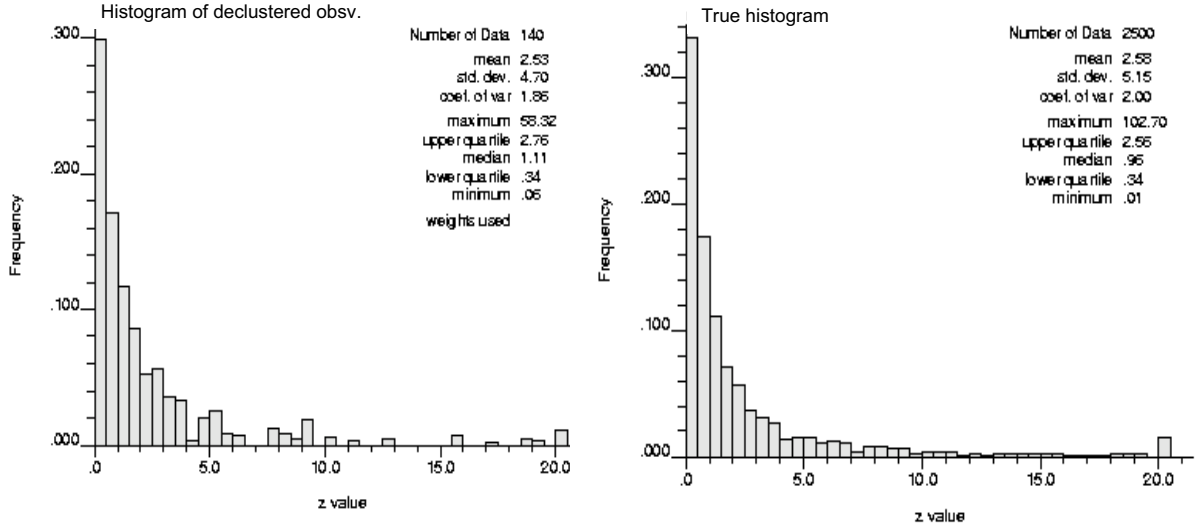



*Figure* 7.4 *Declustered histogram of the* 140 *data values* (*left*) *and the true histogram of the Walker lake data set* (*right*)

*Semivariance and correlation*

Using the Walker-Lake data set of 140 observations we will further illustate the concept of the semivariogram and the correlation function. Figure 7.5 shows scatter plots of the $z(\mathbf{x})$ and $z(\mathbf{x}+\mathbf{h})$ for $|\mathbf{h}| = 1, 5, 10, 20$ units (pixels) apart. For each pair of points the distance $d_i$ to the one-to-one line is can be calculated. The semivariance of a given distance is given by (with $n(h)$ the number of pairs of points that are a distance $h = |\mathbf{h}|$ apart):

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} d_i^2 = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(\mathbf{x}+\mathbf{h}) - z(\mathbf{x})]^2 \qquad (7.3)$$

and the correlation coefficient:

$$\hat{\rho}(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} \frac{z(\mathbf{x}+\mathbf{h})z(\mathbf{x}) - m_{z(\mathbf{x}+\mathbf{h})} m_{z(\mathbf{x})}}{s_{z(\mathbf{x}+\mathbf{h})} s_{z(\mathbf{x})}} \tag{7.4}$$

where $m_{z(\mathbf{x}+\mathbf{h})}, m_{z(\mathbf{x})}$ and $s_{z(\mathbf{x}+\mathbf{h})}, s_{z(\mathbf{x})}$ are the means and variances of the $z(\mathbf{x})$ and $z(\mathbf{x}+\mathbf{h})$ data values respectively. These estimators were already introduced in chapter 5 for data that are not on a grid. Figure 7.6 shows plots of the semivariance and the correlation as a function of distance. These plots are called the semivariogram and the correlogram respectively. If we imagine the data $z$ to be observations from a realisation of a random function $Z(\mathbf{x})$ and this random function is assumed to be intrinsic or wide sense stationary (see chapter 5) then (7.3) and (7.4) are estimators for the semivariance function and the correlation function.

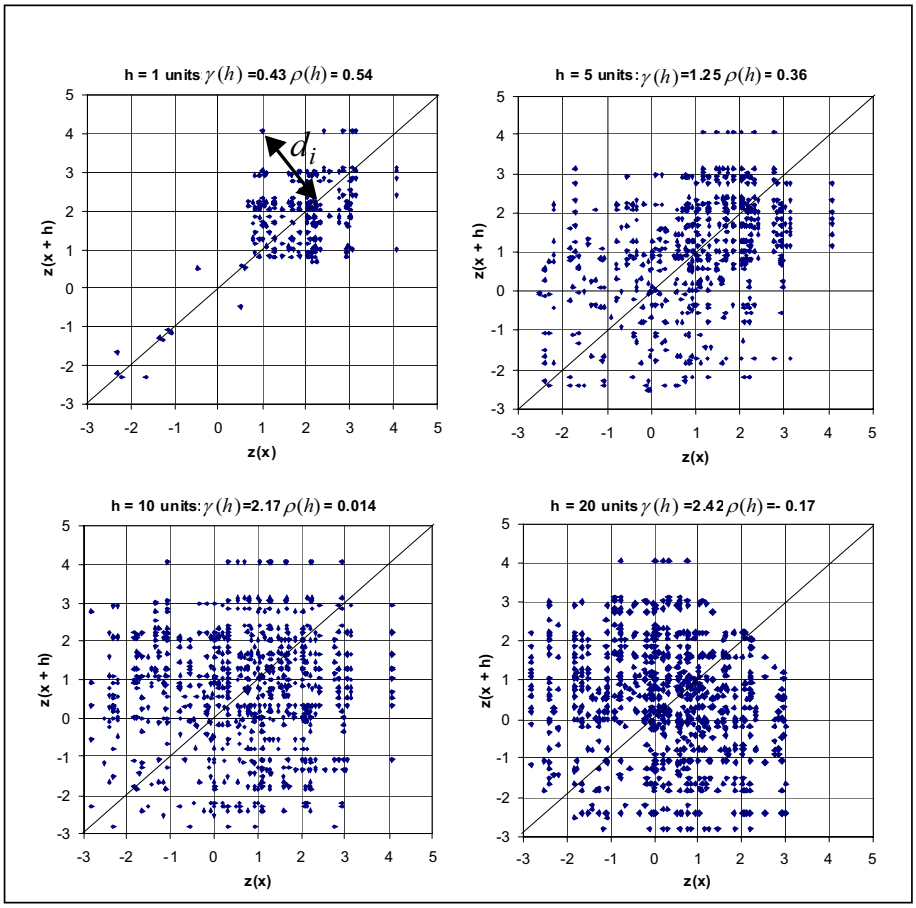



*Figure* 7.4 *Scatter plots of* $z(x)$ *and* $z(x+h)$ *for* $|h| = 1, 5, 10, 20$ *units* (*pixels*) *apart from the Walker lake data set.*

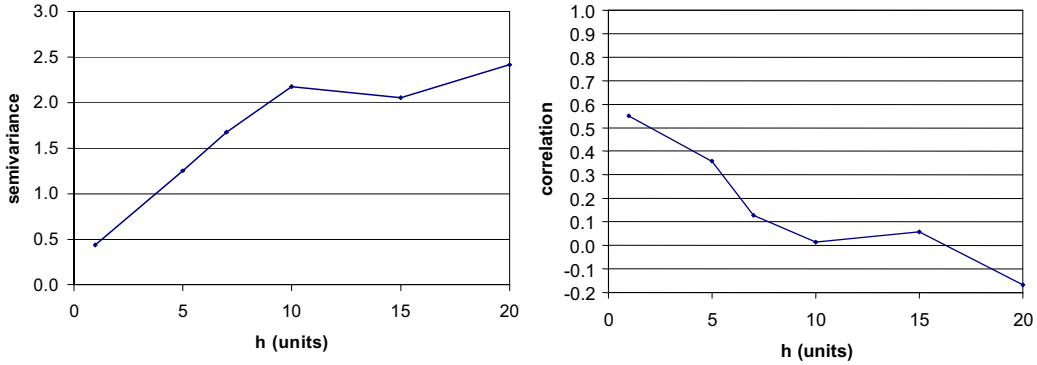



*Figure* 7.5 *Semivariogram and correlogram based on Figure* 7.4.

## 7.3 Spatial interpolation by kriging

Kriging is a collection of methods that can be used for spatial interpolation. Kriging provides optimal linear predictions at non-observed locations by assuming that the unknown spatial variation of the property is a realisation of a random function that has been observed at the data points only. Apart from the prediction, kriging also provides the variance of the prediction error. Here, two kriging variants are discussed: simple kriging and ordinary kriging which are based on slightly different random function models.

### 7.3.1 Simple kriging

#### Theory

The most elementary of kriging methods is called simple kriging and is treated here. Simple kriging is based on a RF that is wide sense stationary, i.e. with the following properties (see also chapter 5):

$$E[Z(\mathbf{x})] = \mu_Z = \text{constant}$$

$$Var[Z(\mathbf{x})] = E[Z(\mathbf{x}) - \mu_Z]^2 = \sigma_Z^2 = \text{constant (and finite)}$$

$$COV[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = E[\{Z(\mathbf{x}_1) - \mu_Z\}\{Z(\mathbf{x}_2) - \mu_Z\}] = C_Z(\mathbf{x}_2 - \mathbf{x}_1) = C_Z(\mathbf{h})$$

Simple kriging is the appropriate kriging method if the RF is second order stationary and the mean of the RF $E[Z(\mathbf{x})] = \mu$ is known without error. With simple kriging a predictor $\widehat{Z}(\mathbf{x}_0)$ is sought that

1. is a linear function of the surrounding data,
2. is unbiased: $E[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$,
3. and has the smallest possible error, i.e. $\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)$ is minimal.

A linear and unbiased predictor is obtained when considering the following weighted average of deviations from the mean:

$$\widehat{Z}(\mathbf{x}_0) = \mu_Z + \sum_{i=1}^{n} \lambda_i [Z(\mathbf{x}_i) - \mu_Z] \tag{7.5}$$

with $Z(\mathbf{x}_i)$ the values of $Z(\mathbf{x})$ at the surrounding observation locations. Usually, not all observed locations are included in the predictor, but only a limited number of locations within a given search neighbourhood. Equation (7.5) is unbiased by definition:

$$E[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \mu_Z + \sum_{i=1}^{n} \lambda_i E[Z(\mathbf{x}_i) - \mu_Z] - E[Z(\mathbf{x}_0)] =$$

$$\mu_Z + \sum_{i=1}^{n} \lambda_i [\mu_Z - \mu_Z] - \mu_Z = 0 \tag{7.6}$$

The weights $\lambda_i$ should be chosen such that the prediction error is minimal. However, as the real value $z(\mathbf{x}_0)$ is unknown, we cannot calculate the prediction error. Therefore, instead of minimizing the prediction error we must be satisfied with minimizing the variance of the prediction error $Var[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]$. Because the predictor is unbiased, the variance of the prediction error can be written as:

$$Var[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = E[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]^2 =$$

$$E\left[ \sum_{i=1}^{n} \lambda_i E[Z(\mathbf{x}_i) - \mu_Z] - [Z(\mathbf{x}_0) - \mu_Z] \right]^2 =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j E\{[Z(\mathbf{x}_i) - \mu_Z][Z(\mathbf{x}_j) - \mu_Z]\} -$$

$$2 \sum_{i=1}^{n} \lambda_i E[\{Z(\mathbf{x}_i) - \mu_Z\}\{Z(\mathbf{x}_0) - \mu_Z\}] + E[Z(\mathbf{x}_0) - \mu_Z]^2 \tag{7.7}$$

Using the definition of the covariance of a second order stationary SF $E[\{Z(\mathbf{x}_i) - \mu_Z\}\{Z(\mathbf{x}_j) - \mu_Z\}] = C_Z(\mathbf{x}_i - \mathbf{x}_j)$ and $C_Z(0) = \sigma_Z^2$, we obtain for the variance of the prediction error:

$$Var[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) - 2 \sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0) + \sigma_Z^2 \tag{7.8}$$

To obtain the mininum value of Equation (7.8) we have to equate all its partial derivatives with

respect to the $\lambda_i$ to zero:

$$\frac{\partial}{\partial \lambda_i} Var[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] \; = 2\sum_{j=1}^{n} \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) - 2C_Z(\mathbf{x}_i - \mathbf{x}_0) = 0$$

$$i = 1,..,n \tag{7.9}$$

This results in the following system of $n$ equations referred to as the *simple kriging system*:

$$\sum_{j=1}^{n} \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) = C_Z(\mathbf{x}_i - \mathbf{x}_0) \qquad i = 1,..,n \tag{7.10}$$

The $n$ unknown values $\lambda_i$ can be uniquely solved from these $n$ equations if all the $\mathbf{x}_i$ are different. The predictor (7.5) with the $\lambda_i$ found from solving (7.10) is the one with the minimum prediction error variance. This variance can be calculated using equation (7.8). However, it can be shown (e.g. de Marsily, 1986, p.290) that the variance of the prediction error can be written in a simpler form as:

$$Var[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] \; = \sigma_Z^2 - \sum_{i=1}^{n} \lambda_i C_Z(\mathbf{x}_i - \mathbf{x}_0) \tag{7.11}$$

The error variance very nicely shows how kriging takes advantage of the spatial dependence of $Z(\mathbf{x}_i)$. If only the marginal probability distribution had been estimated from the data and the spatial coordinates had not been taken into account, the best prediction for every non-observed location would have been the mean $\mu_Z$. Consequently, the variance of the prediction error would have been equal to $\sigma_Z^2$. As the larger kriging weights are positive, it can be seen from (7.11) that the prediction error variance of the kriging predictor is always smaller than the variance of the RF.

To obtain a positive error variance using Equation (7.11) the function $C(\mathbf{h})$ must be positive definite. This means that for all possible $\mathbf{x}_1,...\mathbf{x}_n \in \Re^N$ ($N$ = 1,2 or 3) and for all $\lambda_1,...,\lambda_n \in \Re$ the following inequality must hold:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C_Z(\mathbf{x}_i - \mathbf{x}_j) > 0 \tag{7.12}$$

It is difficult to ensure that this is the case for any covariance function. Therefore, we cannot just estimate a covariance function directly from the data for a limited number of separation distances and then obtain a continuous function by linear interpolation between the experimental points (such as in Figure 7.5) . If such a covariance function were used in (7.10) , Equation (7.11) would not necessarily lead to a positive estimate of the prediction error variance. In fact, there are only a limited number of functions for which it is proven that inequality (7.12) will always hold. So the practical solution used in kriging is to take one of these 'permissible' functions and fit it through the points of the experimental covariance function. Next, the values of the fitted function are used to build the kriging system (7.10) and

to estimate the kriging variance using (7.11). Table 7.1 gives a number of covariance functions that can be used for simple kriging (i.e. using a wide sense stationary RF). Such a table was already introduced in chapter 5 but is repeated for convenience here. Figure 7.6 shows an example of an exponential model that is fitted to the estimated covariances. Of course, in case of second order stationarity the parameter $c$ should be equal to the variance of the RF: $c = \sigma_Z^2$.

*Table*7.1 *Permissible covariance functions for simple kriging*; *his the length of the lag vector*

| | | |
|---|---|---|
| (a) spherical model | $C(h) = \begin{cases} c\left[1 - \frac{3}{2}\left(\frac{h}{a}\right) + \frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{if } h \le a \\ 0 & \text{if } h > a \end{cases}$ | |
| (b) exponential model | $C(h) = c\exp(-h/a)$ | |
| (c) Gaussian model | $C(h) = c\exp(-[h/a]^2)$ | |
| (d) nugget model | $C(h) = \begin{cases} c & \text{if } h = 0 \\ 0 & \text{if } h > 0 \end{cases}$ | |

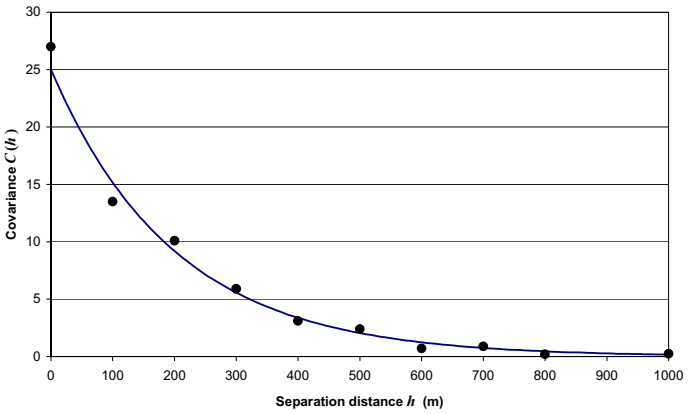



*Figure* 7.6 *Example of an exponential covariance model fitted to estimated covariances*

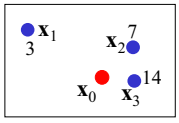
Some remarks should be made about the nugget model. The nugget stems from the mining practice. Imagine that we find occasional gold nuggets in surrounding rock that doesn't contain any gold itself. If we were to estimate the covariance function of gold content from our observation, we would get the nugget model with $c = \sigma_Z^2 = p(1-p)$ (with $p$ the probability of finding a gold nugget).

Any linear combination of a permissible covariance model is a permissible covariance model itself. Often a combination of a nugget model and another model is observed, e.g:

$$C(h) = \begin{cases} c_0 + c_1 & \text{if } h = 0 \\ c_1 \exp(-h/a) & \text{if } h > 0 \end{cases} \tag{7.13}$$

where $c_0 + c_1 = \sigma_Z^2$. In this case $c_0$ is often used to model the part of the variance that is attributable to observation errors or spatial variation that occurs at distances smaller than the minimal distance between observations.
The box below shows a simple numerical example of simple kriging.

Box 3: Simple kriging example




Spatial lay-out

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 3 | 1 | 1 |
| 1 | 3 | 0 | 3 | 4 |
| 2 | 1 | 3 | 0 | 1 |
| 3 | 1 | 4 | 1 | 0 |

Distance table (units)

$$\begin{aligned} \lambda_1 C(\mathbf{x}_1 - \mathbf{x}_1) + \lambda_2 C(\mathbf{x}_1 - \mathbf{x}_2) + \lambda_3 C(\mathbf{x}_1 - \mathbf{x}_3) &= C(\mathbf{x}_1 - \mathbf{x}_0) \\ \lambda_1 C(\mathbf{x}_2 - \mathbf{x}_1) + \lambda_2 C(\mathbf{x}_2 - \mathbf{x}_2) + \lambda_3 C(\mathbf{x}_2 - \mathbf{x}_3) &= C(\mathbf{x}_2 - \mathbf{x}_0) \\ \lambda_1 C(\mathbf{x}_3 - \mathbf{x}_1) + \lambda_2 C(\mathbf{x}_3 - \mathbf{x}_2) + \lambda_3 C(\mathbf{x}_3 - \mathbf{x}_3) &= C(\mathbf{x}_3 - \mathbf{x}_0) \end{aligned}$$

$$C(\mathbf{x}_i - \mathbf{x}_j) = 22.69 e^{\frac{|\mathbf{x}_i - \mathbf{x}_j|}{2}}$$
$$\mu = 7.1$$

$$\begin{aligned} 22.69\lambda_1 + 5.063\lambda_2 + 3.071\lambda_3 &= 5.063 \\ 5.063\lambda_1 + 22.69\lambda_2 + 13.76\lambda_3 &= 13.76 \\ 3.071\lambda_1 + 13.76\lambda_2 + 22.69\lambda_3 &= 13.76 \end{aligned}$$

$$\lambda_1 = 0.0924 \qquad \lambda_2 = 0.357 \qquad \lambda_3 = 0.378$$

$$Z^*(\mathbf{x}_0) = 7.1 + 0.0924 \cdot (-4.1) + 0.357 \cdot (-0.1) + 0.378 \cdot 6.9 = 9.29$$

$$\begin{aligned} &\text{var}[Z^*(\mathbf{x}_0) - Z^*(\mathbf{x}_0)] = \\ &\quad 22.69 - 0.0924 \cdot 5.063 - 0.357 \cdot 13.76 - 0.378 \cdot 13.76 = 12.11 \end{aligned}$$

A simple example of simple kriging. Top left gives the spatial lay-out of the data points and the target location for prediction of the property. Top right shows the table of distances between these locations. The kriging system is shown therafter, with the $\mathbf{x}_i$-$\mathbf{x}_j$ covariances on the left and the $\mathbf{x}_i$-$\mathbf{x}_0$ on the right. Using the assumed mean and covariance function shown, the next boxes show the numerical solution of the kriging equations and the evaluation of the kriging predictor and the kriging variance.

## Practice

The practical application of simple kriging would involve the mapping of some variable observed at a limited number of locations. In practice, the kriging routine would consist of the following steps, which will be illustrated with the Walker-lake dataset:

1. *Estimate the mean and the covariance function from the data*

   The mean value of the Walker-lake data based on the observations and declustering is 2.53.

2. *Fit a permissible covariance model to the experimental semivariogram*

Usually one does not estimate the covariance function but the semivariogram when kriging. The semivariogram is somewhat better suited for estimating data that are irregularly distributed in space. After fitting a semivariogram function that is suited for wide sense stationary processes (See the first four models in Table 7.2), the covariance function can be obtained through Equation (5.19): $C_Z(\mathbf{h}) = \sigma_Z^2 - \gamma_Z(\mathbf{h})$. Figure 7.7 shows the semivariogram of the Walker-lake data set based on 140 data points and the fitted model:

If kriging is used for making maps, the locations where the predictions are made are usually located on a grid. So, when in the following steps we refer to a prediction location $\mathbf{x}_0$ we refer to a location on this grid. Thus, the following steps are repeated for every grid node:

**3**. *Solve the simple kriging system*

Using Equation (7.11) and the covariance function $C_Z(\mathbf{h})$ the $\lambda_i$ are obtained for location $\mathbf{x}_0$.

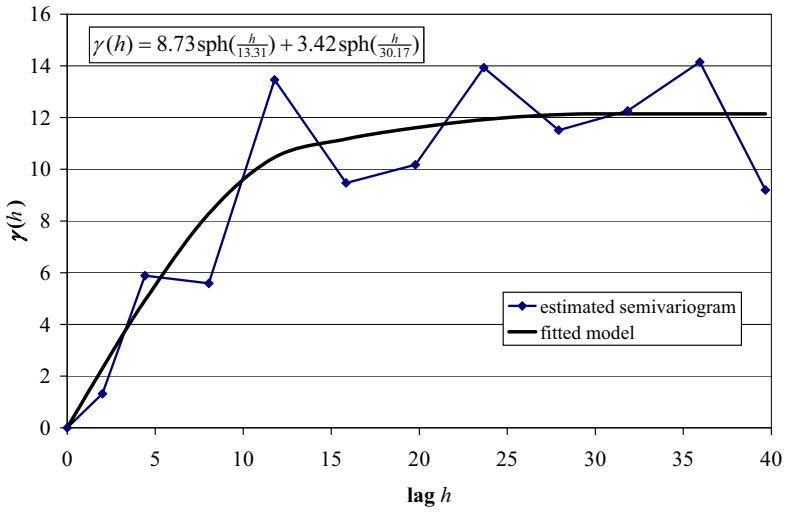
**4**. *Predict the value* $Z(\mathbf{x}_0)$

With the $\lambda_i$, the observed values $z(\mathbf{x}_i)$ and the estimated value of $\mu_Z$ in Equation (7.5) the unknown value of $Z(\mathbf{x}_0)$ is predicted

**5**. *Calculate the variance of the prediction error*

Using $\lambda_i(\mathbf{x}_0)$, $C_Z(\mathbf{h})$ and $\sigma_Z^2$ the variance of the prediction error is calculated with (7.11).

The result is a map of predicted properties on a grid and a map of associated error variances. Figure 7.8 shows the map of kriging predictions and the associated prediction variance or kriging variance:




F*igure* 7.7 *Semivariogram and fitted semivariance function of the* 140 *locations of the Walker lake data set* (*Figure* 2.1); *SPH*() *spherical model*.

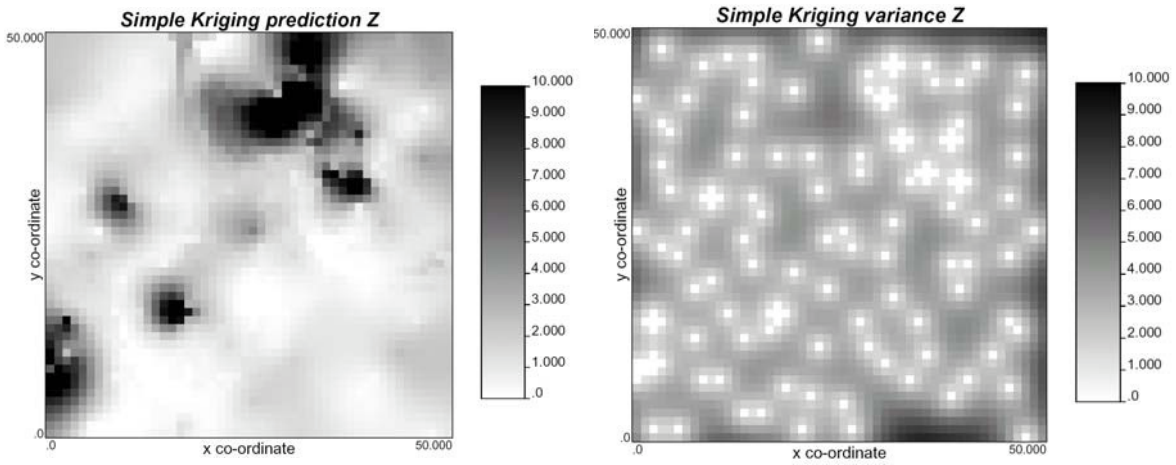



*Figure* 7.8 *Interpolation with simple kriging predictions and the associated kriging variance of theWalker lake data*

## 7.3.2 Ordinary kriging

### Theory

Ordinary kriging can be used if

1. $Z(\mathbf{x})$ is a wide sense stationary RF but the mean of $Z(\mathbf{x})$ is unknown, or
2. $Z(\mathbf{x})$ is an intrinsic RF.

An intrinsic RF has the following properties (see also chapter 5):

$$E[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)] = 0$$

$$E[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)]^2 = 2\gamma(\mathbf{x}_2 - \mathbf{x}_1) = 2\gamma(\mathbf{h})$$

The mean difference between the RVs at any two locations is zero (i.e. constant mean) and the variance of this difference is a function that only depends on the separation vector $\mathbf{h}$. The function $\gamma(\mathbf{h}) = \frac{1}{2}E[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2$ is the semivariogram.

The ordinary kriging predictor is a weighted average of the surrounding observations:

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) \qquad (7.14)$$

with $Z(\mathbf{x}_i)$ the values of $Z(\mathbf{x})$ at the observation locations (usually within a limited search neighbourhood). As with the simple kriging predictor we want (7.14) to be unbiased:

$$E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = E[\sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0)] =$$

$$\sum_{i=1}^{n} \lambda_i E[Z(\mathbf{x}_i)] - E[Z(\mathbf{x}_0)] = 0 \tag{7.15}$$

As the unknown mean is constant, i.e. $E[Z(\mathbf{x}_i)] = E[Z(\mathbf{x}_0)] \,\forall\, \mathbf{x}_i, \mathbf{x}_0$, we find the following "unbiasedness constraint" for the $\lambda_i$:

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{7.16}$$

Apart from being unbiased we also want to have a predictor with a minimum variance of the prediction error. The error variance for predictor (7.14) can be written in terms of the semivariance as (see for instance de Marsily (1986) for a complete derivation):

$$V[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]^2 =$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma_Z(\mathbf{x}_i - \mathbf{x}_j) + 2\sum_{i=1}^{n} \lambda_i \gamma_Z(\mathbf{x}_i - \mathbf{x}_0) \tag{7.17}$$

We want to minimize the error variance subject to the constraint (7.16). In other words, we want to find the set of values $\lambda_i, i = 1,..,n$ for which (7.17) is minimum without violating constraint (7.16). To find these, a mathematical trick is used. First the expression of the error variance is extended as follows:

$$E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]^2 =$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma_Z(\mathbf{x}_i - \mathbf{x}_j) + 2\sum_{i=1}^{n} \lambda_i \gamma_Z(\mathbf{x}_i - \mathbf{x}_0) - 2\nu\left[\sum_{i=1}^{n} \lambda_i - 1\right] \tag{7.18}$$

If the estimator is really unbiased, nothing has happened to the error variance as the added term is zero by definition. The dummy variable $\nu$ is called the *Lagrange multiplier*. It can be shown that if we find the set of $\lambda_i, i = 1,..,n$ and the value of $\nu$ for which (7.18) has its minimum value, we have also have the set of $\lambda_i(\mathbf{x}_0), i = 1,..,n$ for which the error variance of the ordinary kriging predictor is minimal, while at the same time $\sum \lambda_i = 1$. As with simple kriging, the minimum value is found by partial differentation of (7.18) with respect to $\lambda_i, i = 1,..,n$ and $\nu$ and equating the partial derivatives to zero. This results in the following system of $(n + 1)$ linear equations:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j \gamma_Z(\mathbf{x}_i - \mathbf{x}_j) + \nu = \gamma_Z(\mathbf{x}_i - \mathbf{x}_0) \qquad i = 1,..,n \\ &\sum_{i=1}^{n} \lambda_i = 1 \end{aligned} \tag{7.19}$$

Using the langrange multiplier, the value for the (minimum) variance of the prediction error can be conveniently written as:

$$V[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{i=1}^{n} \lambda_i \gamma_Z(\mathbf{x}_i - \mathbf{x}_0) + \nu \qquad (7.20)$$

A unique solution of the system (7.19) and a positive kriging variance is only ensured if thew semivariogram function is "conditionally non-negative definite". This means that for all possible $\mathbf{x}_1, ...\mathbf{x}_n \in \Re^N (N = 1,2 \text{ or } 3)$ and for all $\lambda_1, ..., \lambda_n \in \Re$ such that $\sum_i \lambda_i = 1$, the following inequality must hold:

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma_Z(\mathbf{x}_i - \mathbf{x}_j) \geq 0 \qquad (7.21)$$

This is ensured if one of the permissible semivariogram models (Table 7.2 ,see also chapter 5)) is fitted to the experimental semivariogram data.

*Table* 7.2 *Permisible semivariogram models for ordinary kriging*; *here h denotes the length of the lag vector h.*

| | | |
|---|---|---|
| (a) spherical model | $\gamma(h) = \begin{cases} c\left[\frac{3}{2}(\frac{h}{a}) - \frac{1}{2}(\frac{h}{a})^3\right] & \text{if } h \leq a \\ c & \text{if } h > a \end{cases}$ | |
| (b) exponential model | $\gamma(h) = c[1 - \exp(-h/a)]$ | |
| (c) Gaussian model | $\gamma(h) = c[1 - \exp(-[h/a]^2)]$ | |
| (d) nugget model | $\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ 1 & \text{if } h > 0 \end{cases}$ | |
| (e) power model | $\gamma(h) = ch^{\omega}$ | $0 < \omega < 2$ |

Models (a) to (d) are also permissible in cases the RF is wide sense stationary. The power model, which does not reach a sill can be used in case of an intrinsic RF but not in case of a wide sense stationary RF.

The unknown mean $\mu_Z$ and the langrange multiplier $\nu$ require some further explanation. If all the data are used to obtain predictions at every location, at all locations the same unknown mean $\mu_Z$ is implicitly estimated by the ordinary kriging predictor. The lagrange multiplier represents the additional uncertainty that is added to the kriging prediction by the fact that the mean is unknown and must be estimated. Therefore, if the RF is wide sense stationary, the variance of the prediction error for ordinary kriging is larger than that for simple kriging, the difference being the lagrange multiplier. This can be deduced from substituting in Equation(7.20) $\gamma_Z(\mathbf{h}) = \sigma_Z^2 - C_Z(\mathbf{h})$ and taking into account that $\sum \lambda_i = 1$. This means that,

whenever the mean is not exactly known and has to be estimated from the data it is better to use ordinary kriging, so that the added uncertainty about the mean is taken into account.
Even in simple kriging one rarely uses all data to obtain kriging predictions. Usually only a limited number of data close to the prediction location are used. This is to avoid that the kriging systems become too large and the mapping too slow. The most common way of selecting data is to center an area or volume at the prediction location $\mathbf{x}_0$. Usually the radius is taken about the size of the variogram range. A limited number of data points that fall within the search area are retained for the kriging prediction. This means that the number of data locations becomes a function of the prediction location: $n \equiv n(\mathbf{x}_0)$. Also, if ordinary kriging is used, a local mean is implicitly estimated that changes with $\mathbf{x}_0$. So we have $\mu \equiv \mu(\mathbf{x}_0)$ and $\nu \equiv \nu(\mathbf{x}_0)$ footnote . This shows that, apart from correcting for the uncertainty in the mean and being able to cope with a weaker form of stationarity, ordinary kriging has a third advantage when compared to simple kriging: even though the intrinsic hypothesis assumes that the mean is constant, using ordinary kriging with a search neighbourhood enables one to correct for local deviations in the mean. This makes the ordinary kriging predictor more robust to trends in the data than the simple kriging predictor.

Note:that for briefness of notation we will use $n$ and $\nu$ in the kriging equations, instead of $n(\mathbf{s}_0)$ and $\nu(\mathbf{s}_0)$. The reader should be aware that in most equations that follow, both the number of observations and the lagrange multipliers depend on the prediction location $\mathbf{s}_0$, except for those rare occasions that a global search neighbourhood is used.

In box 4 the ordinary kriging prediction is illustrated using the same example as Box 3. When compared to simple kriging it can be seen that the prediction is slightly different and that the prediction variance is larger.

## Practice

In practice ordinary kriging consists of the following steps (illustrated again with the Walker lake data set):

1. *Estimate the semivariogram*
2. *Fit a permissible semivariogram model*

For every node on the grid repeat:

3. *Solve the kriging equations*

   Using the fitted semivariogram model $\gamma_Z(\mathbf{h})$ in the $n+1$ linear equations (7.19) yields, after solving them, the kriging weights $\lambda_i, i = 1,..,n$ and the lagrange multiplier $\nu$.
4. *Predict the value $Z(\mathbf{x}_0)$*

   With the $\lambda_i$, the observed values $z(\mathbf{x}_i)$ (usually within the search neighbourhood) in equation (7.14) the unknown value of $Z(\mathbf{x}_0)$ is predicted.
5. *Calculate the variance of the prediction error*

   Using $\lambda_i$, $\gamma_Z(\mathbf{h})$ and $\nu$ the variance of the prediction error is calculated with (7.20).

The semivariogram was already shown in Figure 7.7. Figure 7.9 shows the ordinary kriging prediction and the ordinary kriging variance. Due to the large number of observations (140) there are no visual differences between Figure 7.9 and 7.8.

Box 4: Ordinary kriging example

$$\begin{aligned}
&\lambda_1\gamma(\mathbf{x}_1-\mathbf{x}_1)+\lambda_2\gamma(\mathbf{x}_1-\mathbf{x}_2)+\lambda_3\gamma(\mathbf{x}_1-\mathbf{x}_3)+\nu=\gamma(\mathbf{x}_1-\mathbf{x}_0)\\
&\lambda_1\gamma(\mathbf{x}_2-\mathbf{x}_1)+\lambda_2\gamma(\mathbf{x}_2-\mathbf{x}_2)+\lambda_3\gamma(\mathbf{x}_2-\mathbf{x}_3)+\nu=\gamma(\mathbf{x}_2-\mathbf{x}_0)\\
&\lambda_1\gamma(\mathbf{x}_3-\mathbf{x}_1)+\lambda_2\gamma(\mathbf{x}_3-\mathbf{x}_2)+\lambda_3\gamma(\mathbf{x}_3-\mathbf{x}_3)+\nu=\gamma(\mathbf{x}_3-\mathbf{x}_0)\\
&\lambda_1+\lambda_2+\lambda_3=1
\end{aligned}$$

$$\gamma(\mathbf{x}_i-\mathbf{x}_j)=22.69\left(1-e^{-\frac{|\mathbf{x}_i-\mathbf{x}_j|}{2}}\right)$$

$$\begin{aligned}
17.627\lambda_2+19.619\lambda_3+\nu&=17.628\\
16.627\lambda_1+8.930\lambda_3+\nu&=8.930\\
19.619\lambda_1+8.930\lambda_2+\nu&=8.930\\
\lambda_1+\lambda_2+\lambda_3&=1
\end{aligned}$$

$$\lambda_1=0.172\quad\lambda_2=0.381\quad\lambda_3=0.447\quad\nu=2.147$$

$$Z^*(\mathbf{x}_0)=0.172\cdot3+0.381\cdot7+0.447\cdot14=9.44$$

$$\begin{aligned}
&\mathrm{var}[Z^*(\mathbf{x}_0)-Z^*(\mathbf{x}_0)]=\\
&\quad 0.172\cdot17.627+0.381\cdot8.930+0.447\cdot8.930+2.147=12.57
\end{aligned}$$

A simple example of ordinary kriging. For spatial lay-out of the data points and the table of distances between locations we refer to Box 3. The kriging system is shown therafter, with the $\mathbf{x}_i$-$\mathbf{x}_j$ semivariances on the left and the $\mathbf{x}_i$-$\mathbf{x}_0$ semivariances on the right. Using the assumed mean and semivariance function shown, the next boxes show the numerical solution of the kriging equations and the evaluation of the kriging predictor and the kriging variance.

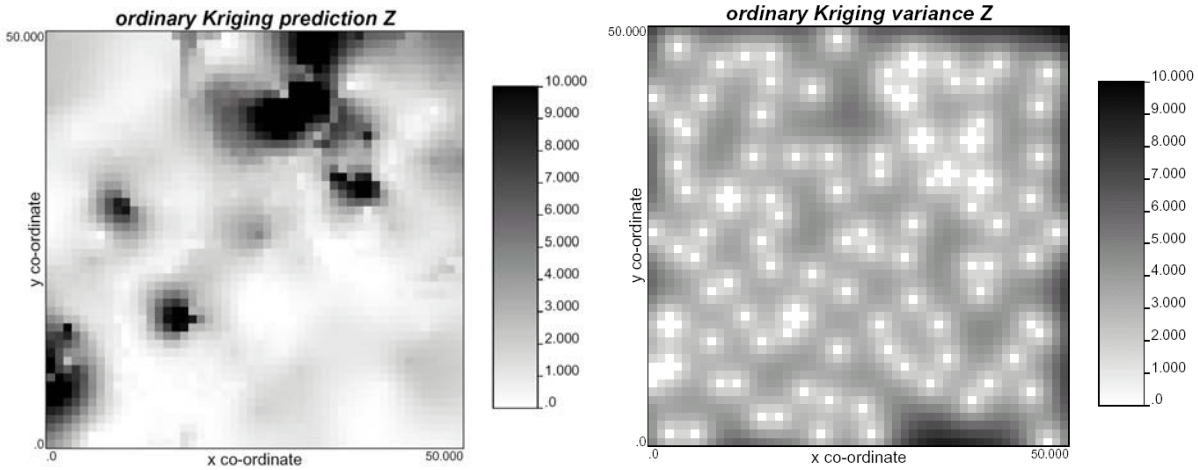



*Figure 7.9 Interpolation with simple kriging predictions and the associated kriging variance of theWalker lake data*

### 7.3.3 Block kriging

Up to known we have been concerned with predicting the values of attributes at the same *support* (averaging volume) as the observations, usually point support. However, in many cases one may be interested in the mean value of the attribute for some area or volume much larger than the support of the observations. For instance, one may be interested in the average porosity of a model block that is used in a numerical groundwater model, or the average precipation of a catchment. These average quantities can be predicted using block kriging. The term "block kriging" is used as opposed to "point kriging" or "punctual kriging" where attributes are predicted at the same support as the observations. Any form of kriging has a point form and a block form. So, there is simple point kriging and simple block kriging and ordinary point kriging and ordinary block kriging etc. Usually, the term "point" is ommited and the term "block" is added only if the block kriging form is used.
Consider the problem of predicting the mean $\overline{Z}$ of the attribute $z$ that varies with spatial co-ordinate $\mathbf{x}$ for some area or volume $D$ with size $|D|$ (length, area or volume):

$$\overline{Z} = \frac{1}{|D|}\int_{\mathbf{x}\in D} Z(\mathbf{x})d\mathbf{x}$$

In case $D$ is a block in three dimensions with upper an lower boundaries boundaries $x_l$, $y_l$, $z_l$, $x_u$, $y_u$, $z_u$ the spatial integral (7.22) stands for

$$\frac{1}{|D|}\int_{\mathbf{x}\in D} Z(\mathbf{x})d\mathbf{x} \equiv$$

$$\frac{1}{|x_u - x_l||y_u - y_l||z_u - z_l|}\int_{z_l}^{z_u}\int_{y_l}^{y_u}\int_{x_l}^{x_u} Z(s_1,s_2,s_3)ds_1 ds_2 ds_3 \qquad (7.23)$$

Of course, the block $D$ can be of any form, in which case a more complicated spatial integral is used (e.g. Figure 7.10 in two dimensions):

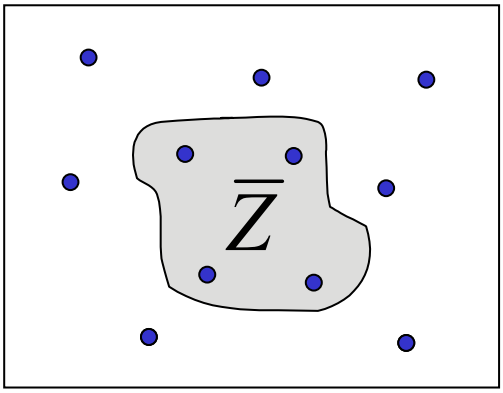



*Figure* 7.10 *Example of block kriging in two dimensions to predict the mean value of Z of some irregular area D*

Similar to point kriging, the unknown value of $\bar{Z}$ can be predicted as linear combination of the observations by assuming that the predictant and the observations are partial realizations of a RF. So, the ordinary bock kriging predictor becomes:

$$\hat{\bar{Z}} = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) \tag{7.24}$$

where the block kriging weights $\lambda_i$ are determined such that $\hat{\bar{Z}}$ is unbiased and the prediction variance $Var[\hat{\bar{Z}} - \bar{Z}]$ is minimal. This is achieved by solving the $\lambda_i$ from the ordinary block kriging system:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j \gamma_Z(\mathbf{x}_i - \mathbf{x}_j) + \nu = \gamma_Z(\mathbf{x}_i, D) \qquad i = 1,..,n \\ &\sum_{i=1}^{n} \lambda_i = 1 \end{aligned} \tag{7.25}$$

It can be seen that the ordinary block kriging system looks almost the same as the ordinary (point) kriging system, except for the term on the right hand side which is the average semivariance between an location $\mathbf{x}_i$ and all the locations inside the area of interest $D$:

$$\gamma_Z(\mathbf{x}_i, D) = \frac{1}{|D|} \int_{\mathbf{x} \in D} \gamma_Z(\mathbf{x}_i - \mathbf{x}) d\mathbf{x} \tag{7.26}$$

When building the block kriging system, the integral in equation (7.26) is usually not solved. Instead, it is approximated by first discretizing the area of interest in a limited number of points. Second, the semivariances are calculated between the observation location and the $N$ points $\mathbf{x}_j$ discretizing $D$ (see Figure 7.10 left figure). Third, the average semivariance is approximated by averaging these semivariances as:

$$\gamma_Z(\mathbf{x}_i, D) \simeq \frac{1}{N} \sum_{j=1}^{N} \gamma_Z(\mathbf{x}_i - \mathbf{x}_j) \tag{7.27}$$

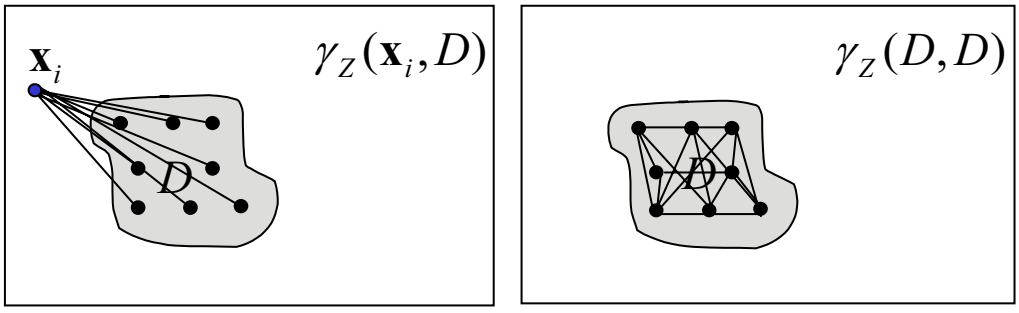



*Figure*7.11 *Numerical approximation of the spatial integrals* (7.26) (*left*) *and* (7.29) (*right*)

The variance of the prediction error is given by

$$Var[\hat{\bar{Z}} - \bar{Z}] = E[\hat{\bar{Z}} - \bar{Z}]^2 = \sum_{i=1}^{n} \lambda_i \gamma_Z(\mathbf{x}_i, D) + \nu - \gamma_Z(D,D) \tag{7.28}$$

where $\gamma_Z(D,D)$ is the average semivariance within the area $D$, i.e. the average semivariance between all locations with $D$:

$$\gamma_Z(D,D) = \frac{1}{|D|^2} \int_{\mathbf{x}_2 \in D} \int_{\mathbf{x}_1 \in D} \gamma(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{7.29}$$

which in practice is approximated by $N$ points $\mathbf{x}_i$ discretizing $D$ as (see also Figure 7.11, right figure)

$$\gamma_Z(D,D) \simeq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma(\mathbf{x}_i - \mathbf{x}_j) \tag{7.30}$$

Figure 7.12 shows the result of block kriging applied to the Walker lake data set with block sizes of 5 × 5 units.

Here we have given the equations for ordinary block kriging. The simple block kriging equations can be deduced in a similar manner from the simple kriging equations (7.10) by replacing the covariance on the right hand side by the point-block covariance $C_Z(\mathbf{x}_i, D)$. The prediction error variance is given by (7.11) with $\sigma_Z^2$ replaced by the within block variance $C_Z(D,D)$ (the average covariance of points within $D$) and $C_Z(\mathbf{x}_i - \mathbf{x}_0)$ by $C_Z(\mathbf{x}_i, D)$. The point-block covariance and the within block covariance are defined as in Equations (7.26) and (7.29) with $\gamma_Z(\mathbf{x}_1 - \mathbf{x}_2)$ replaced by $C_Z(\mathbf{x}_1 - \mathbf{x}_2)$.

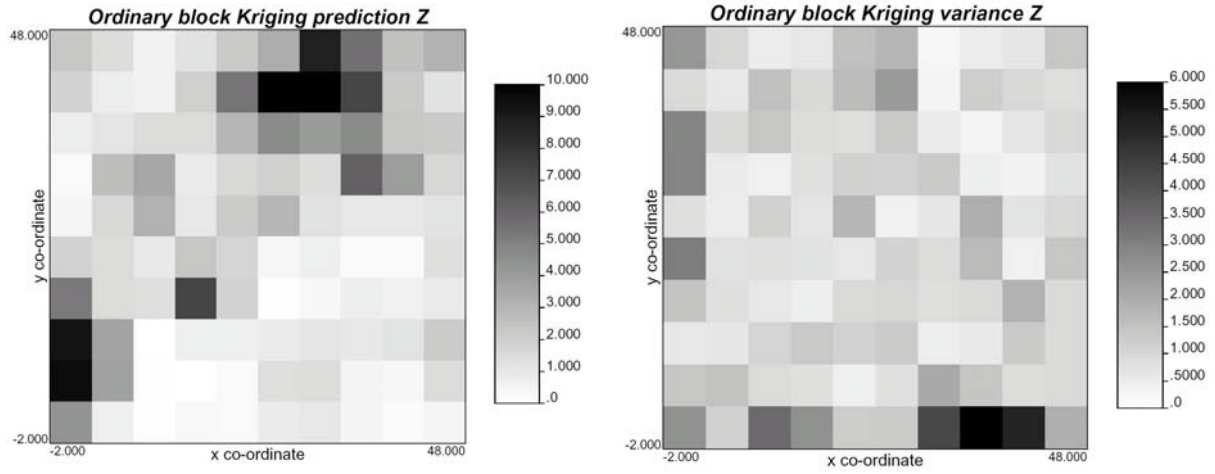



*Figure* 7.12 *Block kriging applied to the Walker lake data set with block sizes of* 5 × 5 *units*

## 7.4 Estimating the local conditional distribution

Kriging can also be used to estimate for each non-observed location the probability distribution $f_z(z;\mathbf{x} \mid z(\mathbf{x}_i), i = 1,..,n)$, i.e the probability distribution given the observed values at the observation locations. Let us return to Figure 5.8. This figure shows conditional random functions. Each realisation is conditioned by the observations, i.e. it passes through the observed value, but is free to vary between observations. The farther away from an observation, the more the realisations differ. This is reflected by the conditional pdf $f_z(z;\mathbf{x} \mid z(\mathbf{x}_i), i = 1,..,n)$ at a given location (two of which are shown in Figure 5.8). The farther away from an observation, the larger the variance of the conditional pdf, which means the more uncertain we are about the actual but unknown value $z(\mathbf{x})$. In the following sections methods are shown that can be used to estimate the conditional pdf $f_z(z;\mathbf{x} \mid z(\mathbf{x}_i), i = 1,..,n)$ through kriging.

### 7.4.1 Multivariate Gaussian random functions

If, apart from being wide sense stationary, the RSF is also multivariate Gaussian distributed then we have:

- The kriging error is Gaussian distributed with mean zero and variance equal to the simple kriging variance $\sigma^2_{SK}(\mathbf{x}_0) = \mathrm{VAR}\{\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)\}$. A 95%-prediction interval would then be given by $[\hat{z}_{SK}(\mathbf{x}_0) - 2\sigma_{SK}(\mathbf{x}_0),\ \hat{z}_{SK}(\mathbf{x}_0) + 2\sigma_{SK}(\mathbf{x}_0)]$, where $\hat{z}_{SK}(\mathbf{x}_0)$ is the simple kriging prediction.
- The conditional cumulative probability distribution function (ccpdf) is Gaussian with mean equal to the simple kriging prediction $\hat{z}_{SK}(\mathbf{x}_0)$ (the dashed line in Figure 5.8) and variance equal to the variance of the simple kriging prediction error $\sigma^2_{SK}(\mathbf{x}_0)$ (the variance over the realisations shown in Figure 5.8):

$$F_{z|z_1..z_n}(z;\mathbf{x}_0 \,|\, z(\mathbf{x}_i), i = 1,..,n) =$$

$$\frac{1}{\sqrt{2\pi\sigma^2_{SK}(\mathbf{x}_0)}} \int_{-\infty}^{z} \exp\left( \frac{-[z - \hat{z}_{SK}(\mathbf{x}_0)]^{\,2}}{\sigma^2_{SK}(\mathbf{x}_0)} \right) dz \qquad (7.31)$$

where $z(\mathbf{x}_1),....,z(\mathbf{x}_n)$ are the observed values at locations $\mathbf{x}_1,....,\mathbf{x}_n$ respectively. The second property is very convenient and the reason why the multivariate Gaussian and stationary RSF is very popular model in geostatistics. After performing simple kriging predictions, one is able to give an estimate of the ccpdf of $Z(\mathbf{x})$ exceeding a given threshold for every location in the domain of interest. For instance, if $Z(\mathbf{x})$ is a concentration of some pollutant in the groundwater and $z_c$ is critical threshold above which the pollutant becomes a health hazard,

simple kriging and Equation (7.31) can be used to map the probability of exceeding this threshold, given the concentrations found at the observation locations. Instead of delineating a single plume based upon some predicted value, several alternative plumes can be delineated, depending on which probability contour is taken as its boundary. This way, both the observed concentration values as well as the local uncertainty are taken into account when mapping the plume. Also, the risk of not treating hazardous groundwater can be weighted against the costs of remediation. For instance, if the risk of not treating hazardous groundwater should be smaller than 5 %, all the water within the 0.05 contour should be treated. Obsviously this results in much higher costs then if, for instance, a 10% risk is deemed acceptable. For a more elaborate discussion about probability distributions and the trade off between risk and costs, we refer to Goovaerts (1997, section 7.4).

### 7.4.2 Log-normal kriging

Many geophysical variables, such as hydraulic conductivity and pollutant concentration, are approximately log-normal distributed. A frequently used RSF model to decribe these variables is the multivariate logGaussian distribition. If $Z(\mathbf{x})$ is multivariate lognormal distributed, the natural logarithm $Y(\mathbf{x}) = \ln(Z(\mathbf{x}))$ is multivariate Gaussian distributed. Log-normal kriging then consists of the following steps:

1. Transform the observations $z(\mathbf{x}_i)$ by taking their logarithms $y(\mathbf{x}_i) = \ln(z(\mathbf{x}_i))$.
2. Estimate the semivariogram $\hat{\gamma}_Y(\mathbf{h})$ from the logtransformed data $y(\mathbf{x}_i)$ and fit a permissible model (note: that mean $m_Y$ must be determined and assumed known if simple kriging is used).
3. Using the semivariogram $\gamma_Y(\mathbf{h})$, the data $y(\mathbf{x}_i)$ (and the mean $m_Y$ in case of simple kriging), the kriging equations are solved to obtain at every non-observed location $\mathbf{x}_0$ the prediction $\hat{Y}_{SK}(\mathbf{x}_0)$ and prediction error variance $\sigma^2_{YSK}(\mathbf{x}_0)$ in cased of simple kriging or $\hat{Y}_{OK}(\mathbf{x}_0)$, $\sigma^2_{YOK}(\mathbf{x}_0)$ in case of ordinary kriging.
4. An unbiased prediction of $Z(\mathbf{x}_0)$ is obtained by the following backtransforms footnote :
   for simple kriging:

$$\hat{Z}(\mathbf{x}_0) = \exp(\hat{Y}_{SK}(\mathbf{x}_0) + \frac{1}{2}\sigma^2_{YSK}) \tag{7.32}$$

   and for ordinary kriging:

$$\hat{Z}(\mathbf{x}_0) = \exp(\hat{Y}_{OK}(\mathbf{x}_0) + \frac{1}{2}\sigma^2_{YOK} - \nu_Y) \tag{7.33}$$

   where $\nu_Y$ is the lagrange multiplier used in the ordinary kriging system.
5. If $Y(\mathbf{x})$ is multivariate Gaussian distributed and stationary, the ccpdf can be calculated from the simple kriging prediction $\hat{y}_{SK}(\mathbf{x}_0)$ and prediction variance as:

$$F_{z|z_{1..}z_n}(z;\mathbf{x}_0\,|\,z(\mathbf{x}_i),i=1,..,n) = \frac{1}{\sqrt{2\pi\sigma^2_{YSK}(\mathbf{x}_0)}}\int_{-\infty}^{z}\exp\left(\frac{-[\ln(z)-\hat{y}_{SK}(\mathbf{x}_0)]^{\,2}}{\sigma^2_{YSK}(\mathbf{x}_0)}\right)dz \tag{7.34}$$

An additional reason why in many geostatistical studies the observations are logtransformed before kriging is that the semivariogram of logtransorms can be better estimated (shows less noise) because of the imposed variance reduction.

### 7.4.3 Kriging normal-score transforms

An even more general transformation is the normal-score transform using the histogram. Through this transform, it is possible to transform any set of observations to univariate Gaussian distributed variables, regardless of the distribution of these observations. A normal score-transform proceeds as follows:

**1**. The $n$ observations are ranked in ascending order:

$$z(\mathbf{x}_i)^{(1)} \leq z(\mathbf{x}_j)^{(2)} \leq .. \leq z(\mathbf{x}_k)^{(r)} \leq .. \leq z(\mathbf{x}_l)^{(n)}$$

where $r = 1,..,n$ are the ranks of the observations.

**2**. The cumulative probability associated with observation $z(\mathbf{x}_k) = z_k$ with rank $r$ is estimated as:

$$\hat{F}(z_k) = r(z_k)/(n+1). \tag{7.35}$$

or in case of declustering

$$\hat{F}(z_k) = \sum_{i=1}^{r(z_k)} w_{r(z_k)}. \tag{7.36}$$

**3**. The associated normal score transform is given by the $\hat{p}_r$-quantile of the standard normal distribution:

$$y_{\text{ns}}(z_k(\mathbf{x}_k)) \;\; = N^{-1}[\hat{F}\{z_k(\mathbf{x}_k)\}] \tag{7.37}$$

where $N(z)$ is the standard Gaussian cumulative distribution function and $N^{-1}(p)$ its inverse.

Figure 7.13 shows graphically how the normal-score transform works. The left figure shows the estimated cumulative distribution of the original (non-transformed) data and the right figure the standard Gaussian cumulative distribution. The dashed lines show how the observed

values $z_k$ are transformed into the normal-score transform $y_{\text{ns}}(z_k)$.

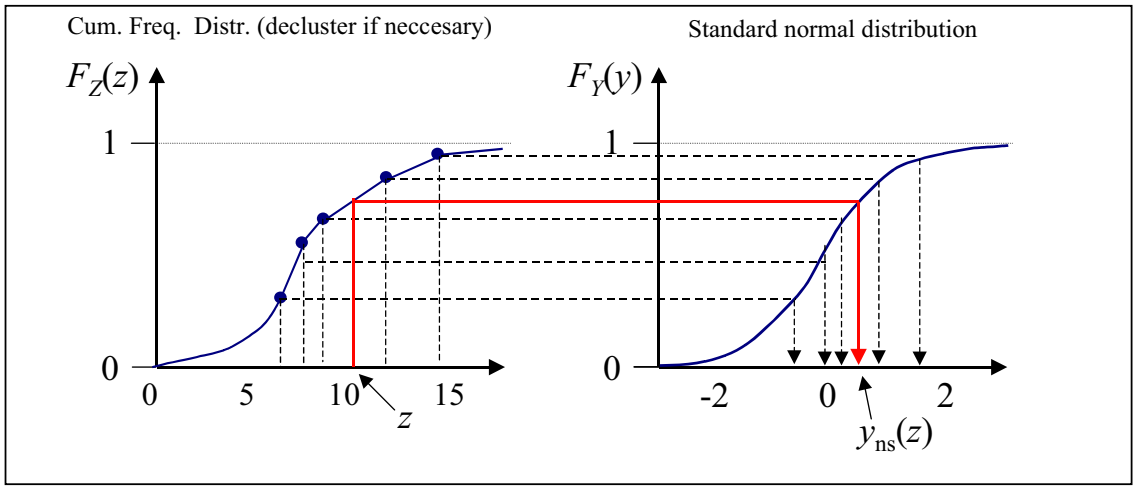



*Figure* 7.13 *Normal score transformation*

If we assume that the normal-score transforms are stationary and multivariate Gaussian distributed (see Goovaerts, 1997 for suggestions how to check this), the local ccpdfs can be obtained through simple kriging as follows:

1. Perform a normal score transform of the observations such as decribed above.
2. Estimate the semivariogram of the normal-score transformed data $y_{\text{ns}}(\mathbf{x}_k) = y_{\text{ns}}(z(\mathbf{x}_k))$ and fit a permissible semivariogram model $\gamma_Y(\mathbf{h})$. By definition, the mean value of the tranformed RSF $Y_{\text{ns}}(\mathbf{x})$ is zero.
3. Use the fitted semivariogram model and the normal-score transforms $y_{\text{ns}}(\mathbf{x}_k)$ in the simple kriging equations (with $m_Y = 0$) to obtain the prediction $\hat{Y}_{SK}(\mathbf{x}_0)$ and the associated prediction error variance $\sigma^2_{YSK}(\mathbf{x}_0)$.
4. The local ccpdf is then given by

$$F_{z|z_{1..z_n}}(z;\mathbf{x}_0\,|z(\mathbf{x}_i), i = 1,..,n) = \Pr[G\{\hat{y}_{SK}(\mathbf{x}_0), \sigma_{YSK}(\mathbf{x}_0)\} \leq y_{\text{ns}}(z)] = \frac{1}{\sqrt{2\pi\sigma^2_{YSK}(\mathbf{x}_0)}} \int_{-\infty}^{z} \exp\left( \frac{-[y_{\text{ns}}(z)] - \hat{y}_{SK}(\mathbf{x}_0)]^2}{\sigma^2_{YSK}(\mathbf{x}_0)} \right) dz \qquad (7.38)$$

where $y_{\text{ns}}(z)$ is the normal-score transform of the value $z$ and $G\{\mu,\sigma\}$ a Gaussian variate with mean $\mu$ and variance $\sigma$.

This is also shown graphically in Figure 7.13. Suppose we want to known for the non-observed location the probability that $Z(\mathbf{x}_0) \leq z$. We first obtain through the transformation the value of $y_{\text{ns}}(z)$. From the simple kriging of transformed data we have at $\mathbf{x}_0$: $\hat{y}_{SK}(\mathbf{x}_0)$ and $\sigma^2_{YSK}(\mathbf{x}_0)$. Finally, we evaluate $\Pr[G\{\hat{y}_{SK}(\mathbf{x}_0), \sigma_{YSK}(\mathbf{x}_0)\} \leq y_{\text{ns}}(z)]$ (Equation 7.38) to obtain the probability.

To calculate the normal-score transform of any given value $z$ (which is not necessarily equal to the value of one of the observations), the resolution of the estimated cpdf $\hat{F}(z)$ must be

increased. Usually, a linear interpolation is used to estimate the values of $\hat{F}(z)$ between two observations (see Figure7.13). Of course, most critical is the extrapolation that must be performed to obtain the lower and upper tails of $\hat{F}(z)$. For instance, if the upper tail of $\hat{F}(z)$ rises too quickly to 1, the probability of high values of $z$ (e.g. a pollutant in groundwater) may be underestimated. Usually a power model is used to extrapolate the lower tail and a hyperbolic model to extrapolate the upper tail. Several models for interpolating between quantiles, as well as rules of thumb about which model to use, are given in Deutsch and Journel (1998) and Goovaerts (1997).

This section is concluded by application of the normal score transform to the Walker lake data set. Figure 7.14 shows the histogram and the semivariogram of of the normal score transforms. It can be seen that semivariogram is less noisy than that of the non-transformed data (Figure 7.7), because transformation decreases the effect of the very large values. The simple kriging predictions and associated variances are shown in Figure 7.15. Figure 7.16 shows the probability that the $z$ exceeds the value of 5 and 10. If these were critical values and the Walker lake data groundwater concentrations Figure 7.16 shows the effect of the critical concentration on the probability of exceeding and through this on the area that must be cleaned up.

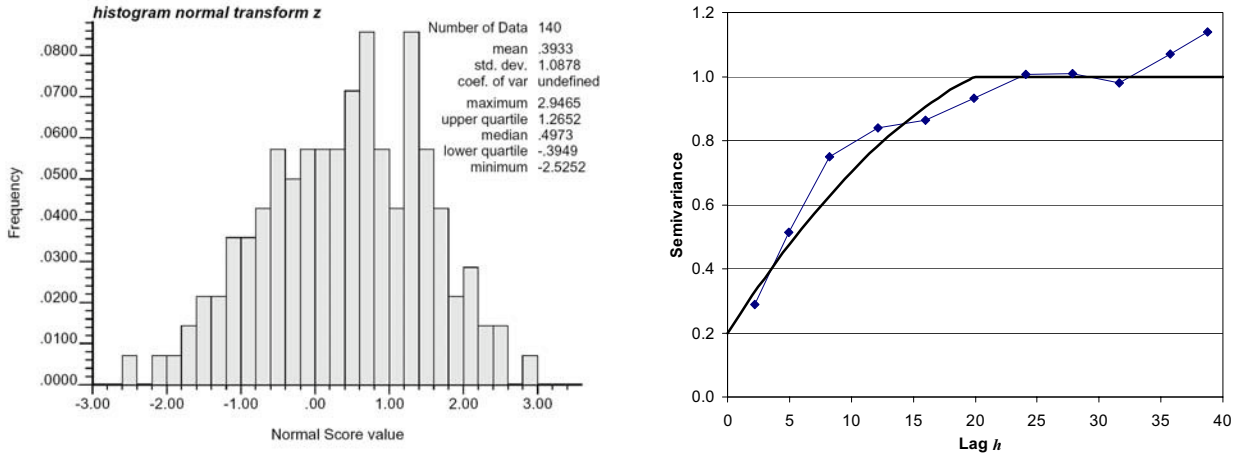



*Figure* 7.14 *Histogram and semivariogram of normal score transforms of the Walker lake data set*; *fitted semivariogram model*: $\gamma(h) = 0.2\text{Nug}(h) + 0.8\text{Sph}(h/19.9)$

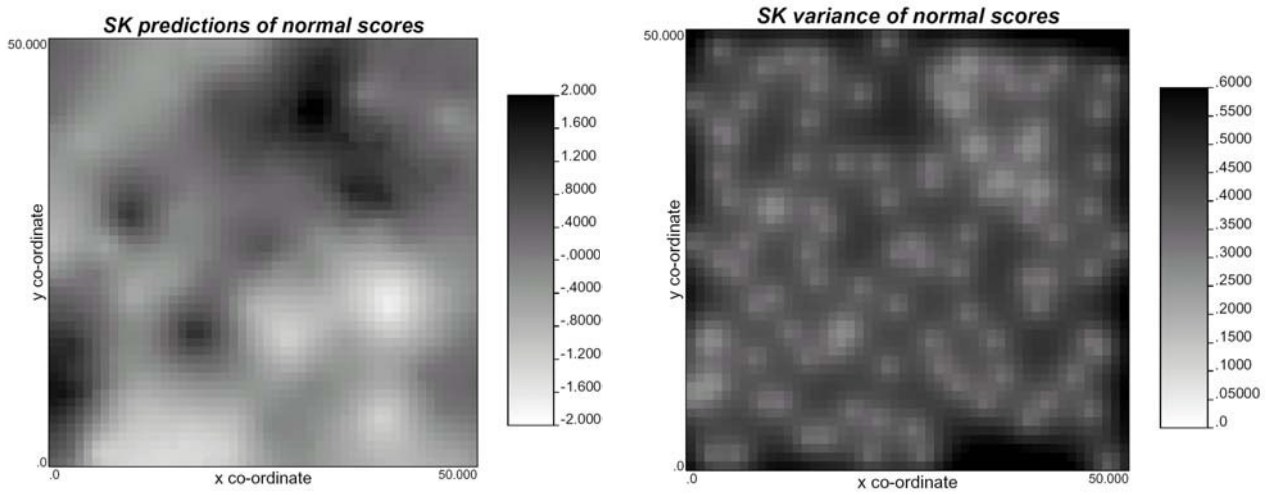



*Figure* 7.15 *Simple kriging results of normal score transforms of the Walker lake data set*

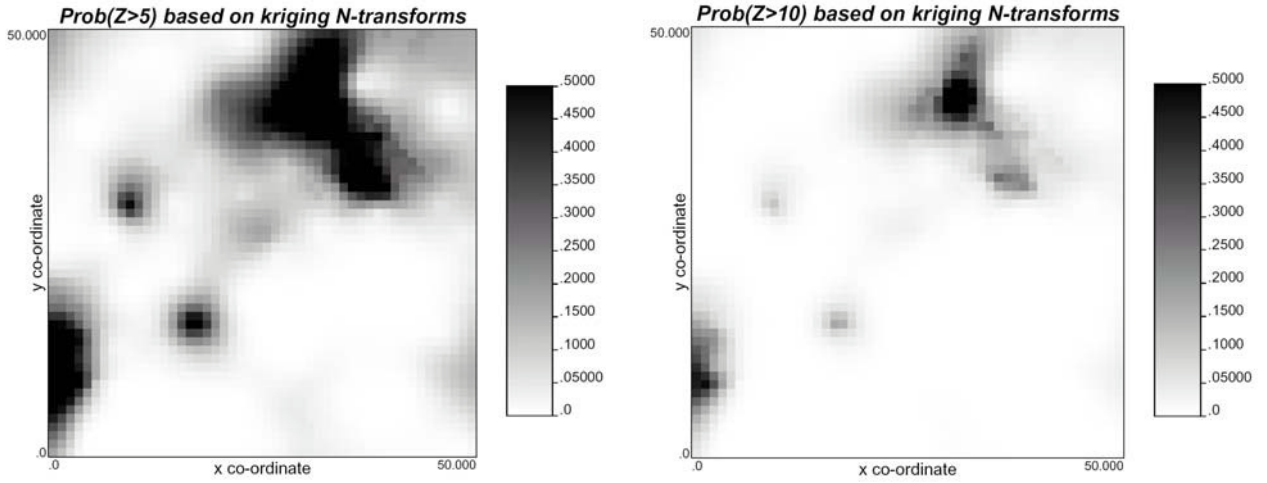



*Figure* 7.16 *Probability of exceeding* 5 *and* 10 *based on normal score simple kriging of the Walker lake data set*

## 7.5 Geostatistical simulation

The third field of application of geostatistics is simulating realisations of the conditional random function $Z(\mathbf{x}|z(\mathbf{x}_i), i = 1,..,n)$. Returning to Figure 5.8: in case of a wide sense stationary and multiGaussian RSF $Z(\mathbf{x})$, simple kriging provides the dashed line, which is the mean of all possible conditional realisations. The aim of *geostatistical simulation* is to generate in the individual conditional realisations. There are two important reasons why sometimes individual realisations of the conditional RSF are preferred over the interpolated map that is provided by kriging:

**1**. kriging provides a so called best linear prediction (it produces values that minimize the variance of the prediction error: $Var[\widehat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)]$), but the resulting maps are much

smoother than reality. This can again be seen from Figure 5.8. The individual realisations are very noisy and rugged while the kriging prediction produces a smoothly varying surface.The noisy realisations have a semivariogram that resembles that of the data, so one can say that the real variation of the property considered is much more like the realisations than the kriging map. This has repercussions if the kriging map is not the end point of the analysis (such as mapping concentrations). For instance, suppose that the goal is to produce a map of hydraulic conductivities that is to be used in a groundwater flow model. To use the kriged map as input in the groundwater flow model would produce flow lines that are probably too smooth also. Especially if the goal is to model groundwater transport, a smooth map of hydraulic conductivity will yield an underestimation of solute spreading. In that case it is better use realisations of the random function as input. Of course, as each realisation has equal probability to be drawn and is therefore an equally viable picture of reality, the question remains: which realisation should then be used? The answer is: not a single realisation should be analysed, but a great number of realisations. This conclusion brings us to the second reason why realisations are often preferred over kriging maps;

**2**. multiple realisations as input for a model can be used for *uncertainty analysis* and *ensemble prediction*. Figure 5.8 shows that usually we only have limited information about reality and we therefore represent our uncertainty about reality with a random function (see also chapters 1 and 5). Returning to the example of hydraulic conductivity, if we are uncertain about the parameters of a groundwater model, we also want to know what the uncertainty is about the model output (heads, fluxes). So instead of analysing a single input of hydraulic conductivity, a large number of conditional realisations of hydraulic conductivity (say 1000) are used as model input. If we use 1000 conditional realisations of hydraulic conductivity as input, we also have 1000 model runs with the groundwater model, producing (in case of a steady state groundwater model) 1000 head fields and 1000 flow fields. From this, it is possible to estimate the probability of hydraulic head at each location, or the probability that a contaminant plume reaches certain sensitive area. This way of modelling is really stochastic modelling, and because we do not produce one prediction, but an ensemble of predictions it is often referred to as *ensemble prediction*. The variance of the output realisations is a measure of our uncertainty about the output (e.g. hydraulic heads) that is caused by our uncertainty (lack of perfect knowledge) about the model parameters (e.g. hydraulic conductivity). So, through this way of stochastic modelling one performs an *uncertainty analysis*: estimating the uncertainty about model output that is caused by uncertainty about model input or model parameters. There are several ways of performing such an analysis, as will be shown extensively in chapter 8. The method described here, i.e. generating realisations of parameters or input variables and analysing them with a numerical model, is called *Monte Carlo simulation*. In Figure 7.17 the method of Monte carlo Simulation for uncertainty analysis is shown schematically.

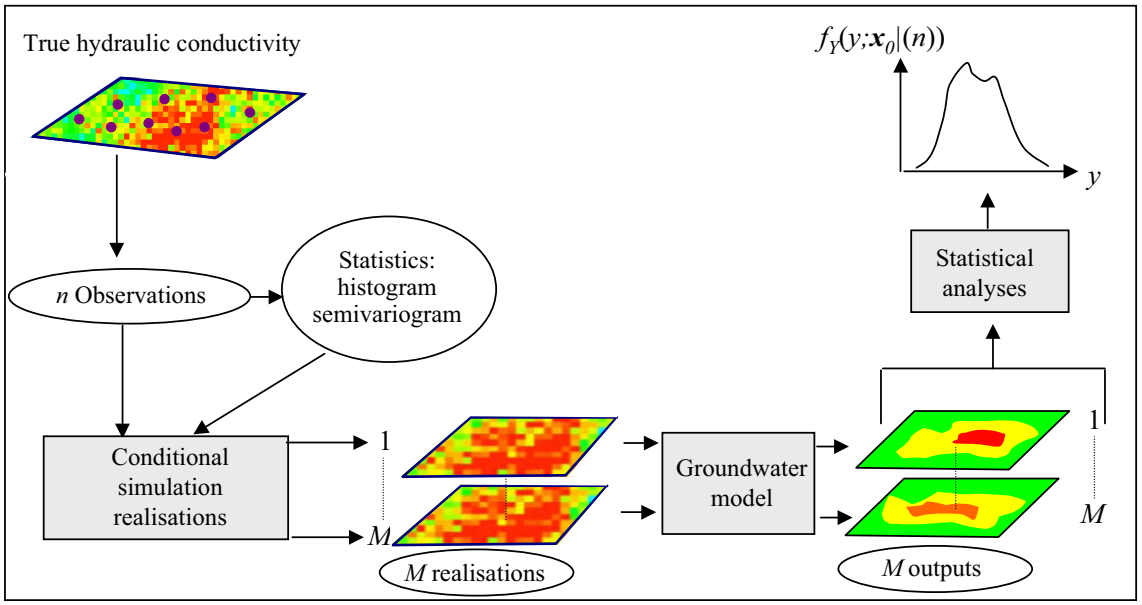



*Figure* 7.17 *Schematic representation of Monte Carlo simulation applied for uncertainty analysis of hydraulic conductivity in groundwater modelling. Hydraulic conductivity is spatially varying and sampled at a limited number of locations. Hydraulic conductivity is modelled as a random space function. Using the observations statistics are estimated that characterise this function* (*histogram*, *senivariogram*). *Next M realisations of this random function are simulated and used in the groundwater model. This yields M realisations of groundwater model output* (*e.g. head fields*). *From these realisations it is possible to obtain for a given location* (*e.g.* $\mathbf{x}_0$) *the probability density function of the ouput variables* (*e.g. head*, *concentration*).

The technique of Monte Carlo simulation is further explained in the next chapter. Here, we focus only on the generation of multiple realisations of the conditional random space function, commonly referred to as (conditional) *geostatistical simulation*. There are quite a few methods for simulating realisations of MultiGaussian random space functions. The most commonly used are LU-decomposition (Alabert, 1987), the turning band method (Mantoglou and Wilson, 1982) and Sequential Gaussian simulation (Goméz-Hernández and Journel, 1993), while there are even more methods for simulating non-Gaussian random functions (e.g. Amstrong and Dowd, 1994). The most flexible simulation algorithm and mostly used nowadays is sequential simulation. Sequential Gaussian simulation (sGs) will be treated here briefly. For a more elaborate description of the method one is referred to Goovaerts (1997) and Deutsch and Journel (1998).

Conditional simulation with sGs needs the mean $\mu_Z$ and the semivariogram of the random space function $\gamma_Z(h)$ and proceeds as follows.

1. The area is divided into a finite number of grid points $N$ (location indices $\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_N$) at which values of the conditional realisations are to be simulated. The grid points are visited in a random order.
2. For the first grid point $\mathbf{x}_1$ a simple kriging is performed from the given data $z(\mathbf{s}_1), .., z(\mathbf{s}_n)$ yielding the prediction $\hat{Z}_{SK}(\mathbf{x}_1)$ and the prediction variance $\sigma^2_{SK}(\mathbf{x}_1)$. Under the assumption

that $Z(\mathbf{x})$ is stationary and multiGaussian the conditional cumulative distribution is Gaussian:

$$F_Z(z, \mathbf{x}_1 | z(\mathbf{s}_1), .., z(\mathbf{s}_n)) \; = N\{z; \hat{Z}_{SK}(\mathbf{x}_1), \sigma_{SK}(\mathbf{x}_1)\} \tag{7.39}$$

.

3. A random value $P$ between zero and one is drawn from a uniform distribution $U[0,1]$. Using the inverse of the conditional distribution (7.39) the random quantile $P$ is used to draw a random value $Z$:

$$Z(\mathbf{x}_1) \; = N^{-1}\{P; \hat{Z}_{SK}(\mathbf{x}_1), \sigma_{SK}(\mathbf{x}_1)\} \tag{7.40}$$

4. For the second grid point $\mathbf{x}_2$ a simple kriging is performed using the data $z(\mathbf{s}_1), .., z(\mathbf{s}_n)$ and the previously simulated value $z(\mathbf{x}_1)$ in the kriging equations (so the previously simulated value is now treated as a data point). This yields the prediction $\hat{Z}_{SK}(\mathbf{x}_2)$ and the prediction variance $\sigma^2_{SK}(\mathbf{x}_2)$ from which the conditional cumulative distribution $F_Z(z, \mathbf{x}_2 | z(\mathbf{x}_1), z(\mathbf{s}_1), .., z(\mathbf{s}_n)) \; = N\{z; \hat{Z}_{SK}(\mathbf{x}_2), \sigma_{SK}(\mathbf{x}_2)\}$ is build.
5. A random value $P$ between zero and one is drawn from a uniform distribution $U[0,1]$ and using the inverse of the conditional distribution $N^{-1}\{P; \hat{Z}_{SK}(\mathbf{x}_2), \sigma_{SK}(\mathbf{x}_2)\}$ the random quantile $P$ is used to draw a random value $Z(\mathbf{x}_2)$.
6. For the third grid point $\mathbf{x}_3$ a simple kriging is performed using the data $z(\mathbf{s}_1), .., z(\mathbf{s}_n)$ and the previously simulated values $z(\mathbf{x}_1), z(\mathbf{x}_2)$ in the kriging equations yielding $F_Z(z, \mathbf{x}_3 | z(\mathbf{x}_1), z(\mathbf{x}_2), z(\mathbf{s}_1), .., z(\mathbf{s}_n)) \; = N\{z; \hat{Z}_{SK}(\mathbf{x}_3), \sigma_{SK}(\mathbf{x}_3)\}$.
7. Using a random value $P$ drawn from a uniform distribution $U[0,1]$ the random variable $Z(\mathbf{x}_3)$ is drawn and added to the data set.
8. Steps 6 and 7 are repeated adding more and more simulated values to the conditioning data set until all values on the grid have been simulated: the last simple kriging exercise thus yields the conditional probability $F_Z(z, \mathbf{x}_N | z(\mathbf{x}_1), ..., z(\mathbf{x}_N), z(\mathbf{s}_1), .., z(\mathbf{s}_n))$.

It can be shown heuristically that by construction this procedure produces a draw (realisation) from the multivariate conditional distribution $F_Z(z(\mathbf{x}_1), ..., z(\mathbf{x}_N) | z(\mathbf{s}_1), .., z(\mathbf{s}_n))$ (Goméz-Hernández and Journel, 1993; Goovaerts, 1997), i.e. a realisation from the conditional randum function $Z((\mathbf{x}) | z(\mathbf{s}_1), .., z(\mathbf{s}_n))$. To simulate another realisation the above procedure is repeated using a different random path over the grid nodes and drawing different random numbers for the quantiles $P \sim U[0,1]$. Unconditional realisations of the random function $Z(\mathbf{x})$ can also be simulated by starting at the first grid point with a draw from the Gaussian distribution $N\{z; \mu_Z, \sigma_Z\}$ and conditioning at every step on previously simulated points only. Obviously, the number of conditioning points and thus the size of the kriging system to be solved increases as the simulation proceeds. This would lead to unacceptably large computer storage requirements and computation times. To avoid this, a search area is used, usually with a radius equal to the semivariogram range, while only a limited number of observations and previously simulated points in the search radius are used in the kriging system (Deustch and Journel, 1998).

Obviously, the assumption underlying the simulation algorithm is that the RSF $Z(\mathbf{x})$ is stationary and multiGaussian. For a RSF to be multiGaussian it should at least have a univariate Gaussian distribution $f_Z(\mathbf{x}) = N\{z; \mu_Z, \sigma_Z\}$. So, if this method is applied, for instance, to the Walker-lake data set, a normal score transformation is required. The simulation procedure for a realisation of $Z((\mathbf{x})|z(\mathbf{s}_1),..,z(\mathbf{s}_n))$ would then involve the following steps:

1. Perform a normal score transform of the observations $.y_{\text{ns}}(z_k(\mathbf{s}_i)) = N^{-1}[\hat{F}\{z_k(\mathbf{x}_i)\}]$ (see Figure 7.13).
2. Estimate the semivariogram of the normal-score transformed data $y_{\text{ns}}(\mathbf{x}_i)$ and fit a permissible semivariogram model $\gamma_Y(\mathbf{h})$. By definition, the mean value of the tranformed RSF $Y_{\text{ns}}(\mathbf{x})$ is zero.
3. Assuming $Y_{\text{ns}}(\mathbf{x})$ to be stationary and multiGaussian, simulate a realisation of the conditional random function $Y_{\text{ns}}(\mathbf{x}|y_{\text{ns}}(\mathbf{x}_1),..,y_{\text{ns}}(\mathbf{x}_N))$ using sequential Gaussian simulation.
4. Back-transform the simulated values $(z(\mathbf{x}) = \hat{F}^{-1}[N\{y_{\text{ns}}(z_k(\mathbf{x}))\}]$, i.e.reversing the arrows in Figure 7.13) to obtain a realisation of the conditional random function $Z((\mathbf{x})|z(\mathbf{s}_1),..,z(\mathbf{s}_n))$.

In the geostatistical toolbox of Deutsch and Journel (1998) the simulation program sgsim performs the normal score transform, sequential simulation and the back transform all together. The parameters of the semivariogram of transforms $\gamma_Y(\mathbf{h})$ have to be provided separately. Figure 7.18 shows two realisations of the conditional random function based on the Walker lake data.

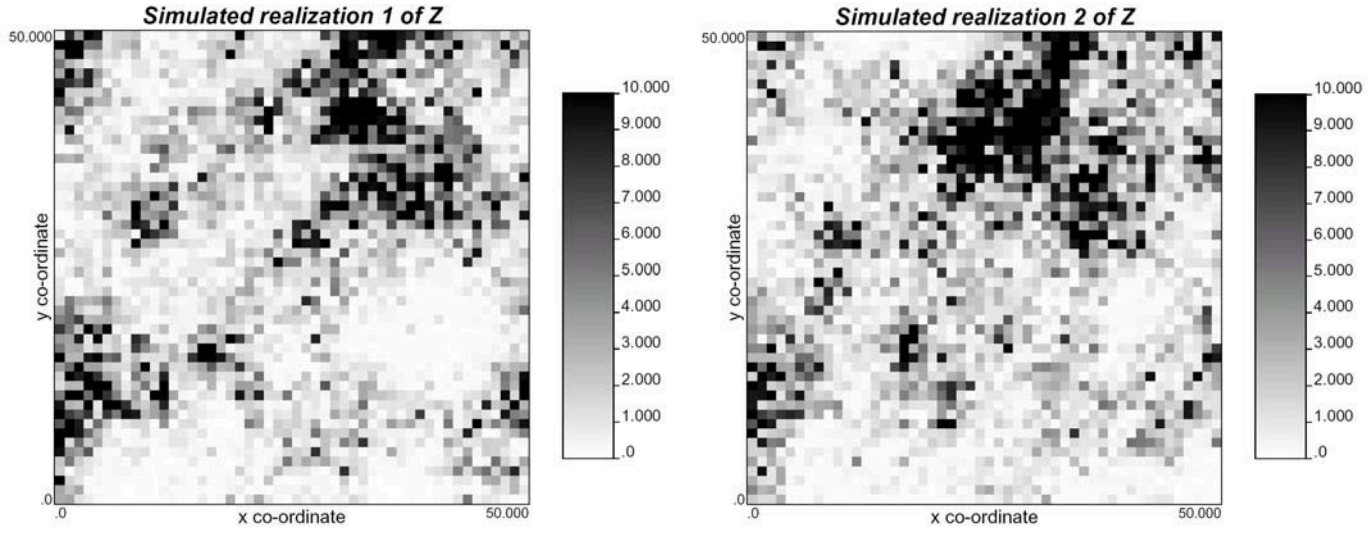



*Figure*7.18 *Two simulated realisations of a conditional random function based on the Walker lake data set*

## 7.6 More geostatistics

In this chapter the basic geostatistical methods have been presented. Naturally, the area of geostatistics is much more extensive. More advanced geostatistical methods are presented in various textbooks, such as that of Cressie (1993), Rivoirard (1994), Goovaerts (1997), Chilès and Delfiner (1999), and Christakos (2000). More advanced geostatistical methods are concerned with:

- kriging in case of non-stationary random functions;
- kriging using auxiliary information;
- estimating conditional probabilities of non-Gaussian random functions;
- simulating realisations of non-Gaussian random functions (e.g. positively skewed variables such a s rainfall; categorical data such as texture classes);
- geostatistical methods applied to space-time random functions;
- geostatistics applied to random functions defined on other metric spaces such as a sphere or river networks;
- Bayesian geostatistics, i.e using various forms of a priori information about the random function and formally updating this prior information with observations.

One is referred to above references for elaborate descriptions of these methods.

## 7.7 Exercises

Consider a square area size 10×10 units. Data points are located at locations (2,3), (3,8) and (7,9) with values of a property $z$ of 3, 8, 5 respectively. The property is modelled with a stationary and isotropic multivariate Gaussian random space function $Z(\mathbf{x})$ with mean $\mu_Z = 6$ and exponential semivariogram $\gamma(h) = 20\exp(-h/2)$.

1. Predict the value of $Z(\mathbf{x})$ at $\mathbf{x}_0 = (5,5)$ using simple kriging.
2. Predict the value of $Z(\mathbf{x})$ at $\mathbf{x}_0 = (5,5)$ using ordinary kriging.
3. Calculate the probability that $Z(5,5) > 10$.
4. Predict the average $\bar{Z}$ value of the 10×10 area using block kriging. For calculating the necessary point-block semivariances $\gamma(\mathbf{x},D)$ and average block semivariance $\gamma(D,D)$ discretise the block with four points at locations (2,2), (2,8), (8,2), (8,8).

# 8. Forward stochastic modelling

## 8.1 Introduction

In previous chapters methods were introduced for stochastic modelling of single variables, time series and spatial fields. A hydrological property that is represented by a random variable or a random function can be the target itself, e.g. flood events (chapter 4), groundwater head series (chapter 6) and areas of high concentration in groundwater (chapter 7). Often however, we have imperfect knowledge about some hydrological property that is used as parameter, input series, boundary condition or initial condition in a hydrological model. In that case, interest is focussed on the probability distribution or some uncertainty measure of the model output, given the uncertainty about the model input. This chapter is focussed on deriving these probability distributions or uncertainty measures.

More formally, consider a random variable $Z$ that is used as input[4] for some hydrological model $g$ to produce an output variable $Y$, which is also stochastic:

$$Y = g(Z) \tag{8.1}$$

The problem to solve is then: given that we know the probability distribution of $Z$ $f_Z(z)$ or some of its moments (e.g. mean and variance), what is the probability distribution $f_Y(y)$ of $Y$ or its moments? This problem is called *forward stochastic modelling*, as opposed to backward or *inverse (stochastic) modelling*. In the latter case we have observations of $Y$ and the unknown value of some deterministic parameter $z$ is estimated from these observations or, if $Z$ is stochastic, its conditional probability distribution. $f_Z(z \mid y)$.

Obviously, the problem of forward stochastic modelling can be put in more general terms, i.e. in case the input or the output are random functions of time, space or space-time, vectors of more random variables or even vector random functions. Also, he function $g()$ can have various forms, such a an explicit scalar or vector function, a differential equation or the outcome of a numerical model. Based on the form of g() and the form of $Z$ the following types of relations are considered in the framework of forward stochastic modelling (see Heuvelink (1998) for a good monograph about the subject):

- explicit functions of one random variable;
- explicit functions of multiple random variables;
- explicit vector functions;
- explicit functions of random functions of time, space or space-time;
- differential equations with a random parameter;
- stochastic differential equations.

[4] We use "input" here, but we mean in fact (see chapter 1 for system theory definitions) "input variables", "parameters", "boundary conditions" or "initial conditions".

In the following sections each of these problems is treated. For each problem type, a number of solution techniques are presented, where for each solution technique the conditions are given that should be met for its application.

## 8.2 Explicit functions of one random variable

Consider the relation between two random variables as shown in Equation (8.1).

### a) Derived distributions

Goal:

- the probability density function $f_Y(y)$.

Requirements:

- the probability density $f_Z(z)$ of $Z$ is known;
- the function $g(Z)$ is monotonous (only increasing or only decreasing), differentiable and can be inverted.

The cumulative distribution function of $Y$ can be obtained from the distribution function of $Z$ as follows:

$$F_Y(y) = F_Z(g^{-1}(y)) \tag{8.2}$$

while the probability density function (pdf) of $Y$ is related to the pdf of $Z$ as (Papoulis, 1991):

$$f_Y(y) = \left| \frac{d[g^{-1}(y)]}{dy} \right| f_Z(g^{-1}(y)) \tag{8.3}$$

where $g^{-1}(y)$ is the inverse function and the term $|\,.\,|$ the absolute value of its derivative. The term $|\,.\,|$ ensures that the area under $f_Y(y)$ is equal to 1.

Example Take the relation between water height above a weir crest $h$ and the discharge $q$ that is used to measure discharge with a weir (this could also be a rating curve for some river):

$$q = ah^b \tag{8.4}$$

Now suppose that the water height is observed with some error making it stochastic with pdf  The inverse of this relation and its derivative are given as:

$$g^{-1} = \left( \frac{q}{a} \right)^{\frac{1}{b}} \tag{8.5}$$

$$\left(g^{-1}\right)' = \frac{1}{b}\left(\frac{q}{a}\right)^{\frac{1-b}{b}} \tag{8.6}$$

The probability density function of discharge $f_H(h)$ then is given by:

$$f_Q(q) = \frac{1}{b}\left(\frac{q}{a}\right)^{\frac{1-b}{b}} f_H\left(\left(\frac{q}{a}\right)^{\frac{1}{b}}\right) \tag{8.7}$$

**b) Derived moments**

Goal:

- the moments of *Y*, e.g. $\mu_Y$ and $\sigma_Y^2$.

Requirements:

- the probability density $f_Z(z)$ of $Z$ is known.

The first two moments are then obtained through (see also 3.24):

$$\mu_Y = \int_{-\infty}^{\infty} g(z) f_Z(z) dz \tag{8.8}$$

$$\sigma_Y^2 = \int_{-\infty}^{\infty} g^2(z) f_Z(z) dz - \mu_Z^2 \tag{8.9}$$

Example Consider the same rating curve (Equation 8.4) with *H* following a uniform distribution between upper and lower values $h_u$ and $h_l$:

$$f_H(h) = \frac{1}{h_u - h_l} \tag{8.10}$$

The mean then becomes

$$\mu_Q = \int_{-\infty}^{\infty} \frac{ah^b}{h_u - h_l} dh = \frac{1}{h_u - h_l} \int_{h_l}^{h_u} ah^b dh = \frac{a}{(b+1)(h_u - h_l)}\left[h_u^{b+1} - h_l^{b+1}\right] \tag{8.11}$$

and the variance is given by:

$$\sigma_Q^2 = \frac{1}{h_u - h_l} \int_{h_l}^{h_u} a^2 h^{2b} dh = \frac{a^2}{(2b+1)(h_u - h_l)}\left[h_u^{2b+1} - h_l^{2b+1}\right] - \mu_Z^2 \tag{8.12}$$

In case that (8.7) and (8.8) cannot be evaluated analytically, the integrals can of course be solved numerically using for instance Euler-type integration methods.

**c) Monte Carlo simulation**

Goal:

- the probability density function $f_Y(y)$ or its moments.

Requirements:

- the probability density $f_Z(z)$ of $Z$ is known.

The principle of Monte Carlo simulation has been explained before in chapter 7, but is repeated here. Monte Carlo simulation is the advised method if the probability distribution of the input is known, if the complete distribution of model output is required and if the derived density approach (a) cannot be applied or if (8.2) cannot be evaluated analytically. Monte Carlo simulation proceeds as follows:

1. Draw a realisation $z_i$ of $Z$ using the pdf $f_Z(z)$. This is achieved by calculating the distribution function from the pdf

$$F_z(z) = \Pr[Z \le z] = \int_{-\infty}^{z} f_{Z'}(z')dz \tag{8.13}$$

   drawing a uniform deviate $u_i$ between 0 and 1 using a pseudo random number generator (e.g. Press et al, 1986), and converting $u_i$ using the inverse $z_i = F_Z^{-1}(u_i)$ (see Figure 8.1).

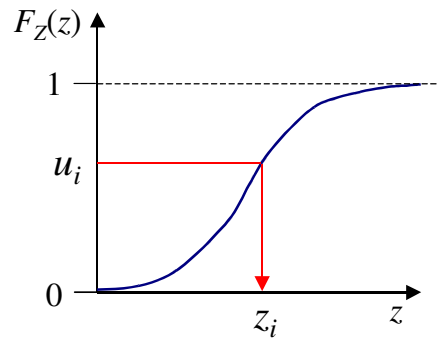



*Figure 8.1 Drawing a random number from a given distribution function*

2. Calculate the realisation $y_i$ of $Y$ by inserting $z_i$: $y_i = g(z_i)$.
3. Repeat steps 1 and 2 a large number of times (typically order 1000 to 10000 draws are necessary).
4. From the $M$ simulated output realisations $y_i$, $i$=1,..,$M$ the probability density function or cumulative distribution function of $Y$ can be estimated.

Example Consider again the rating curve (8.4) with parameter values $a$=5 and $b$=1.5, with $Q$ in m$^3$/d and with $H$ in m following a Gaussian distribution with mean 0.3 m and standard error of 0.02 m. Figure 8.2 shows the cumulative distribution function estimated from 1000 realisations of $Q$ calculated from 1000 simulated realisations of $H$. Also shown is the exact cumulative distribution function calculated using (8.2). It can be seen that both distributions are very close.

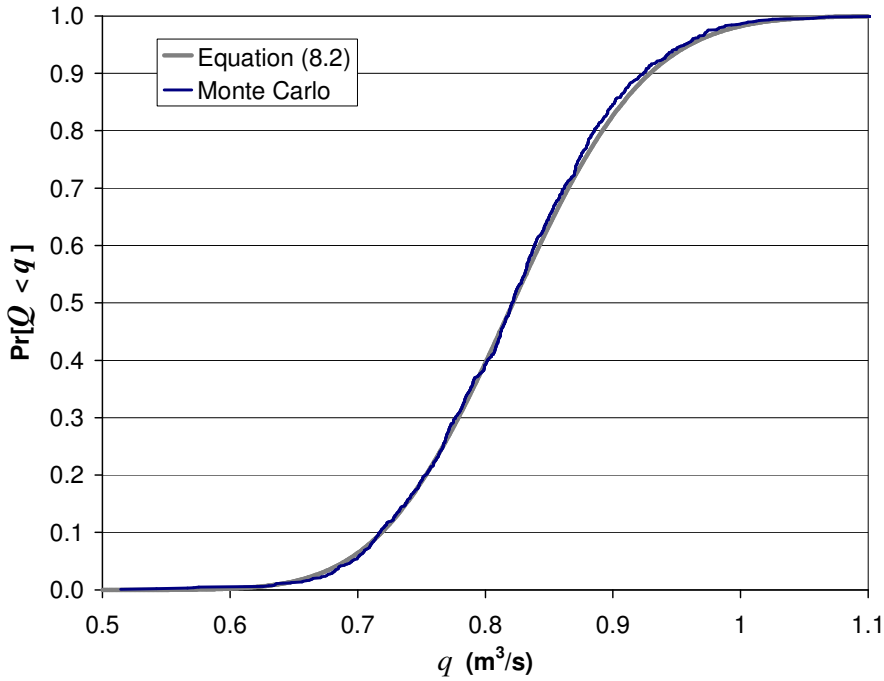



*Figure 8.2 Cumulative distribution functions: exact and estimated from Monte Carlo simulation.*

The Monte Carlo simulation presented here uses *simple random sampling*: values of $U$ are drawn from the entire range 0-1. To limit the number of realisations needed to accurately estimate the pdf of model output, a technique called *stratified random sampling* can be used. In that case, the interval 0-1 is divided into a finite number of intervals, preferably of equal width (e.g. 0-0.1, 0.1-0.2,..,0.9-1 in case of 10 intervals). In each interval a number of values of $U$ and the associated $Z$ are drawn. The result of this procedure is that the drawn realisations of $Z$ are more evenly spread over the value range, and that less realisations are necessary to obtain accurate estimates of $f_Y(y)$.

### d) Taylor expansion approximation

Goal:

- the moments of $Y$, e.g. $\mu_Y$ and $\sigma_Y^2$.

Requirements:

- the moments of $Z$, e.g. $\mu_Z$ and $\sigma_Z^2$, are known;
- the variance $\sigma_Z^2$ should not be too large.

Consider the Taylor expansion of the function $g(Z)$ around the value $g(\mu_Z)$:

$$Y = g(Z) = g(\mu_Z) + \left( \left. \frac{dg(z)}{dz} \right|_{z=\mu_Z} \right)(Z - \mu_Z) + \frac{1}{2}\left( \left. \frac{d^2 g(z)}{dz^2} \right|_{z=\mu_Z} \right)(Z - \mu_Z)^2 + \frac{1}{6}\left( \left. \frac{d^3 g(z)}{dz^3} \right|_{z=\mu_Z} \right)(Z - \mu_Z)^3 + \ldots \quad (8.14)$$

The *first order Taylor approximation* only considers the first two terms. The expected value is then approximated as:

$$\mu_Y = E[Y] \approx g(\mu_Z) + \left( \left. \frac{dg(z)}{dz} \right|_{z=\mu_Z} \right) E[(Z - \mu_Z)] = g(\mu_Z) \quad (8.15)$$

and the variance

$$\sigma_Y^2 = E[(Y - \mu_Y)^2] \approx \left( \left. \frac{dg(z)}{dz} \right|_{z=\mu_Z} \right)^2 E[(Z - \mu_Z)^2] = \left( \left. \frac{dg(z)}{dz} \right|_{z=\mu_Z} \right)^2 \sigma_Z^2 \quad (8.16)$$

Keeping the first three terms of Equation (8.14) and taking expectations yields the *second order Taylor approximation*.

The mean becomes:

$$\mu_Y \approx g(\mu_Z) + \frac{1}{2}\left( \left. \frac{d^2 g(z)}{dz^2} \right|_{z=\mu_Z} \right)\sigma_Z^2 \quad (8.17)$$

The general expression for the variance is very large, but can be simplified in case $Z$ is Gaussian (see Heuvelink, 1998). Here only the expression for Gaussian $Z$ is shown. For the full expression one is referred to Heuvelink (1998):

$$\sigma_Y^2 \approx \left( \left. \frac{dg(z)}{dz} \right|_{z=\mu_Z} \right)^2 \sigma_Z^2 + \frac{1}{4}\left( \left. \frac{d^2 g(z)}{dz^2} \right|_{z=\mu_Z} \right)^2 \sigma_Z^4 \quad (8.18)$$

Example One of the requirements for the Taylor approximation to work is that the variance of Z is not too large. To test this the first and second order Taylor approximations are applied to the rating curve $Q = aH^b$ for increasing variance of $H$. The derivatives that are necessary for this analysis are:

$$\alpha = \left( \left. \frac{dg(z)}{dz} \right|_{z=\mu_Z} \right) = abh^{b-1} \Big|_{h=\mu_H} = ab\mu_H^{b-1} \tag{8.19}$$

$$\beta = \left( \left. \frac{d^2 g(z)}{dz^2} \right|_{z=\mu_Z} \right) = ab(b-1)h^{b-2} \Big|_{h=\mu_H} = ab(b-1)\mu_H^{b-2} \tag{8.20}$$

with the first order Taylor approximation:

$$\mu_Q \approx a\mu_H^b \tag{8.21}$$

$$\sigma_Q^2 \approx \alpha^2 \sigma_H^2 \tag{8.22}$$

and the second order Taylor approximation:

$$\mu_Q \approx a\mu_H^b + \frac{\beta}{2}\sigma_H^2 \tag{8.23}$$

$$\sigma_Q^2 \approx \alpha^2 \sigma_H^2 + \frac{\beta^2}{4}\sigma_H^4 \tag{8.24}$$

To be able to analyse a large range of variances, the mean $\mu_H$ is set to 0.8 m (was 0.3 m). With $a$=5 and $b$=1.5 we have $\alpha = 6.708$ and $\beta = 4.193$. Figure 8.3 shows a plot of $\mu_Q$ and $\sigma_Q$ as a function of the standard deviation $\sigma_H$ as obtained from Monte Carlo simulation (1000 realisations) and with first and second order Taylor analysis. Clearly the Taylor approximation fails in estimating the mean if the variance becomes too large, although the second order methods performs much better than the first. In this example the variance is approximated accurately with both methods.

At this time it is convenient to remark that the methods presented in this chapter can also be viewed from the point of prediction errors. So, instead of having a mean $\mu_Z$ and variance $\sigma_Z^2$ of a stochastic input variable $Z$, we have a predictor $\hat{Z}$ of $Z$ and the prediction error variance $\sigma_{\hat{Z}}^2 = Var[\hat{Z} - Z]$. If the prediction error is unbiased, i.e. $E[\hat{Z} - Z] = 0$, then the same equations can be used as above, but with the mean $\mu_Z$ replaced by the prediction $\hat{z}$ and the variance $\sigma_Z^2$ by the error variance $\sigma_{\hat{Z}}^2$. From the point of error analysis the mean value of $Q$ then becomes:

$$\mu_Q \approx a\hat{h}^b + \frac{\beta}{2}\sigma_{\hat{H}}^2 \tag{8.25}$$

Equation (8.25) and Figure 8.3 show that in case of non-linear models unbiased (and even optimal predictions) of the model input do not yield unbiased (and certainly not optimal) predictions of the model output (see the remarks in Chapter 1). Adding higher order correction terms such as in (8.25) produce better results.

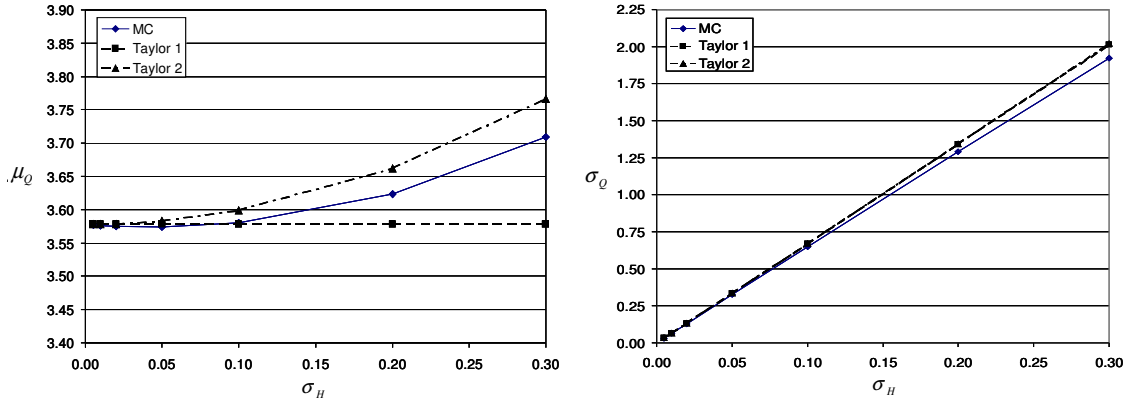



*Figure 8.3* $\mu_Q$ (left) and $\sigma_Q$ *(right) as a function of the standard deviation* $\sigma_H$ *as obtained from Monte Carlo simulation (1000 realisations) and the first and second order Taylor approximation.*

As a final remark: if the moments of $Y$ are required, but $g()$ is not differentiable or the variance of $Z$ is large, then Monte Carlo simulation could be used to derive the moments of $Y$. However, this means that some distribution type of $Z$ should be assumed.

## 8.3 Explicit functions of multiple random variables

The following explicit function of multiple random variables is considered:

$$Y = g(Z_1,..,Z_m) \tag{8.26}$$

Depending on what is aked about $Y$ , what is known about $Z_1,..,Z_m$ and the form of $g()$ a number of different methods can be distinguished:

### a) Derived distribution in the linear and Gaussian case

Goal:

- the probability density function $f_Y(y)$ .

Requirements:

- the joint probability density $f(z_1,..,z_m)$ of $Z_1,..,Z_m$ is known and multivariate Gaussian;
- the function g() is linear:

$$Y = a + \sum_{i=1}^{n} b_i Z_i \tag{8.27}$$

In the linear and multiGaussian case, the random variable *Y* is also Gaussian distributed. The multivariate Gaussian distribution of $Z_1,..,Z_m$ is completely described by the mean values $\mu_1,..,\mu_m$, the variances $\sigma_1^2,..,\sigma_m^2$ and the correlation coefficients $\rho_{ij}, i, j = 1,..,m$ with $\rho_{ij} = 1$ if $i = j$. The mean and variance of *Y* can then be obtained by:

$$\mu_Y = a + \sum_{i=1}^{m} b_i \mu_i \tag{8.28}$$

$$\sigma_Y^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} b_i b_j \rho_{ij} \sigma_i \sigma_j \tag{8.29}$$

Note that in case the $Z_i$ are not MultiGaussiasian that (8.28) and (8.29) are still valid expressions for the mean and the variance. However, in this case the mean $\mu_Y$ and the variance $\sigma_Y^2$ are not sufficient to characterise the complete pdf of *Y*.

**b) Derived distribution in the non-linear and Gaussian case**

Goal:

- the probability density function $f_Y(y)$.

Requirements:

- the joint probability density $f(z_1,.., z_m)$ of $Z_1,..,Z_m$ is known and multivariate Gaussian.

In case $Y = g(z_1,.., z_m)$ is non-linear we have to derive the distribution of *Y* through Monte Carlo simulation. To achieve this we have to draw realisations from the joint distribution $f(z_1,.., z_m)$. If this joint distribution is multivariate Gaussian this is possible through a technique called Cholesky decomposition (see box 5). The method then consists of:

1. Draw *M* realisations of the set of random variables $z_1^{(k)},.., z_m^{(k)}, k = 1,.., M$ from $f(z_1,.., z_m)$ using simulation by Cholesky-decomposition.
2. Use the *M* sets $z_1^{(k)},.., z_m^{(k)}, k = 1,.., M$ as input for the function *g*() to get *M* values of *y*: $y^{(k)}, k = 1,.., M$.
3. Estimate the distribution function or probability density function of *Y* from $y^{(k)}, k = 1,.., M$.

In case the joint distribution of $f(z_1,.., z_m)$ is not multivariate Gaussian, a solution is to apply a transformation to each of the variables $Z_1,..,Z_m$: $X_1 = \mathrm{Tr}(Z_1),..., X_m = \mathrm{Tr}(Z_m)$, such that we can assume $X_1,..., X_m$ multivariate Gaussian with $\mu_1 = \mu_2 =,..,= \mu_m = 0$,

$\sigma_1^2 = \sigma_2^2 = ,.., \sigma_m^2 = 1$. If we assume additionally that the correlation coefficients are unaltered by the transformation (note that this is generally not the case!), then realisations of $X_1,...,X_m$ can be simulated by Cholesky decomposition. The simulated realisations of $X_1,...,X_m$ are subsequently back transformed to realisations of $Z_1,..,Z_m$, which can then be used in the Monte Carlo analysis described above.

**c) Derived moments**

Goal:

- the moments of *Y*, e.g. $\mu_Y$ and $\sigma_Y^2$.

Requirements:

- the joint probability density $f(z_1,.., z_m)$ of $Z_1,..,Z_m$ is known.

The first two moments of *Y* are then obtained through:

$$\mu_Y = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{-\infty} g(z_1,.., z_m) f(z_1,.., z_m) dz_1 \cdots d_m \tag{8.30}$$

$$\sigma_Y^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{-\infty} g^2(z_1,.., z_m) f(z_1,.., z_m) dz_1 \cdots d_m - \mu_Z^2 \tag{8.31}$$

In practice it is highly unlikely that $f(z_1,.., z_m)$ is known, other than under the assumption that it is multivariate Gaussian. Also, evaluating the integrals, even numerically is likely to be very cumbersome. So, in practice this problem will be solved by assuming $f(z_1,.., z_m)$ to be multiGaussian (at least after transformation) and use Monte Carlo simulation as explained under (b).

**Box 5 Simulation by Cholesky-decomposition**

The goal is to simulate realisations of the set of random variables $Z_1,..,Z_m$ with multivariate Gaussian joint distribution $f(z_1,.., z_m)$, with parameters $\mu_1,.., \mu_m$, $\sigma_1^2,.., \sigma_m^2$ and $\rho_{ij}, i, j = 1,.., m$ with $\rho_{ij} = 1$ if $i = j$. The following steps are taken:

1. a vector of mean values is defined: $\boldsymbol{\mu} = (\mu_1, \mu_2,.., \mu_m)^T$;
2. the covariance matrix **C** is constructed with element $[C_{ij}]$ given by:

$$[C_{ij}] = \rho_{ij} \sigma_i \sigma_j ; \tag{8.32}$$

3. the covariance matrix is decomposed in a lower and upper triangular matrix that are each others transpose:

$$\mathbf{C} = \mathbf{LU} \text{ with } \mathbf{L} = \mathbf{U}^{\mathrm{T}}; \tag{8.33}$$

This operation is called Cholesky decomposition (a special form of LU-decomposition, so that the technique is also referred to as simulation by LU-decomposition). A routine to perform this operation can for instance be found in Press et al. (1986).

4. A realisation of the random vector $\mathbf{z} = (Z_1, Z_2, .., Z_m)^{\mathrm{T}}$ can now be simulated by simulating a vector $\mathbf{x} = (X_1, X_2, .., X_m)^{\mathrm{T}}$ of independent standard Gaussian random variables $X_1, ..., X_m$ using a random number generator (see 8.2) and performing the transformation:

$$\mathbf{z} = \boldsymbol{\mu} + \mathbf{Lx} \tag{8.34}$$

That (8.34) yield the right variables can be seen as follows. First, (8.33) yields linear transformations of Gaussian random variables so the simulated variables are Gaussian. Second, they have the correct mean value as:

$$E[\mathbf{z}] = E[\boldsymbol{\mu}] + E[\mathbf{Lx}] = \boldsymbol{\mu} + \mathbf{L}E[\mathbf{x}] = \boldsymbol{\mu} \tag{8.35}$$

And the correct covariance structure

$$\begin{aligned} E[(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}] &= E[\mathbf{Lx}(\mathbf{Lx})^{\mathrm{T}}] = E[\mathbf{Lxx}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}] = \mathbf{L}E[\mathbf{xx}^{\mathrm{T}}]\mathbf{L}^{\mathrm{T}} \\ &= \mathbf{LIL}^{\mathrm{T}} = \mathbf{LL}^{\mathrm{T}} = \mathbf{LU} = \mathbf{C} \end{aligned} \tag{8.36}$$

So the simulated variables are indeed drawn from a multivariate Gaussian distribution with the preposed statistics.

Note that this method can also be used to simulate realisations of multiGaussian random space functions on a grid, i.e. as an alternative to sequential simulation. In that case the random vector contains the values of the random space function at the grid nodes $\mathbf{z} = (Z(\mathbf{x}_1), Z(\mathbf{x}_2), .., Z(\mathbf{x}_m))^{\mathrm{T}}$, the mean is constant and the covariance matrix is constructed as:

$$[C_{ij}] = C_Z(\mathbf{x}_i, \mathbf{x}_j); \tag{8.37}$$

**d) Taylor expansion approximation**

Goal:

- the moments of *Y*, e.g. $\mu_Y$ and $\sigma_Y^2$.

Requirements:

- the joint moments of $Z_1,..,Z_m$ are known up to a certain order, e.g.: $\mu_1,..,\mu_m$, $\sigma_1^2,..,\sigma_m^2$ and $\rho_{ij}, i, j = 1,..,m$ ;
- the variances $\sigma_1^2,..,\sigma_m^2$ should not be too large.

We first define a vector $\boldsymbol{\mu} = (\mu_1, \mu_2,.., \mu_m)^{\mathrm{T}}$ that contains the mean of the *m* random input variables. Next, we consider the Taylor expansion of the function *g*(*Z*) around the value $g(\boldsymbol{\mu}) = g(\mu_1, \mu_2,.., \mu_m)$ :

$$\begin{aligned} Y = g(Z) = g(\boldsymbol{\mu}) &+ \sum_{i=1}^{m}\left(\frac{dg(z)}{dz_i}(\boldsymbol{\mu})\right)(Z_i - \mu_i) \\ &+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{d^2 g(z)}{dz_i dz_j}(\boldsymbol{\mu})\right)(Z_i - \mu_i)(Z_j - \mu_j) \\ &+ \frac{1}{6}\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{d^3 g(z)}{dz_i dz_j dz_k}(\boldsymbol{\mu})\right)(Z_i - \mu_i)(Z_j - \mu_j)(Z_k - \mu_k) + ... \end{aligned} \tag{8.38}$$

The *first order Taylor approximation* only considers the first two terms. The expected value is then approximated as:

$$\mu_Y = E[Y] \approx g(\boldsymbol{\mu}) + \sum_{i=1}^{m}\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)E[(Z_i - \mu_i)] = g(\boldsymbol{\mu}) \tag{8.39}$$

and the variance

$$\begin{aligned} \sigma_Y^2 = E[(Y - E[Y]^2] &\approx E\left[\left(g(\boldsymbol{\mu}) + \sum_{i=1}^{m}\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)(Z_i - \mu_i) - g(\boldsymbol{\mu})\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^{m}\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)(Z_i - \mu_i)\right)\left(\sum_{j=1}^{m}\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)(Z_j - \mu_j)\right)\right] \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)E[(Z_i - \mu_i)(Z_j - \mu_j)] \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)\left(\frac{dg(z)}{dz}(\boldsymbol{\mu})\right)\rho_{ij}\sigma_i\sigma_j \end{aligned} \tag{8.40}$$

Keeping the first three terms of Equation (8.38) and taking expectations yields the *second order Taylor approximation*. We will only show the expression for the mean here. For the variance one is referred to Heuvelink (1998).

$$\mu_Y = E[Y] \approx g(\boldsymbol{\mu}) + \sum_{i=1}^{m}\left(\frac{dg(z)}{dz_i}(\boldsymbol{\mu})\right)E[(Z_i - \mu_i)] + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{d^2 g(z)}{dz_i dz_j}(\boldsymbol{\mu})\right)E[(Z_i - \mu_i)(Z_j - \mu_j)] = g(\boldsymbol{\mu}) + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{d^2 g(z)}{dz_i dz_j}(\boldsymbol{\mu})\right)\rho_{ij}\sigma_i\sigma_j \tag{8.41}$$

Example Consider the weir equation or rating curve $Q = Ah^B$, where $A$ and $B$ are random variables with statistics $\mu_A, \sigma_A^2, \mu_B, \sigma_B^2$ and $\rho_{AB}$. The first order Taylor approximation of the mean becomes:

$$E[Q] \approx \mu_A h^{\mu_B} \tag{8.42}$$

and the second order approximation:

$$E[Q] \approx \mu_A h^{\mu_B} + \frac{1}{2}(\mu_A h^{\mu_B}(\ln h)^2)\sigma_B^2 + (h^{\mu_B}\ln h)\rho_{AB}\sigma_A\sigma_B \tag{8.43}$$

The variance from the first order Taylor analysis is given by:

$$\sigma_Q^2 \approx (h^{2\mu_B})\sigma_B^2 + (4\mu_A h^{2\mu_B}\ln h)\rho_{AB}\sigma_A\sigma_B + (2\mu_A^2 h^{2\mu_B}(\ln h)^2)\sigma_B^2 \tag{8.44}$$

As can be seen, these expressions quickly become quite extensive, especially if due to larger variances $\sigma_A^2, \sigma_B^2$ higher order terms have to be included. The alternative then is to use Monte Carlo simulation by jointly simulating realisations of the variables $A$ and $B$ using Cholesky decomposition. Of course, this means that some joint probability distribution for these random variables has to be assumed.

## 8.4 Spatial, temporal or spatio-temporal integrals of random functions

We consider the following relationship (here we consider space, but it could also be time or space-time):

$$Y_D = \int_{\mathbf{x}\in D} g[Z(\mathbf{x})]d\mathbf{x} \tag{8.45}$$

**a) Simple averaging of moments**

Goal:

- the moments of *Y*, e.g. $\mu_Y$ and $\sigma_Y^2$, $COV(Y_{D_1}, Y_{D_2})$.

Conditions:

- the function g[] is linear;
- the random function $Z(\mathbf{x})$ is wide sense stationary (see chapter 5);
- the mean $\mu_Z$ and covariance function $C_Z(\mathbf{x}_1, \mathbf{x}_2) = C_Z(\mathbf{x}_2 - \mathbf{x}_1)$ are known.

If the function g[] is linear, e.g. $g[Z(\mathbf{x})] = a + bZ(\mathbf{x})$, the moments of $Y_D$ can be evaluated by spatial integration of the moments of $Z(\mathbf{x})$ (see also section 5.6 and 7.3.3).

The mean of $Y_D$ is given by

$$\begin{aligned} E[Y_D] &= E\left[\int_{\mathbf{x}\in D} a + bZ(\mathbf{x})d\mathbf{x}\right] = a + bE\left[\int_{\mathbf{x}\in D} Z(\mathbf{x})d\mathbf{x}\right] \\ &= a + b\int_{\mathbf{x}\in D} E[Z(\mathbf{x})]d\mathbf{x} = a + b\int_{\mathbf{x}\in D} \mu_Z d\mathbf{x} = a + b|D|\mu_Z \end{aligned} \tag{8.46}$$

and the variance by:

$$\begin{aligned} E[(Y_D - E[Y_D])^2] &= E\left[\left(\int_{\mathbf{x}\in D} a + bZ(\mathbf{x})d\mathbf{x} - a - b|D|\mu_Z\right)^2\right] \\ &= b^2 E\left[\left(\int_{\mathbf{x}\in D} Z(\mathbf{x})d\mathbf{x} - |D|\mu_Z\right)^2\right] \\ &= b^2 E\left[\left(\int_{\mathbf{x}\in D} (Z(\mathbf{x}) - \mu_Z)d\mathbf{x}\right)^2\right] \\ &= b^2 E\left[\int_{\mathbf{x}_1\in D} (Z(\mathbf{x}_1) - \mu_Z)d\mathbf{x}_1 \int_{\mathbf{x}_2\in D} (Z(\mathbf{x}_2) - \mu_Z)d\mathbf{x}_2\right] \\ &= b^2 E\left[\int_{\mathbf{x}_2\in D}\int_{\mathbf{x}_1\in D} (Z(\mathbf{x}_1) - \mu_Z)(Z(\mathbf{x}_2) - \mu_Z)d\mathbf{x}_1 d\mathbf{x}_2\right] \\ &= b^2 \int_{\mathbf{x}_2\in D}\int_{\mathbf{x}_1\in D} E[(Z(\mathbf{x}_1) - \mu_Z)(Z(\mathbf{x}_2) - \mu_Z)]d\mathbf{x}_1 d\mathbf{x}_2 \\ &= b^2 \int_{\mathbf{x}_2\in D}\int_{\mathbf{x}_1\in D} C_Z(\mathbf{x}_1, \mathbf{x}_2)d\mathbf{x}_1 d\mathbf{x}_2 \end{aligned} \tag{8.47}$$

By the same type of derivation the covariance between spatial averages of two domains can be derived (see also section 5.6 and 7.3.3):

$$COV(Y_{D_1}, Y_{D_2}) = b^2 \int_{\mathbf{x}_2 \in D_2} \int_{\mathbf{x}_1 \in D_1} C_Z(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{8.48}$$

The spatial integrals can be solved either analytically in certain cases (e.g. Vanmarcke, 1983), but are usually approximated numerically as is explained in section 7.3.3.

Note that if the random function $Z(\mathbf{x})$ is wide sense stationary and multiGaussian, $Y_D$ will be Gaussian distributed also and its probability density function is given through the mean (8.46) and the variance (8.47).

**b) Monte Carlo simulation**

Goal:

- the moments of $Y$, e.g. $\mu_Y$ and $\sigma_Y^2$, $COV(Y_{D_1}, Y_{D_2})$ or its probability density $f_Y(y)$.

Conditions:

- the multivariate probability density function of $Z(\mathbf{x})$ is known.

If g() is non-linear or we are interested in the complete probability density function geostatistical simulation in a Monte Carlo analysis is the appropriate method. The following steps are taken:

1. generate $M$ realisations $z(\mathbf{x})^{(k)}, k = 1,..,m$ of the random function $Z(\mathbf{x})$ using geostatistical simulation on a fine grid discretising the domain $D$. If $Z(\mathbf{x})$ is non-Gaussian, a transformation to a Gaussian distribution is in order, after which sequential Gaussian simulation can be applied (see sections 7.4.3 and 7.5 for elaborate descriptions of normal-transforms and geostatistical simulation respectively);
2. the $M$ realisations are used as input for the spatial integral (8.45) yielding $M$ results $y_D^{(k)}, k = 1,..,M$;
3. from the simulated values $y_D^{(k)}, k = 1,..,M$ the moments and the probability density function of $Y_D$ can be estimated.

If the random function is observed at a number of locations, conditional realisations should be drawn (see also section 7.5). This allows one to investigate the effect of additional sampling of $Z$ on the uncertainty about $Y_D$.

## 8.5 Vector functions

Consider a vector of model inputs (parameters, input variables, boundary conditions etc) that are random variables: $\mathbf{z} = (Z_1,...,Z_m)^{\mathrm{T}}$. The vector of model inputs is linked to a vector of model outputs $\mathbf{y} = (Y_1,...,Y_n)^{\mathrm{T}}$ through a functional relationship $g()$:

$$\mathbf{y} = g(\mathbf{z}) \tag{8.49}$$

The goal is to get the joint pdf or the moments of **y**.

### a) First order analysis

Required:

- the statistical moments of **y**: $\boldsymbol{\mu}_{\mathbf{y}} = E[\mathbf{y}] = (E[Y_1],..., E[Y_n])^{\mathrm{T}}$ and the covariance matrix $C_{\mathbf{yy}} = E[(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})^{\mathrm{T}}]$.

Conditions:

- The statistical moments of **z** should be known: $\boldsymbol{\mu}_{\mathbf{z}} = E[\mathbf{z}] = (\mu_1,..\mu_m)^{\mathrm{T}}$, $C_{\mathbf{zz}} = E[(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{z}})^{\mathrm{T}}]$.
- The variances of the elements of **z** $\sigma_1^2,..,\sigma_m^2$ should not be too large.

The first order analysis is in fact the first order Taylor approximation (see 8.3) applied to each of the elements of **y**. The first order approximation is given by:

$$E[\mathbf{y}] \approx g(\boldsymbol{\mu}_{\mathbf{Z}}) \tag{8.50}$$

To obtain the covariance matrix of **y** the covariance matrix of **z** is constructed. This is an $m \times m$ matrix with the following elements (with $\rho_{ij}$ the correlation between $Z_i$ and $Z_j$):

$$[C_{zz}(i, j)] = \rho_{ij}\sigma_i\sigma_j \tag{8.51}$$

Also, the *sensitivity matrix* or *Jacobian* is required. This $n \times m$ matrix gives the derivatives of element $Y_i$ with respect to input $Z_j$ and has the following form:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \cdots & \frac{\partial y_1}{\partial z_m} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \frac{\partial y_n}{\partial z_1} & \cdots & \cdots & \frac{\partial y_n}{\partial z_m} \end{pmatrix} \tag{8.52}$$

Sometimes it is possible to construct this matrix analytically, i.e. if the vector function $g(\mathbf{z})$ consist for each element $y_i$ of $\mathbf{y}$ of a separate explicit and differentiable function $g_i(z_1,..,z_m)$. However, usually this is not the case and $g(\mathbf{z})$ may represent some numerical model, e.g. a groundwater model, where the elements $y_i$ of $\mathbf{y}$ are state elements, perhaps defined at some grid. In this case, a sensitivity analysis must be performed by running the the model $g()$ $m$+1 times: one baseline run where the model inputs are set at there mean values (i.e. Equation 8.50) and one run for each model in inputs $z_j$ where $z_j$ is slightly changed, e.g. $z_j = \mu_j + \Delta z_j$. From these runs the changes in the values of the elements of $y_i$, e.g. $\Delta y_i$, are calculated and the derivatives are subsequently estimated as:

$$\frac{\partial y_i}{\partial z_j} \approx \frac{\Delta y_i}{\Delta z_j} \tag{8.53}$$

With the sensitivity matrix and the covariance matrix a first order approximation of the covariance matrix of $\mathbf{y}$ can be provided:

$$\mathbf{C}_{\mathbf{yy}} = \mathbf{J}\mathbf{C}_{\mathbf{zz}}\mathbf{J}^{\mathrm{T}} \tag{8.54}$$

Some additional remarks about this method:

- Above equations have been developed for stochastic input variables with prescribed means and variances. Of course, as is the case with the Taylor approximation in sections 8.2 and 8.3, this method can also be used as a first order approximation of a prediction error covariance. In that case the prediction equation becomes:

  $$\hat{\mathbf{y}} \approx g(\hat{\mathbf{z}}) \tag{8.55}$$

  with $\hat{\mathbf{z}} = (\hat{Z}_1,..,\hat{Z}_m)$ the predicted values of the model inputs and $\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}} = E[(\mathbf{z}-\hat{\mathbf{z}})(\mathbf{z}-\hat{\mathbf{z}})^{\mathrm{T}}]$ the covariance matrix of the prediction errors; and similarly for $\mathbf{y}$. Equation (8.54) then becomes:

  $$\mathbf{C}_{\hat{\mathbf{y}}\hat{\mathbf{y}}} = \mathbf{J}\mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}}\mathbf{J}^{\mathrm{T}} \tag{8.56}$$

- If the function g() is a matrix multiplication

$$\mathbf{y} = \mathbf{A}\mathbf{z} \tag{8.57}$$

the system is linear, the elements of the sensitivity matrix are exactly the elements of the matrix **A**, i.e. $[j_{ij}] = [a_{ij}]$ , and an exact equation for the covariance of **y** is given by:

$$\mathbf{C}_{\mathbf{yy}} = \mathbf{A}\mathbf{C}_{\mathbf{zz}}\mathbf{A}^{\mathrm{T}} \tag{8.58}$$

If on top of that **z** has a multivariate Gaussian distribution then **y** is also multivariate Gaussian and the derived mean and variance of **y** completely determine its probability distribution (using Equation 3.87 with $\boldsymbol{\mu}_{\mathbf{y}}$ and $\mathbf{C}_{\mathbf{yy}}$).

- This method can also be used for transient models. Suppose that the following model applies:

$$\mathbf{y}(t) = g(\mathbf{z}, t) \tag{8.59}$$

where **y** is the outcome of some dynamic model, e.g. a transient groundwater model, with stochastic input or parameters **z**. Then $m+1$ transient runs can be performed, i.e. the baseline and one for each perturbed parameter, and for each time that one requires the sensitivity can be determined:

$$\frac{\partial y_i}{\partial z_j}(t) \approx \frac{\Delta y_i(t)}{\Delta z_j} \tag{8.60}$$

and the covariance of $\mathbf{y}(t)$ be approximated at each time as:

$$\mathbf{C}_{\mathbf{yy}}(t) = \mathbf{J}(t)\mathbf{C}_{\mathbf{zz}}\mathbf{J}^{\mathrm{T}}(t) \tag{8.61}$$

**b) Monte Carlo simulation**

In case of strong non-linearity of $g()$ or large variances of the elements of **z**, the linear approximation would no longer work. In that case Monte Carlo simulation is to be applied, as described before: 1) $M$ realisations of **z** are simulated (e.g. using Cholesky decomposition as shown in Box 5); 2) the $M$ simulated realisations **z** are used as input for $g(\mathbf{z})$ yielding $M$ realisations **y** ; 3) the statistics of **y** can be estimated from the $M$ realisations of **y**.

## 8.6 Differential equations with a random variable

Consider a partial differential equation with two random parameters, e.g. the groundwater equation in two spatial dimensions with a homogenous and isotropic but random transmissivity $T$ and a homogenous storage coefficient $S$:

$$S\frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(T\frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial H}{\partial y}\right) = T\left(\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2}\right) \tag{8.62}$$

What can immediately be seen from (8.62) is that, even though the transmissivity $T$ is random, it can be placed outside the derivatives because it does not change with space. If the transmissivity is a single random variable there are effectively two way of obtaining the statistics of the random head $H$, depending on whether an analytical solution is available.

1. If an analytical solution is available, Equation (8.62) can be solved for given $T$ and $S$ (as if they were deterministic) to produce an explicit relation between $H(\mathbf{x},t)$ and $S$ and $T$. This relation would generally be non-linear and the a Taylor approximation could be used to derive the statistics of $H(\mathbf{x},t)$ from the joint statistics of $S$ and $T$: $\mu_S, \sigma_S^2, \mu_T, \sigma_T^2$ and $\rho_{ST}$. If a Taylor approximation is not appropriate, e.g. because the variances of $S$ and $T$ are large, then realisations of $S$ and $T$ can be simulated assuming some joint distribution $f(s,T)$ of $S$ and $T$ (if multiGaussian Cholosky decomposition can be used). These realisations can then be plugged into the analytical solution of (8.62) to produce realisations of $H(\mathbf{x},t)$ from which its statistics can be obtained.
2. If no analytical solution can be obtained, then Monte Carlo simulation in combination with a numerical method, e.g. finite elements or finite difference, is the only option. A large number $M$ of realisations of $S$ and $T$ are simulated assuming a joint distribution $f(s,T)$ of $S$ and $T$ (if multiGaussian Cholosky decomposition can be used). The $M$ simulated realisations are used as parameters in the equations solved by the finite difference or finite element scheme. The numerical solution is obtained for each simulated parameter set to yield $M$ realisations of $H^{(k)}(\mathbf{x}_i,t_j)$, $k$=1,..,$M$ at a finite number of points in space and time. From these, the statistics (mean, variance, spatial and temporal covariance) of $H(\mathbf{x},t)$ can be obtained.

So, in short: if random parameters are involved in differential equations that are not random functions of space or time, they can treated as deterministic while solving the differential equation and analysed as stochastic variables afterwards; that is if an analytical solution can be found.

Example Consider the following differential equation describing the concentration of some pollutant in a lake as a function of time:

$$v\frac{dC}{dt} = -vKC + q_{in} \tag{8.63}$$

where $v$ is the volume of the lake (assumed constant and known) $q_{in}$ is the constant an known input load and $K$ is a random decay coefficient. The solution to this equation is (with known initial concentration $c(t)$=0):

$$C(t) = \frac{q_{in}}{vK}\left(1 - e^{-Kt}\right) \tag{8.64}$$

Now the statistics of $C(t)$ can be derived from those of $K$. For instance, using a first order Taylor approximation (see section 8.2) the following relation can be derived for the variance:

$$\sigma_C^2(t) = \frac{q_{in}^2}{v^2}\frac{\left(e^{-\mu_K t}(1+\mu_K t)-1\right)^2}{\mu_K^4}\sigma_K^2(t) \tag{8.65}$$

Figure 8.65 shows the development of the mean concentration with time as well as the confidence band of one standard deviation based on a first order Taylor analysis and the assumption that $C$ is Gaussian distributed. The following parameters are used: $q_{in}/v = 100$ (mg m$^{-3}$year$^{-1}$), $\mu_K = 0.5$(year$^{-1}$), $\sigma_K^2 = 0.01$ (year$^{-2}$).

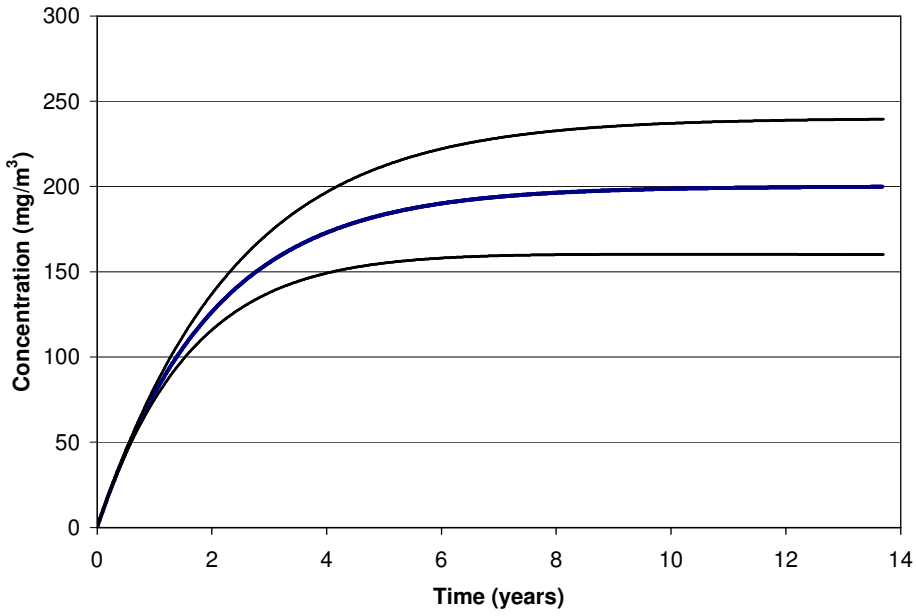



*Figure 8.4 Evolution of concentration of pollutant in a lake described by Equation (8.64) with random decay rate K. Mean concentration (central line) and one standard deviation prediction intervals (outer lines) are approximated by a first order Taylor analysis; parameters:* $q_{in}/v = 100$ (mg m$^{-3}$year$^{-1}$), $\mu_K = 0.5$(year$^{-1}$), $\sigma_K^2 = 0.01$ (year$^{-2}$).

## 8.7 Stochastic differential equations

As a last case, consider the following differential equations:

1. Transient groundwater equation for two-dimensional flow with heterogeneous storage coefficient and transmissivity, described as random space functions:

$$S(x,y)\frac{\partial H}{\partial t} = \frac{\partial}{\partial x}\left(T(x,y)\frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(T(x,y)\frac{\partial H}{\partial y}\right) \tag{8.66}$$

9.* Evolution of lake concentration with decay rate as a random function of time:

$$v\frac{dC}{dt} = -vK(t)C + q_{in} \tag{8.67}$$

In both of these cases we cannot hope to find an analytical solution given a particular realisation of the random functions $S(\mathbf{x})$, $T(\mathbf{x})$ or $K(t)$, as the behaviour of random functions is generally wild and not described by a simple analytical expression. There are two alternatives to solve these *stochastic differential equations*:

The first alternative is to assume some form of stationarity about the random functions and then develop differential equations in terms of the moments of the dependent variables and the moments of the random inputs. As the latter are assumed (wide sense) stationary the moments are constant, such that analytical solutions may be feasible.

The second alternative is to use Monte Carlo simulation, i.e. simulate realisations of the random function and use these as input for the differential equations. The differential equations are subsequently solved for each realisation with a numerical scheme such as finite difference or fine elements.

In the following examples of both approaches are given. The field of stochastic subsurface hydrology is the most advanced in terms of analytical stochastic analysis of parameter heterogeneity, with many papers in various hydrological journals, in particular in Water Resources Research. Although not very recent, extensive overviews of the advances in this field can be found in Dagan (1989) and Gelhar (1993). The applications in these books mostly pertain to flow and transport in porous media described with wide sense stationary stochastic functions and assuming infinite domains with uniform flow. Since the appearance of these books, advances have been made on finding (quasi)-analytical solutions for finite domains, non-uniform flow, random boundary conditions, unsaturated flow, two-phase flow, non-stationary (even fractal) random media and fractal porous media. A more recent book with some advanced topics is written by Zhang (2002).

**Box 6 Mean square differentiable random functions and white noise**

In both approaches, we would like to use the rules of standard differential and integral calculus. For these standard rules to apply, the random functions involved have to be *mean square integrable*, i.e. the following limit expressions should exists (a similar expression can be formulated for random functions of space):

$$\lim_{\tau \to 0} E\left[\left(\frac{Z(t+\tau)-Z(t)}{\tau}-\frac{dZ}{dt}\right)^2\right]=0 \tag{8.68}$$

So, averaged over all possible realisations, the quadratic difference between the derivative of the random function and its finite difference approximation should approach to zero if we look at increasingly smaller intervals $\tau$. If (8.68) exists, and we interpret the differential *dZ*/*dt* in the mean square sense, normal calculus rules can thus be applied.

A necessary and sufficient condition for this to apply is that the following equation is finite (Gelhar, 1993, p. 53):

$$\int_{-\infty}^{\infty} \omega^2 S_Z(\omega) d\omega < \infty \tag{8.69}$$

where $S_Z(\omega)$ is the spectral density function of the random function *Z*(*t*). An alternative necessary and sufficient condition is that the second derivative if the covariance function at lags approaching zero is finite (Gelhar, 1993):

$$\left.\frac{d^2 C_Z(\tau)}{d\tau^2}\right|_{\tau=0} < \infty \tag{8.70}$$

From equation (8.69) it can be seen that if aB stationary random function is mean square differentiable, it must have a spectral density that goes to zero more rapidly than $|\omega^{-3}|$. This means that the random function should be limited with respect to the higher frequency components. From Table 5.3 it can be seen that this is not the case for instance the exponential covariance function. However, Gelhar (1993, p. 53) shows how cutting off the spectrum at some maximum frequency $\omega_m$ does produce a random function that is mean square differentiable, while this weeding out of the high frequency components only slightly decreases the variance of *Z*(*t*). Such a cut off is also physically defendable, as many hydrological variables such as porosity, hydraulic conductivity or rainfall intensity are only defined as continuous variables above a certain minimum averaging in space or time. A natural cut-off for conductivity for instance could be the Representative Elementary Volume (REV) as defined by Bear (1972). The random function used has no validity for variations within this volume, or, in ters of the covariance function (8.70): it is not applied at lag differences smaller than the REV size. Given this assumption it can

then be assumed that the regular limited-band stationary random functions used are mean square differentiable.

An important exception is the white noise process. Not only is this process not mean square differentiable, it also has infinite variance. This can be seen as follows. White noise can be defined as the formal derivative of Brownian motion:

$$W(t) \equiv \lim_{\tau \to 0} \frac{B(t+\tau) - B(t)}{\tau} \tag{8.71}$$

Brownian motion is a process for which the variance of the difference between two values sampled at two different times is proportional to the time interval between the sample times (where the mean of the difference is zero):

$$\begin{aligned} &E[B(t+\Delta t) - B(t)] = 0 \\ &Var[B(t+\Delta t) - B(t)] \propto \Delta t \qquad \forall t, \Delta t \end{aligned} \tag{8.72}$$

Using this definition in (8.71) one obtains:

$$\begin{aligned} \text{var}[W(t)] &\equiv E\left[\left(\lim_{\tau \to 0} \frac{B(t+\tau) - B(t)}{\tau^2}\right)^2\right] \\ &= \lim_{\tau \to 0} \frac{E[(B(t+\tau) - B(t))^2]}{\tau^2} = \lim_{\tau \to 0} \frac{\tau}{\tau^2} = \infty \end{aligned} \tag{8.73}$$

White noise is usually not used as a model for measurable variables, but for random influences that are virtually uncorrelated in space and time, e.g. measurement noise or fluctuations in input, parameters and boundary conditions which have much smaller space or time scales than the target variable of interest. For instance, looking at the equation modelling the concentration in a lake (Equation 8.67): if the random process $K(t)$ is a stationary process with a correlation scale comparable to the dynamics of the concentration variations (e.g. it may depend on temperature, which fluctuates in the order of weeks or even months; see Figure 8.4), then it would be treated with the methods described in this section. However, suppose it is meant to incorporate the influence of turbulence on the decay rate. In that case it has a time scale of seconds and is thus best modelled as a white noise process. In that case, the methods described here no longer apply and (8.67) is solved using a special form of stochastic calculus (called Ito calculus) (e.g. Gardiner, 1983).

**a) Small perturbation analysis**

First order perturbation analysis has been widely used to analyse the effect of unknown heterogeneity in hydraulic parameters on the uncertainty about hydraulic heads, fluxes and concentration plumes. The applicability of the method and the tractability of analytical solutions rest on the following conditions:

- The logarithm of hydraulic conductivity $Y$=ln$K$ is modelled as (locally) wide sense stationary random function. If a process is locally wide sense stationary it means that the mean of the random function may be a function of space, but it changes very slowly compared to the integral scale (see chapter 5) of the random function. So if $Y'(\mathbf{x}) = Y(\mathbf{x}) - \mu_Y(\mathbf{x})$ is the wide sense stationary residual with integral scale $I_Y$ then the condition to be met is (for two dimensions):

$$I_Y\left(\frac{\partial \mu_Y}{\partial x} + \frac{\partial \mu_Y}{\partial y}\right) << 1; \tag{8.74}$$

- The flow takes place in an infinite domain and is uniform or at least slowly varying in space. In that case the gradient of hydraulic head $\nabla H$ can also be assumed a (locally) wide sense stationary random function (that is if hydraulic conductivity is locally wide sense stationary). It means that the following condition should hold (with $I_{\nabla H}$ the integral scale of the head gradient):

$$I_{\nabla H}\left(\frac{\partial \mu_{\nabla H}}{\partial x} + \frac{\partial \mu_{\nabla H}}{\partial y}\right) << 1; \tag{8.75}$$

- The variance of the logarithm of hydraulic conductivity $\sigma_Y^2$ is small. As a rule of thumb one usually requires that $\sigma_Y^2 \leq 1$.

Above conditions are formulated in terms of ln$K$ because it is more convenient to work with ln$K$ than $K$. There are three reasons why this is so (Gelhar, 1993). First, it is generally accepted that the logarithm of $K$ follows a Gaussian distribution (e.g. Freeze, 1975), so that the assumption of wide sense stationary ln$K$ means that the mean and the covariance function are sufficient to describe its multivariate probability (see chapter 5). Second, the variation in ln$K$ is smaller than that of $K$, which makes the small perturbation approach applicable to a wider range of variances of $K$. The third reason is that the logarithm of hydraulic conductivity appears naturally in the equation for groundwater flow (as will be shown hereafter (Equation 8.77)).

The following explanation of the workings of the small perturbation approximation is taken from Gelhar (1993). We start with the steady-state groundwater flow equation in multiple dimensions:

$$\frac{\partial}{\partial x_i}\left(K\frac{\partial H}{\partial x_i}\right)=0;\ i=1,N,N=1,2\text{ or }3. \tag{8.76}$$

This equation can be rewritten as (with $Y$=ln$K$):

$$\begin{aligned}
&\frac{\partial K}{\partial x_i}\frac{\partial H}{\partial x_i}+K\frac{\partial^2 H}{\partial x_i^2}=0 \ \Rightarrow \frac{\partial K}{K\partial x_i}\frac{\partial H}{\partial x_i}+\frac{\partial^2 H}{\partial x_i^2}=0\\
&\Rightarrow(\text{as }\frac{\partial K}{K}=\partial(\ln K))\Rightarrow\frac{\partial^2 H}{\partial x_i^2}+\frac{\partial \ln K}{\partial x_i}\frac{\partial H}{\partial x_i}=0\\
&\Rightarrow\frac{\partial^2 H}{\partial x_i^2}+\frac{\partial Y}{\partial x_i}\frac{\partial H}{\partial x_i}=0
\end{aligned} \tag{8.77}$$

Equation (8.77) show that variations in hydraulic head are driven by variations in lnK. The random functions for logconductivity and hydraulic head are written as a mean value and a perturbation as follows:

$$Y(\mathbf{x})=\mu_Y(\mathbf{x})+Y'(\mathbf{x});\ \ E[Y(\mathbf{x})]=\mu_Y(\mathbf{x});E[Y'(\mathbf{x})]=0 \tag{8.78}$$

$$H(\mathbf{x})=\mu_H(\mathbf{x})+H'(\mathbf{x});\ \ E[H(\mathbf{x})]=\mu_H(\mathbf{x});E[H'(\mathbf{x})]=0 \tag{8.79}$$

Substitution of these expansions into (8.77) yields:

$$\begin{aligned}
&\frac{\partial^2(\mu_H+H')}{\partial x_i^2}+\frac{\partial(\mu_Y+Y)}{\partial x_i}\frac{\partial(\mu_H+H')}{\partial x_i}=0\\
&\Rightarrow\frac{\partial^2(\mu_H)}{\partial x_i^2}+\frac{\partial^2(H')}{\partial x_i^2}+\frac{\partial\mu_Y}{\partial x_i}\frac{\partial\mu_H}{\partial x_i}+\frac{\partial\mu_Y}{\partial x_i}\frac{\partial H'}{\partial x_i}+\frac{\partial Y'}{\partial x_i}\frac{\partial\mu_H}{\partial x_i}+\frac{\partial Y'}{\partial x_i}\frac{\partial H'}{\partial x_i}=0
\end{aligned} \tag{8.80}$$

Taking the expected value of this equation and considering that $E[H']$=$E[Y']$ = 0 yields the equation for the mean:

$$\frac{\partial^2(\mu_H)}{\partial x_i^2}+\frac{\partial\mu_Y}{\partial x_i}\frac{\partial\mu_H}{\partial x_i}+E\left[\frac{\partial Y'}{\partial x_i}\frac{\partial H'}{\partial x_i}\right]=0 \tag{8.81}$$

Subtracting this equation from (8.80) yields the equation for the head perturbation:

$$\frac{\partial^2(H')}{\partial x_i^2}+\frac{\partial\mu_Y}{\partial x_i}\frac{\partial H'}{\partial x_i}+\frac{\partial Y'}{\partial x_i}\frac{\partial\mu_H}{\partial x_i}=E\left[\frac{\partial Y'}{\partial x_i}\frac{\partial H'}{\partial x_i}\right]-\frac{\partial Y'}{\partial x_i}\frac{\partial H'}{\partial x_i} \tag{8.82}$$

In the small perturbation approach it is assumed that perturbations *Y'* and *H'* are small. Hence, the products of these terms are even smaller. If these second order terms are neglected the following equations result for the mean and the perturbation:

$$\frac{\partial^2(\mu_H)}{\partial x_i^2} + \frac{\partial \mu_Y}{\partial x_i}\frac{\partial \mu_H}{\partial x_i} = 0 \tag{8.83}$$

$$\frac{\partial^2(H')}{\partial x_i^2} + \frac{\partial \mu_Y}{\partial x_i}\frac{\partial H'}{\partial x_i} + \frac{\partial Y'}{\partial x_i}\frac{\partial \mu_H}{\partial x_i} = 0 \tag{8.84}$$

To show how these equations can be solved we consider the one-dimensional case (Figure 8.5):

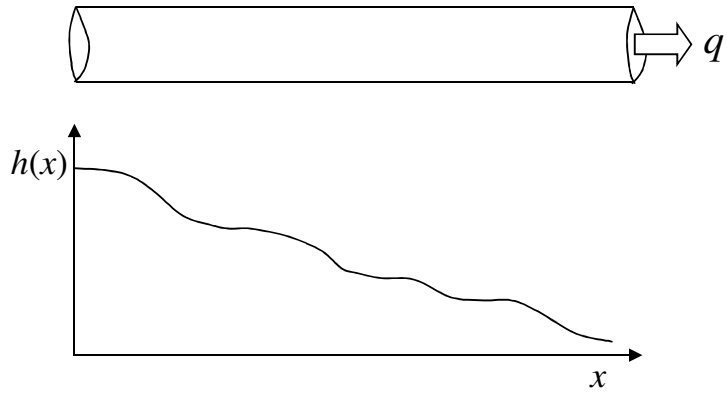



*Figure 8.5 One-dimensional flow in a tube with random conductivity and constant and known flux q.*

Figure 8.5 shows schematically the random head variation in one-dimensional flow in a tube due to random variation of hydraulic conductivity. Assuming that logconductivity ln*K* is described by a wide sense stationary function with constant mean, the equation for the mean is given by (8.83) with $\partial \mu_Y / \partial x = 0$:

$$\frac{\partial^2(\mu_H)}{\partial x^2} = 0 \tag{8.85}$$

It can be seen that the mean groundwater head is decribed by Laplace equation. Integrating one time yields:

$$\frac{\partial(\mu_H)}{\partial x} = c \tag{8.86}$$

To find the value of the constant $c$ we first return to Darcy's law and assume that the constant flux $q$ is known:

$$q = -K\frac{\partial H}{\partial x} \Rightarrow \frac{\partial H}{\partial x} = -\frac{q}{K}$$
$$\Rightarrow \frac{\partial H}{\partial x} = -qe^{-Y} \tag{8.87}$$

The Taylor approximnation of exp(-*Y*) around $\mu_Y$ is is given by:

$$e^{-Y} = e^{-\mu_Y} - e^{-\mu_Y}(Y-\mu_Y) + \frac{1}{2}e^{-\mu_Y}(Y-\mu_Y)^2 - ..... \tag{8.88}$$

Keeping the second order terms and taking expectations then yields:

$$\frac{\partial \mu_H}{\partial x} = -qe^{-\mu_Y}(1+\frac{1}{2}\sigma_Y^2) \tag{8.89}$$

Combining equation (8.86) and (8.89) then yields the value for the constant *c*. Integrating (8.89) once more yields the solution for the mean of *H*, if the value of $h(x=0) = h_0$ is known:

$$H(x) = h_0 - qe^{-\mu_Y}(1+\frac{1}{2}\sigma_Y^2)x \tag{8.90}$$

The variance can be obtained by writing the perturbations as spectral representations (see Equation 5.30):

$$Y'(x) = \int_{-\infty}^{\infty} e^{i\omega x} dX_Y(\omega) \tag{8.91}$$

$$H'(x) = \int_{-\infty}^{\infty} e^{i\omega x} dX_H(\omega) \tag{8.92}$$

where $dX_Y(\omega)$ and $dX_H(\omega)$ are the complex random amplitudes for frequency $\omega$ belonging respectively to the ln*K*(*x*) and *H*(*x*) random functions. The equation for the variance (8.84) with $\partial\mu_Y/\partial x = 0$ and setting $J = \partial\mu_H/\partial x$ becomes:

$$\frac{\partial^2(H')}{\partial x_i^2} = -J\frac{\partial Y'}{\partial x} \tag{8.93}$$

Substitution of the spectral representations in (8.93) gives:

$$\frac{\partial^2}{\partial x^2}\int_{-\infty}^{\infty} e^{i\omega x} dX_H(\omega) = -J\frac{\partial}{\partial x}\int_{-\infty}^{\infty} e^{i\omega x} dX_Y(\omega) \tag{8.94}$$

As the integrals are with respect to the frequencies and not the $x$, the differentials and the integrals can be interchanged leading to:

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} e^{i\omega x} dX_H(\omega) = -J \int_{-\infty}^{\infty} \frac{\partial}{\partial x} e^{i\omega x} dX_Y(\omega) \\ &\Rightarrow \int_{-\infty}^{\infty} -\omega^2 e^{i\omega x} dX_H(\omega) = -J \int_{-\infty}^{\infty} i\omega e^{i\omega x} dX_Y(\omega) \end{aligned} \tag{8.95}$$

Differentiation of (8.95) with respect to $\omega$ leads to the following relationship

$$\omega^2 dX_H(\omega) = Ji\omega dX_Y(\omega) \tag{8.96}$$

or

$$dX_H(\omega) = \frac{-JdX_Y(\omega)}{i\omega} \tag{8.97}$$

The following relationship holds between the complex amplitute and its complex conjugate:

$$E[dX(\omega_1)dX^*(\omega_2)] = \begin{cases} S(\omega)d\omega & \text{if } \omega_1 = \omega_2 = \omega \\ 0 & \text{if } \omega_1 \neq \omega_2 \end{cases} \tag{8.98}$$

By multiplying Equation (8.97) by its complex conjugate (noting that $dX_Y(\omega)$ and $dX_H(\omega)$ are independent) and taking the expectation, the following equation is obtained that relates the spectrum of hydraulic head to that of logconductivity:

$$S_H(\omega) = \frac{J^2 S_Y(\omega)}{\omega^2} \tag{8.99}$$

The variance of logconductivity is then obtained by applying equation (5.26) :

$$\sigma_H^2 = \int_{-\infty}^{\infty} S_H(\omega)d\omega = J^2 \int_{-\infty}^{\infty} \frac{S_Y(\omega)}{\omega^2} d\omega \tag{8.100}$$

An analytical solution of the integral (8.100) can be obtained when the covariance function is the hole effect model (see table 5.3), having the spectrum:

$$S_Y(\omega) = \frac{a^3 \sigma_Y^2 \omega^2}{\pi(1 + a^2\omega^2)^2} \tag{8.101}$$

The variance then becomes:

$$\sigma_H^2 = J^2 a^2 \sigma_Y^2 \tag{8.102}$$

Figure 8.6 shows the expected value of the head as given for $\mu_Y = 0, \sigma_Y^2 = 1$, $q$ = 0.5 m/d and $h(x_0)$ = 100 m for a domain length of 100 m (i.e. the head gradient is –0.75 m/m). Also given are the 95% confidence intervals of $H(x)$ as calculated from (8.102) with correlation parameter scale $a$ = 10 m (i.e. var[$H(x)$] = 28.13 m$^2$). Figure 8.6 shows that, apart from being approximations, the solutions also contain an inconsistency. To obtain a closed form solution to the expected head (8.90) we need to know $h(x_0)$. So the actual confidence interval close to $x_0$ should decrease to zero as indicated by the dashed line in Figure 8.6. Solution (8.102) is therefore only valid for large enough $x$, where the influence of the boundary condition is diminished. Boundary effects are treated extensively by Gelhar (1993).

Similar analyses can be performed on flow and transport in higher dimensions. Gelhar (1993) presents many examples of these. Solutions that are valid for larger variances can be found in Dagan (1998).

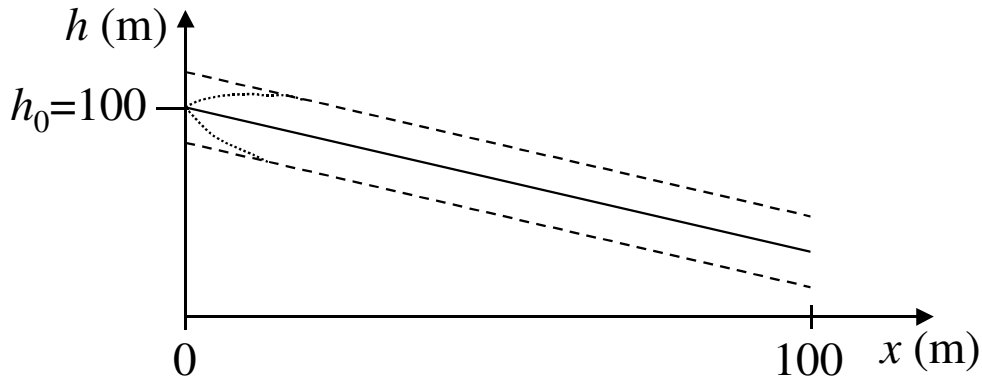



*Figure 8.6 Expected value of hydraulic head (solid line) calculated with Equation (8.90) and the 95%-confidence interval (dashed lines) according to Equation (8.102); the true confidence interval close to* x=*0 is given by the dotted lines.*

**b) Monte Carlo simulation**

If some of the conditions required for small perturbation analysis are not met, Monte Carlo analysis is an alternative. Consider two-dimensional flow in a heterogeneous porous media (Equation 8.66). Suppose that transmissivity is a random function of space and we are interested in obtaining the probability distribution of hydraulic head (or its moments, e.g. mean and variance) at each location. Monte Carlo simulation then proceeds as follows (see also section 7.5):

1) simulate a large number $M$ of realisations of the transmissivity random function $T(\mathbf{x})$. In case data on transmissivity are present, conditional geostatististical simulation is used, otherwise unconditional simulation. If the pdf $T(\mathbf{x})$ is not Gaussian, it could be transformed to a Gaussian pdf using a normal-score transform (see section 7.4.3).

Assuming a multivariate Gaussian distribution, the random function $Y(\mathbf{x}) = G[T(\mathbf{x})]$, can be simulated using sequential Gaussian simulation (sGs) (see section 7.5);

2) the $M$ realisations $y^{(k)}(\mathbf{x}), k = 1,..,M$ are transformed back to conductivity fields by taling the inverse of the normal-score transform: $T^{(k)}(\mathbf{x}) = G^{-1}[y^{(k)}(\mathbf{x})]$;
3) the numerical groundwater model solving (8.66) for the required boundary conditions is run $M$ times with each of the simulated transmissivity fields $T^{(k)}(\mathbf{x})$ as input, yielding $M$ realisations of hydraulic head: $h^{(k)}(\mathbf{x}), k = 1,..,M$ ;
4) from the $M$ realisations of hydraulic head the statistics (pdf, mean, variance, spatial covariance) of the (conditional or unconditional) head random function $H(\mathbf{x})$ can be estimated.

The great advantage of Monte Carlo simulation compared to analytical stochastic analysis is that it is very general. It can be applied for any model under any type of boundary condition, large variances of heterogenous parameter fields and, given the right geostatistical simulation procedure, for any kind of random function that is used. Also, conditioning on observations can be achieved quite naturally using Monte Carlo analysis, either directly (e.g. conditioning on transmissivity measurements by conditional simulation) or indirectly through inverse methods (e.g. conditioning on head and concentration measurements: Gómez-Hernández et al, 1998). Another advantage, one that is not often spoken of, is that Monte Carlo simulation is a technique that is very simple to apply and requires much less mathematical background than the analytical approaches.

A disadvantage of Monte Carlo simulation is obviously the required computation time, especially when models are large (contain a large number of grid points). As will be shown in the example hereafter, quite a large number of realisations have to be analysed to obtain stable estimates of higher order moments and the tail of the probability distributions, especially if the number of grid nodes is large. A considerable disadvantage of Monte Carlo simulation is that it will yield in case of many realisations an accurate estimate of the uncertainty, but it does not reveal in a straightforward manner what are the major causes of this uncertainty. For instance, the small perturbation analysis reveals through Equation (8.102) that apart from the variance of ln$K$ (which is obvious), also the mean head gradient and the correlation scale of ln$K$ determine the variance (i.e. uncertainty) of hydraulic head. Such information could only be obtained through monte Carlo analysis by performing a sensitivity analysis on top of it, e.g. repeating the Monte Carlo experiments for various settings of mean head gradient and variance and correlation scale of ln$K$. In large-scale problems, this is generally not feasible due to the enormous computation times involved.

In general, if insight into the nature of uncertainty propagation in some hydrological compartment is required (a more academic question), then analytical approaches are often used. Unconditional simulation is then applied to compare its results with the analytical solutions, i.e. to check whether the assumptions made to arrive at closed form analytical solutions are valid. In applied studies, where many of the assumptions underlying analytical approaches are not met, where measurements have to be taken into account and

where the goal is a single uncertainty measure of the model output, Monte carlo analysis (using conditional simulation) is the approach that is usually followed.

Example As an example two-dimensional steady state flow in a square domain of 1000×1000 m is considered. The boundary of the domain has a prescribed hydraulic head of 0.0 m. Groundwater recharge takes place at a rate of 1 mm per day. This model represents an idealized version of an island in the sea. Groundwater flow takes place through an aquifer with a constant thickness of 10 m. The domain is modelled with a numerical groundwater model (MODFLOW, McDonald and Harbaugh, 2000) using a constant cell size of 50 m. Depth averaged hydraulic conductivity is heterogenous and isotropic and its logarithm ln$k$ is modelled with a stationary and isotropic multivariate Gaussian random field $Y(\mathbf{x})$ with mean $\mu_Y = \ln 10$ (geometric mean: $K_G = \exp(\mu_Y) = 10\,\text{m/d}$) a variance $\sigma_Y^2 = 2$ and a spherical covariance function (Table 5.1) with parameter $a$=500 m. A single realisation (simulated with the program SGSIM from GSLIB (Deutsch and Journel, 1998)) of hydraulic conductivity is used as reality. Figure 8.7 shows the simulated reality (left) and the associated head field (right) as calculated by the groundwater model.

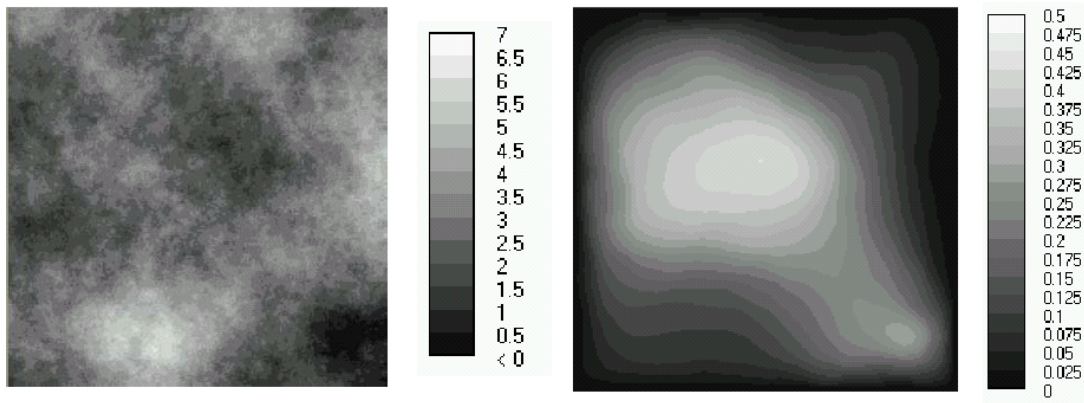



*Figure 8.7 Simulated reality: logarithm of hydraulic conductivity (left) and hydraulic head (m)*

If the hydraulic head is calculated with an assumed homogeneous hydraulic conductivity of $k = K_G = 10\,\text{m/d}$ it can be seen from Figure 8.8 that the results are far from the assumed reality. Hence, some form of uncertainty analysis using a random function to model heterogeneity is in order.

In the following the results from a Monte Carlo analysis are demonstrated. To investigate the effect of the number of observations on the head uncertainty, four sets of datapoints are sampled from the "real" field of hydraulic conductivity of Figure 8.7. In a real application usually a limited budget is available to take samples or to perform well or pumping tests to measure the depth averaged hydraulic conductivity at a number of locations. The four sets of data contain 9, 16, 25 and 36 observations respectively. The observations are taken on a square regular grid (3×3, 4×4, 5×5 and 6×6). Monte Carlo analysis was thus performed five times: one time for unconditional simulation, and four

times with conditional simulation using the four data sets. The results are shown in Figures 8.9 and 8.10. Figure 8.9 shows examples of the realisations of $\ln K(\mathbf{x})$ and associated hydraulic head calculated with the $\ln K(\mathbf{x})$ realisations as input. Realisations in the top row relate to unconditional simulation, the middle row to conditional simulation with 9 observations and the bottom row to conditional simulation with 36 observations. Figure 8.10 shows the mean head field and its standard deviation as obtained from 1000 realisations. Again, the upper row presents the results from conditional simulation and the middle and lower rows from conditional simulation with 9 and 36 observations respectively. These figures show that as the number of conditioning data increases, that both the individual realisations as well as the mean head field are closer to the true one. Also, the variance decreases quite significantly if the number of conditioning data increases. This is also shown in Figure 8.11 where the average head standard deviation of the model domain is plotted against the number of conditioning points. Obviously, the effect of conditioning data on the reduction of head variance depends on the correlation scale. If the correlation scale is small, many observations are needed to achieve a given reduction in uncertainty.

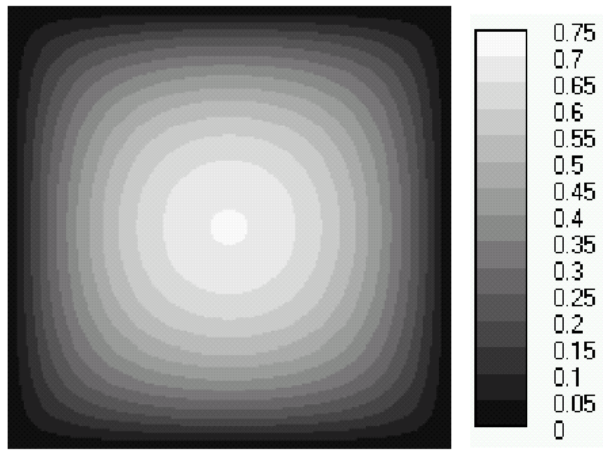



*Figure 8.8 Hydraulic head (m) calculated with a presumed homogenous porous media with hydraulic conductivity* $k = K_G = 10\,\text{m/d}$

Previously, the required number of realisations was discussed. To show the effect of the number of realisations on the accuracy of the uncertainty estimates, Figure 8.12 is added. This figure shows the estimate of the mean head and the head standard deviation in case of unconditional simulation using 100 instead of 1000 realisations. It can be seen that 100 realisations is sufficient to obtain accurate (stable estinates) of the mean, but far from sufficient to estimate the head variance. In fact, even in case of 1000 realisations the estimate of the standard deviation still looks asymmetric for the unconditional case. This suggests that even more realisations are needed.[5]

[5] In case of groundwater flow with and random hydraulic conductivity, generally the largest variance between hydraulic head realisations occurs when the correlation length is such that zones of high conductivity cross the domain with 50% probability, such that half of the realisations show shortcuts and half don't. For stationary and multivariate Gaussian random fields this is generally the case for correlation lengths between 0.5 and 1 times the domain size.

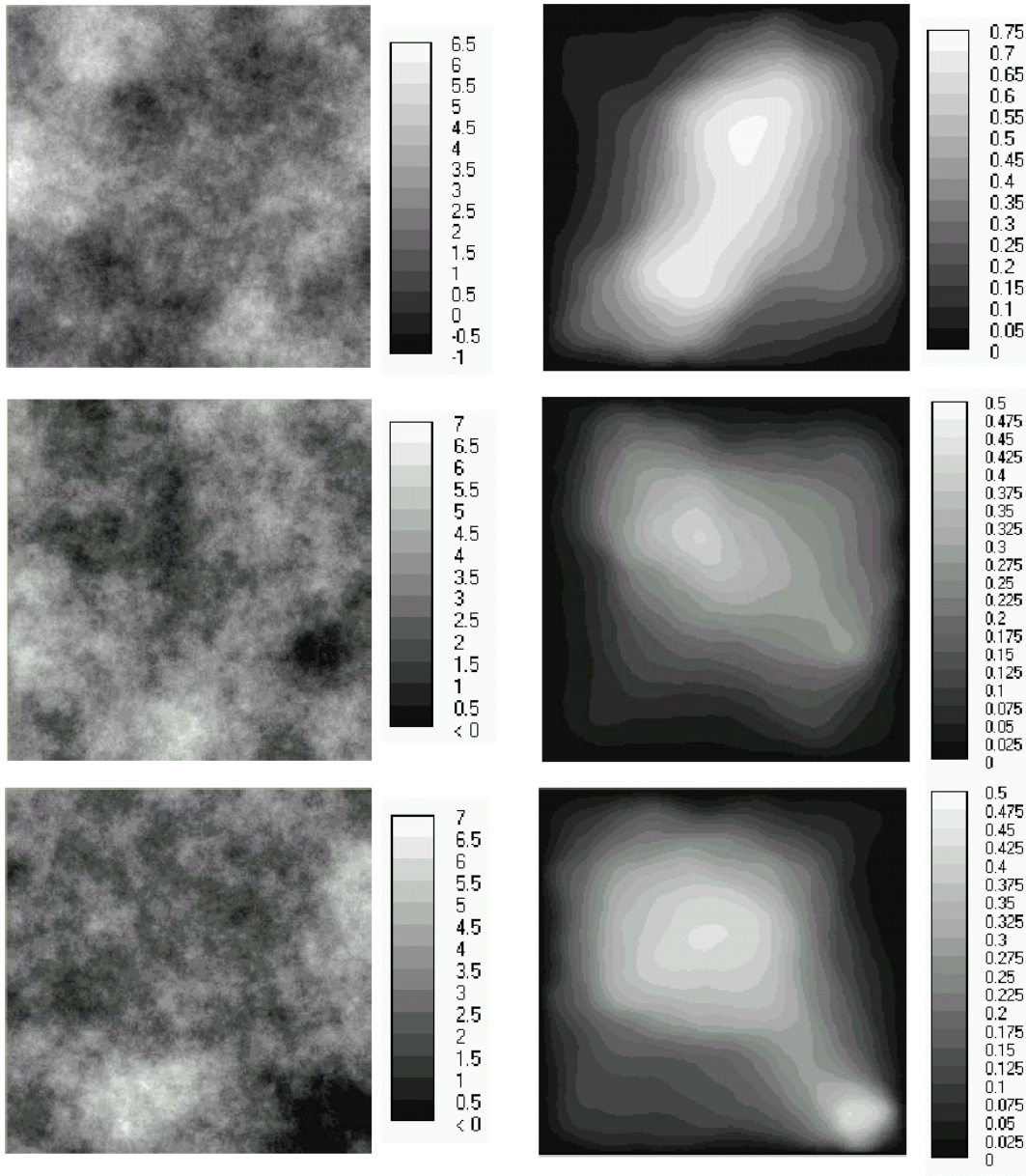



*Figure 8.9 Realisations of lnK(**x**) (left column) and associated hydraulic head (right column in m/d); top row: unconditional simulation; middle row: conditional simulation with 9 observations; bottom row: conditional simulation with 36 observations.*

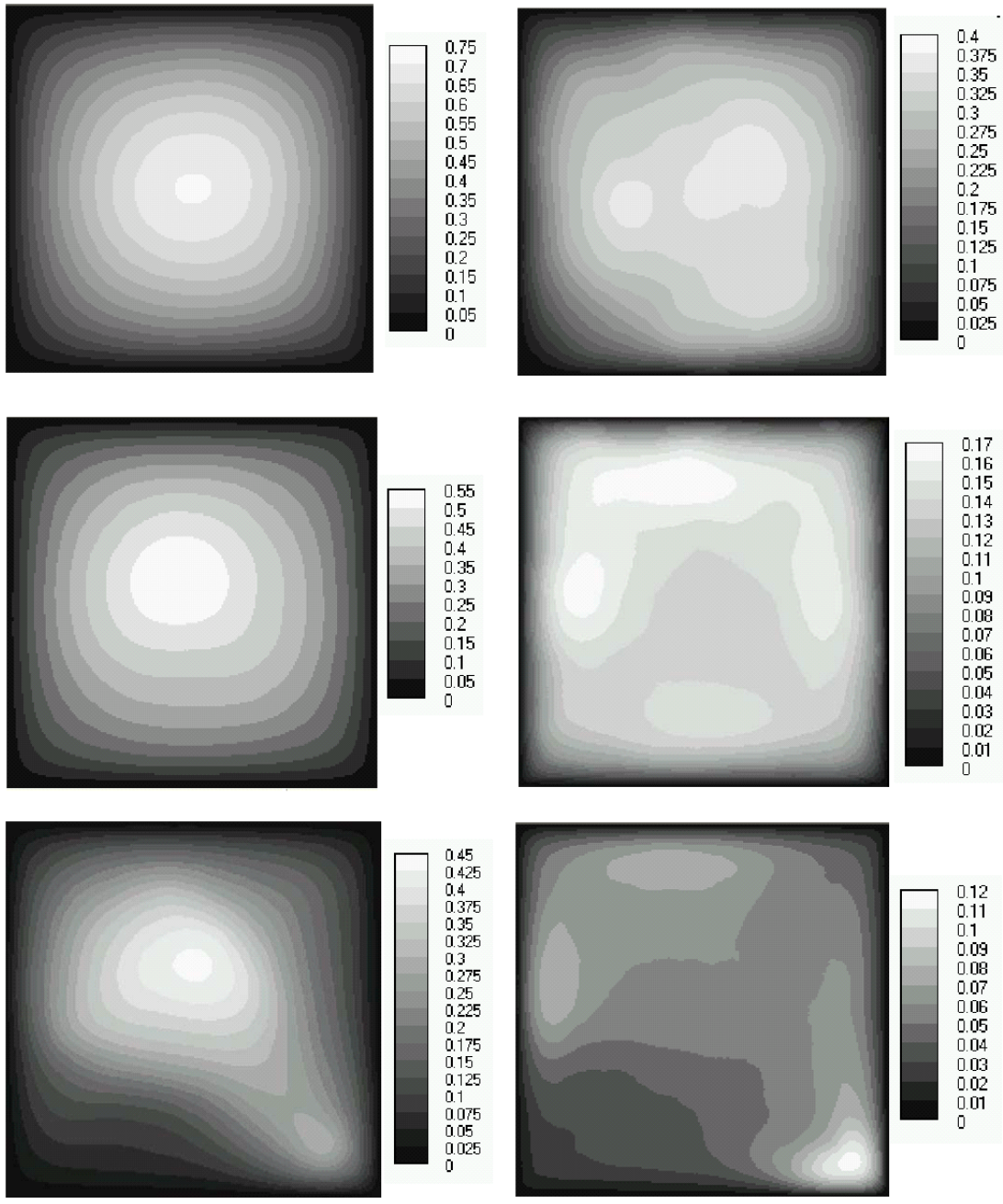



*Figure 8.10 Mean head (left colum) and head standard deviation (right column) obtained from 1000 realisation of lnK(**x**); top row: unconditional simulation; middle row: conditional simulation with 9 observations; bottom row: conditional simulation with 36 observations.*

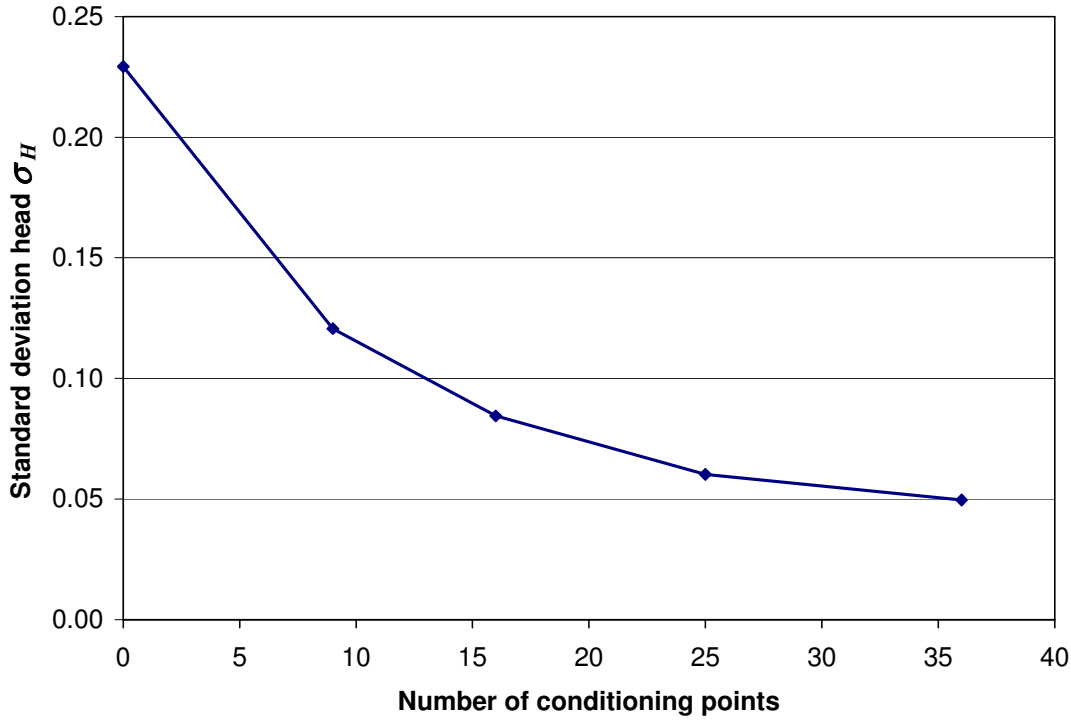



*Figure 8.11 Relationship between the domain average standard deviation of hydraulic head and the number of observations used for conditioning (zero observations means unconditional simulation).*

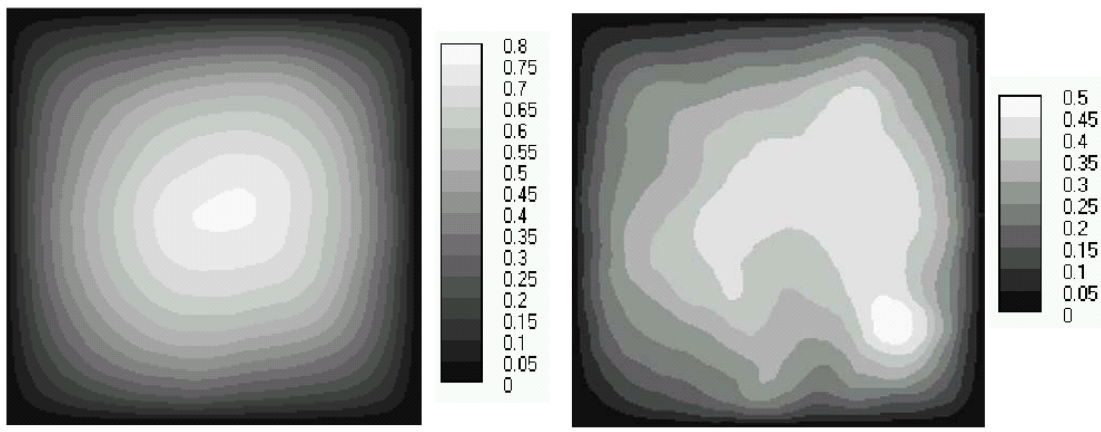



*Figure 8.12 Mean head and head standard deviation in case of unconditional simulation using 100 instead of 1000 realisations.*

## 8.8 Exercises

8.1. Consider the following fitted relation between concentration of some pollutant $C$ and discharge $Q$:

$$C = a - b\ln(Q)$$

with $Q$ a random variable with pdf $f_Q(q)$. Give an expression for the pdf of $C$?

8.2. For the same relationship as given in question 21: The mean and variance of $Q$ are 1.5 $m^3/s$ and 0.01 $m^6/s^2$, respectively. $C$ is expressed in mg/l and $a$ = 200 mg/l and $b$=20 mg/l. What are the mean and variance of $C$, based on a first order Taylor approximation? What is the mean of $C$ based on a second order Taylor approximation?

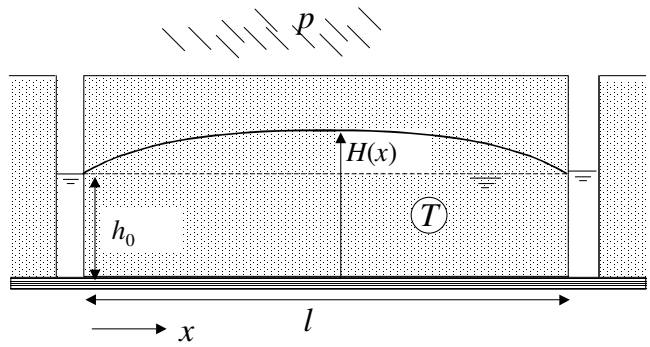
8.3. Consider the same relationship as used in Question 21, but now with known discharge $q$ and random parameters $A$ and $B$:

$$C = A - B\ln(q)$$

with the following statistics: $A: \mu_A = 200, \sigma_A = 10$ and $B$: $\mu_B = 20, \sigma_A = 1$ and correlation coefficient $\rho_{AB} = 0.7$. Calculate the mean and the variance of $C$?

8.4. The long term average variation of the phreatic level between two water courses (see Figure) at a distance $l$ apart is given by the following equation:

$$H(x) = h_0 + \frac{p}{T}(lx - x^2) \quad 0 \le x \le l$$




8.5.

with

| | |
|---|---|
| $H(x)$: | groundwater elevation (m) as a function of location $x$ |
| $L$: | distance between water courses: 200 m |
| $h_0$: | water levels in water courses: 1 m |
| $p$: | long year average groundwater recharge (0.001 m/d) |
| $T$: | Transmissivity of the aquifer ($m^2$/d). |

The transmissivity is not exactly known and is treated as random variable with mean $\mu_T = 10\,\mathrm{m}^2/\mathrm{d}$ and a variance of $\sigma_T^2 = 4\,\mathrm{m}^4/\mathrm{d}^2$.

a. Use a first order Taylor approximation to estimate the mean $\mu_H(x)$ and the variance $\sigma_H^2(x)$ of hydraulic head for $x$ = 0, 25, 50, 75,100,125,150,175,200 m. Make a plot of mean head and the 95%-confidence interval, assuming $H(x)$ Gaussian distributed.
b. Explain why the width of the confidence interval changes with location $x$.

8.6. Consider the following differential equation describing the concentration $C$ of some pollutant in a lake of known volume $v$, with constant but random influx $Q_{\text{in}}$ and known decay coefficient $k$:

$$v\frac{dC}{dt} = -kC + Q_{\text{in}}$$

Derive an expression for the mean and the variance of the concentration.

# 9. Optimal state prediction and the Kalman filter

## 9.1 Introduction

As discussed in the previous chapters, the dynamic behavior of hydrological systems is often described by a model. This can either be a time series model, or a differential equations based on physical processes. Due to simplification, discretization and schematization, the model is only an approximation of reality. Moreover, parameters and input variables are not exactly known. They are subject to uncertainty. One of the objectives of stochastic hydrology is to quantify statistically the difference between the model predictions and reality, or in other words to quantify the predictive uncertainty.
A system is called causal, if the system variables depend only on past and present input. For example, a river discharge at some time ($t=\tau$) depends on rainfall up to that time ($-\infty<t\leq\tau$), while future rainfall ($t>\tau$) does not effect the discharge at $t=\tau$. Likewise, the system's future behavior (for $t=\tau+\ell$, $\ell>0$) only depends on the system variables at time ($t=\tau$) and the future inputs ($\tau<t\leq\tau+\ell$). In causal systems we can define the state of the system (see also section 1.3.1) as a set of values of the systems variables, such that all information of the past of the system, relevant for the future behavior is embedded in this set of values. According to this definition, the future behavior of the system is completely determined by the present state and the future input. An example of a state in the field of dynamic groundwater flow is the spatial pressure distribution at some point is time. In the governing differential equation for groundwater flow, the state appears as initial condition and the input is formed by the boundary conditions.
At successive points in time, we can predict the unknown state (seen here as the unknown realization of set of random variables, chapters 3 and 5) and the corresponding uncertainty, for instance with stochastic modeling or time series modeling. At discrete points in time and space, the state of the system can be observed and we can compare the state prediction with the observation. If all elements of the state are observed and observation errors can be neglected, we can replace the state prediction by the observation. However, in practice often the observation error may be significant. Moreover not all elements of the state can be observed directly. Therefore, we have to deal with two sources of information to predict the inknown state, the model prediction and the observation, both with uncertainty. The aim of optimal state prediction is that we optimally combine both sources of information, given the uncertainty of both sources. In this chapter we focus on the linear Kalman Filter, which is a powerful method in optimal state prediction. First, in section 9.2, we discuss the principles of Kalman Filtering and we present the associated state estimation algorithm. In sections 9.3 and 9.4 it is demonstrated how the Kalman Filter can be applied to respectively time series models and spatially distributed process models. Finally in section 9.5 some applications in hydrological practice are discussed.

## 9.2 Principles of Kalman filtering

### 9.2.1 State equation and measurement equation.

In this chapter we restrict ourselves to systems of which the evolution of the state can be described by the linear state equation (9.1) in discrete time:

$$\mathbf{z_t} = \mathbf{A}_t\mathbf{z}_{t-1} + \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t \tag{9.1}$$

Where $\mathbf{z_t}$ the state vector at time *t*.
$\mathbf{A}_t$ the parameter matrix, relating the state vector at time *t* to the state vector at time *t*-1.
$\mathbf{u}_t$ the input vector, representing the known input (driving forces) at time *t*.
$\mathbf{B}_t$ the parameter matrix, relating the state vector at time **t** to input at time *t*
$\mathbf{w}_t$ the system noise vector, representing all influences that are not described by the model (first two terms on the right hand side of (9.1)). The system noise includes also unknown inputs, and errors in the parameter matrices. Sometimes in literature the system noise is referred to as model error.

Systems that can be described by matrices that are independent of time ($\mathbf{A}_t$=**A** and $\mathbf{B}_t$=**B)** are called time invariant systems.

The observation process can be described by the measurement equation (9.2):

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{z}_t + \mathbf{v}_t \tag{9.2}$$

Where $\mathbf{y}_t$ is the measurement vector at time *t*, containing the observations.
$\mathbf{H}_t$ the measurement matrix, relating the state at time *t* to the measurement vector at time *t*
$\mathbf{v}_t$ the measurement error vector at time *t.*

If the monitoring system doesn't change with time, the measurement matrix is time invariant $\mathbf{H}_t = \mathbf{H}$.

### 9.2.2 Optimal state prediction algorithm.

In the optimal state prediction algorithm, commonly referred to as the Kalman filter, we use the following definitions and assumptions:

- the conditional prediction of the state vector at time *t* given observations up to and including time *t* and the corresponding error covariance matrix are denoted as:

$$\hat{\mathbf{z}}_\mathbf{t} = \mathbf{E}[\mathbf{z}_{t|t}] \text{ and } \mathbf{P}_t = \mathbf{E}[(\mathbf{z}_t - \hat{\mathbf{z}}_t)(\mathbf{z}_t - \hat{\mathbf{z}}_t)^{\mathrm{T}}] = \mathbf{E}[\mathbf{e}_{t|t}\mathbf{e}_{t|t}^{\mathrm{T}}] \tag{9.3}$$

- The conditional prediction of the state vector at time $t$ given observations up to and including time $t$-1 and the corresponding error covariance matrix are denoted as:

$$\bar{\mathbf{z}}_t = \mathbf{E}[\mathbf{z}_{t|t-1}] \text{ and } \mathbf{M}_t = \mathbf{E}[(\mathbf{z}_t - \bar{\mathbf{z}}_t)(\mathbf{z}_t - \bar{\mathbf{z}}_t)^{\mathrm{T}}] = \mathbf{E}[\mathbf{e}_{t|t-1}\mathbf{e}_{t|t-1}^{\mathrm{T}}] \tag{9.4}$$

Furthermore, we assume the system noise as well as the measurement error to be mutually independent white noise processes with known statistics:

$$\begin{aligned}
&\mathbf{E}[\mathbf{w}_t] = \mathbf{0} \\
&\mathbf{E}[\mathbf{w}_t \mathbf{w}_\tau^{\mathrm{T}}] = \begin{cases} \mathbf{Q}_t & \text{if} \quad t = \tau \\ \mathbf{0} & \text{if} \quad t \neq \tau \end{cases} \\
&\mathbf{E}[\mathbf{v}_t] = \mathbf{0} \\
&\mathbf{E}[\mathbf{v}_t \mathbf{v}_\tau^{\mathrm{T}}] = \begin{cases} \mathbf{R}_t & \text{if} \quad t = \tau \\ \mathbf{0} & \text{if} \quad t \neq \tau \end{cases} \\
&\mathbf{E}[\mathbf{w}_t \mathbf{v}_\tau^{\mathrm{T}}] = \mathbf{0}
\end{aligned} \tag{9.5}$$

Let the conditional prediction of the state vector at time $t$-1 given observations up to time $t$-1 ($\hat{\mathbf{z}}_{t-1}$) and the corresponding error covariance matrix ($\mathbf{P}_{t-1}$) be known. Given the model (known matrices $\mathbf{A}_t$ and $\mathbf{B}_t$) and the known input $\mathbf{u}_t$ , the conditional prediction of the state at time $t$ given the observations up to time $t$-1 follows from equation (9.1):

$$\bar{\mathbf{z}}_t = \mathbf{A}_t \hat{\mathbf{z}}_{t-1} + \mathbf{B}_t \mathbf{u}_t \tag{9.6}$$

Note that from the statistics (9.5) it follows that $\mathbf{E}[\mathbf{w}_{t|t-1}] = \mathbf{0}$.

The conditional prediction (9.6) is called the time up-date. The prediction error is the p is obtained by subtracting (9.6) from (9.1):

$$\mathbf{e}_{t|t-1} = \mathbf{z}_t - \bar{\mathbf{z}}_{t-1} = \mathbf{A}(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1}) + \mathbf{w}_t = \mathbf{A}\mathbf{e}_{t-1|t-1} + \mathbf{w}_t \tag{9.7}$$

From (9.7) is can be seen that the error of the prediction is a function of the error of the conditional estimate at the previous time step ($t$-1) and the system noise at time **t.** If the model (matrices $\mathbf{A}_t$ and $B_t$) are well calibrated, the expected value of the error terms is equal to **0**. The covariance of the prediction error can be derived from (9.7):

$$\mathbf{M}_t = \mathbf{E}[\mathbf{e}_{t|t-1}\mathbf{e}_{t|t-1}^{\mathrm{T}}] = \mathbf{A}_t \left\{\mathbf{E}[\mathbf{e}_{t-1|t-1}\mathbf{e}_{t-1|t-1}^{\mathrm{T}}]\right\}\mathbf{A}_t^{\mathrm{T}} + \mathbf{Q}_t = \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^{\mathrm{T}} + \mathbf{Q}_t \tag{9.8}$$

At time $t$ the observations $\mathbf{y}_t$ become available, and we like to correct the prediction with these observations. This can be done by calculating the conditional prediction of the state vector at time $t$ given observations up to and including time $t$. It can be proven that we get

the minimum error covariance if the correction step (also called measurement update) is done by:

$$\hat{\mathbf{z}}_{t|t} = \bar{\mathbf{z}}_{t|t-1} + \mathbf{K}_t\{\mathbf{y}_t - \mathbf{H}_t\bar{\mathbf{z}}_{t|t-1}\} \tag{9.9}$$

With

$$\mathbf{K}_t = \mathbf{H}_t\mathbf{M}_t\{\mathbf{H}_t\mathbf{M}_t\mathbf{H}_t^{\mathrm{T}} - \mathbf{R}_t\} \tag{9.10}$$

The matrix $\mathbf{K}_t$ is called the Kalman gain.

Finally with the error statistics (9.5) and the equations (9.9) and (9.10) it can be proven that the covariance matrix of the conditional state estimate at time **t** given observations up to time *t* equals:

$$\mathbf{P}_t = \{\mathbf{I} - \mathbf{K}_t\mathbf{H}_t\}\mathbf{M}_t \tag{9.11}$$

The measurement up-date is the optimal linear estimate of the state at time *t* given observations up to time *t* (minimal error variance). With the measurement up-date (9.9) and the corresponding covariance matrix (9.11) we can repeat the calculation for time step *t*+1. The state prediction algorithm is summarized in the figure below.

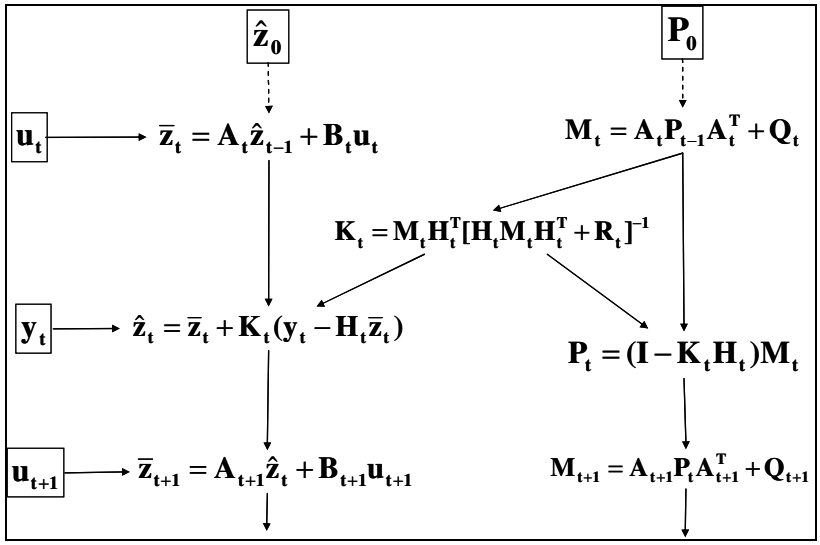



*Figure 9.1 Linear Kalman Filter algorithm*

Note that:

- In hydrology we often we have time invariant systems ($\mathbf{A}_t$= **A**, $\mathbf{B}_t$=**B**, $\mathbf{Q}_t$=**Q** and $\mathbf{R}_t$=**R**).
- In practice, the matrices $\mathbf{A}_t$, $\mathbf{B}_t$, $\mathbf{Q}_t$ and $\mathbf{R}_t$ have to be estimated or calibrated. However, the calibration is beyond the scope of this chapter. More information can be found in in for instance Van Geer et al. (1992).

## 9.3 Kalman filtering and time series

### 9.3.1 Kalman filter algorithm for transferfunction/noise models.

In this section we apply the Kalman Filter to a simple transfer/noise model (see chapter 6). As will be seen, this leads to a simple scalar example that is very informative about the properties of the Kalman filter. Consider the transfer/noise model:

$$Z_t = \delta_1 Z_{t-1} + \omega_0 X_t + a_t \tag{9.12}$$

Equation (9.12) is a scalar form of the state equation (9.1) where:

$$\mathbf{z}_t = Z_t, \quad \mathbf{A} = \delta_1, \quad \mathbf{B} = \omega_0, \quad \mathbf{u}_t = X_t, \quad \mathbf{w}_t = a_t, \quad \mathbf{Q}_t = \sigma_a^2 \tag{9.13}$$

In chapter 6 we neglected measurement errors, but here we account for measurement errors. At observation times the measurement equation is:

$$Y_t = Z_t + v_t \tag{9.14}$$

Where we have:

$$\mathbf{y}_t = Y_t, \quad \mathbf{H}_t = 1, \quad \mathbf{v}_t = v_t, \quad \mathbf{R}_t = \sigma_v^2 \tag{9.15}$$

Suppose we have an observation at each time step and the input $X_t$ as well as the initial conditions $\hat{Z}_{t-1}$ and $P_{t-1}$ are known. Then, substitution of the quantities (9.13) and (9.15) into the algorithm of section 9.2 yields:

| | |
|---|---|
| Time update: | $\overline{Z}_t = \delta_1 \hat{Z}_{t-1} + \omega_0 X_t$ |
| | $M_t = \phi_1^2 P_{t-1} + \sigma_a^2$ |
| Measurement up-date: | $K_t = \dfrac{M_t}{M_t + \sigma_v^2}$ |
| | $\hat{Z}_t = \overline{Z}_t + K_t\{Y_t - \overline{Z}_t\}$ |
| | $P_t = \{1 - K_t\}M_t$ |

(9.16)

### 9.3.2 Properties of the linear Kalman Filter applied to a simple Transferfunction/noise model.

Balance

From (9.16) we can see that the Kalman Filter balances the information from the model prediction and the observation.

- One extreme is that we have perfect observations ($\sigma_v^2 = 0$). It follows from (9.16) that $K_t$=1 and therefore: $\hat{Z}_t = Y_t$. The best estimate of the state at time *t* is equal to the observation at that time.
- The other extreme is that we have a perfect model ($\sigma_a^2 = 0$), it follows that the Kalman gain $K_t$=0 and $\hat{Z}_t = \overline{Z}_t$. In this case we keep the model prediction and the observation does not contain any additional information.
- If the uncertainty of both the model and the observation are equal ($\sigma_a^2 = \sigma_v^2$) it follows that $K_t$ = ½ and the best state estimate is the average of the model prediction and the observation.

In general, it can be stated that $0 \le K_t \le 1$. If we are more uncertain about the model, the Kalman gain will tend to 1, and if the observations are subject to large uncertainty, the Kalman gain tends to 0. Therefore we can consider the Kalman gain as a mechanism that balances the model and the observation, according to the relative uncertainty. Because the Kalman gain is always in between 0 and 1, the measurement up-date ($\hat{Z}_t$) is always in between the time up-date ($\overline{Z}_t$) and the observation ($Y_t$).

Magnitude of the measurement up-date.

From the algorithm (9.16) it can also be seen that the magnitude of the measurement up-date depends on the difference between the time up-date and the observation (= innovation $Y_t - \overline{Z}_t$). If the observation is far away from the time up-date ($|Y_t - \overline{Z}_t|$ is large), the correction in the measurement up-date is large. If the observation is close to the time up-date ($|Y_t - \overline{Z}_t|$ is small), the correction in the measurement up-date is small.

Uncertainty reduction

Because the Kalman gain is a positive number in between 0 and 1, the last equation from algorithm (9.16) shows that the variance of the measurement up-date is always smaller than (or equal to) the variance of the time up-date ($P_t \le M_t$). Also it can be proven that the variance of the measurement up-date is always smaller than the variance of the measurement error ($P_t \le \sigma_v^2$). Combining the model with observations yields a reduction of the uncertainty relative to the prediction with the model (time up-date) as well as to the observation.

Non observed time steps.

The algorithm (9.16) is valid if we have observations at each time step. In practice, we may not have observations at certain points in time and we can't calculate the measurement up-date. At non-observed time steps, the algorithm reduces to (see also the forecast in chapter 6):

$$\begin{aligned} \overline{Z}_t &= \delta_1 \hat{Z}_{t-1} + \omega_0 X_t \\ M_t &= \delta_1^2 P_{t-1} + \sigma_a^2 \end{aligned} \qquad (9.17)$$

$$\hat{Z}_t = \overline{Z}_t$$
$$P_t = M_t$$

In figure 9.2 we show the effect of observed and non-observed time steps. In between the observation times, the error variance grows until the next observation time. At the observation time the error variance drops below the variance of the observation error.

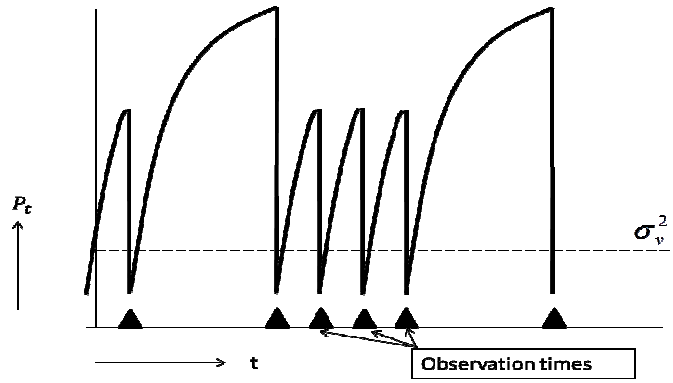



*Figure 9.2. Error variance ($P_t$) at observed and non-observed time steps.*

State estimation with Kalman Filter compared to open loop

The effect of state prediction with Kalman Filtering can be illustrated by comparison with an open loop prediction. An open loop prediction is a state prediction without using observations. Figure 9.3 shows the comparison for a time series model of the form (9.12). In this illustration the Kalman Filter uses observations at each time step. Figure 9.3.

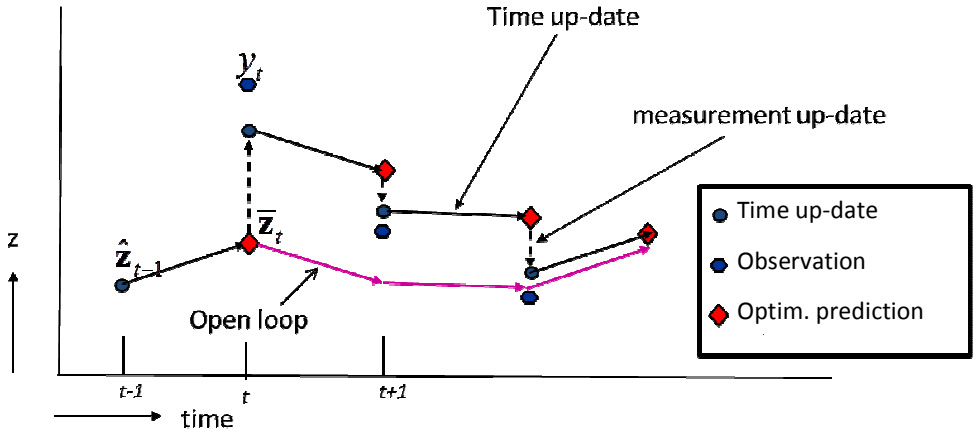



*Figure 9.3. Comparison state prediction using Kalman Filtering and open loop.*

From figure 9.3 it can be seen that at each observation time, the optimal prediction (measurement up-date) is in between the time up-date and the observation. The open loop prediction uses only the input $X_t$ and it doesn't use the observations. The time up-date is

similar as in the Kalman Filter, but there is no measurement up-date. The most important consequence is that for the open loop the errors propagate in time. If at some point in time the prediction of $Z_t$ is too small, it tends to persist in being too small for the next time steps. The errors in the open loop prediction are correlated in time, whereas for the Kalman Filter, it can be proven that the errors for the measurement up-date are uncorrelated in time and the predictions are closer to the observations.

## 9.4 Example groundwater head time series.

Many time series of shallow groundwater head show a seasonal pattern, which is driven by the seasonal behavior of precipitation and evapotranspiration (see figure 9.4). The groundwater observation frequency is 24 time per year, which yields an observation interval of ca. 15.2 days. This interval is taken as the time step in the time series model. For each time step, we calculate the average precipitation excess from the available observations of precipitation and evapotranspiration. The groundwater head is described by the Transfer/noise model:

$$\begin{aligned} H_{1,t} &= \delta_1 H_{1,t-1} + \omega_0 P_t \\ n_t - c &= \phi_1 (n_{t-1} - c) + a_t \\ H_t &= H_{1,t} + n_t \end{aligned} \qquad (9.18)$$

With:

$H_t$ : the groundwater depth below surface at time $t$,

$p_t$ : the precipitation surplus (precipitation minus evaporation) during the time step from *t-1* to $t$,

$H_{1,t}$ : the component of the groundwater head due to the precipitation excess

$n_t$ : the noise component at time $t$,

$a_t$ : the white noise at time $t$,

$\delta_1$, $\omega_0$, $\phi_1$ : parameters of the transfer/noise model,

$c$ : the long term average groundwater depth in case $p_t$=0.

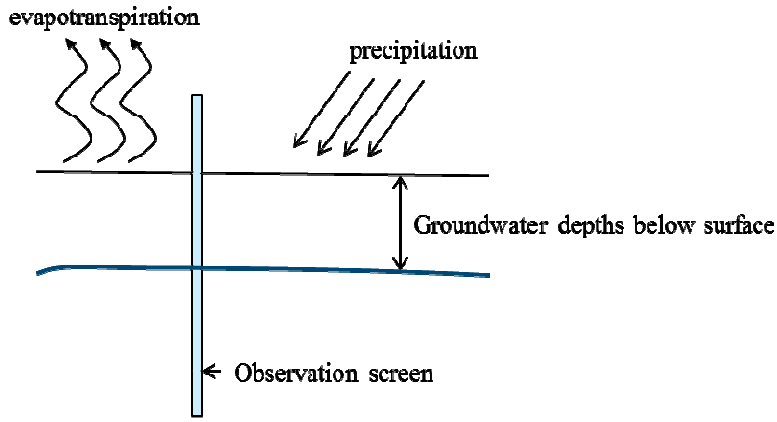



*Figure 9.4 Situation around a groundwater observation.*

To formulate the transfer/noise model in term of the state equation of the Kalman Filter, we define the state vector, the input vector and the system noise vector respectively as:

$$\mathbf{z}_t = \begin{bmatrix} H_{1,t} \\ n_t \end{bmatrix}, \qquad \mathbf{u}_t = [P_t] \quad \text{and} \quad \mathbf{w}_t = [a_t] \tag{9.20}$$

Using the definitions (9.13) the state equation of the form (9.1) for the model (9.12) becomes:

$$\mathbf{z}_t = \begin{bmatrix} \delta_1 & 0 \\ 0 & \phi_1 \end{bmatrix} \mathbf{z}_{t-1} + \begin{bmatrix} \omega_0 \\ 0 \end{bmatrix} \mathbf{u}_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathbf{w}_t \tag{9.21}$$

We have observation of $h_t$ every time step. The measurement equation is:

$$Y_t = H_t + v_t \tag{9.22}$$

Where: $Y_t$ is the observation at time $t$
$v_t$ is the measurement error at time $t$.

The measurement error standard deviation is dependent on the equipment used. For monitoring shallow groundwater we estimated the measurement error standard deviation $\sigma_v = 1.0\,cm$. The parameters of the time series model, including the standard deviation of the system noise are estimated as:

$$\delta_1 = 0.865, \quad \omega_0 = 0.515(cm/mm), \quad \phi_1 = 0.619(days), \quad c = 23(cm), \quad \sigma_a = 20.4(cm)$$

In Figure 9.5 the results of the state prediction with respectively open loop and Kalman Filter are visualized. The open loop estimate is the forecast using the observations of the precipitation excess and the transfer function. In the open loop, the groundwater observations are not used, whereas in the state prediction of the Kalman Filter, we have an observation of the groundwater head every third time step.

Although both predictions in figure 9.5 resemble the seasonal pattern quite well, it is clear that the Kalman Filter predictionis more close to the observations. This effect will be even stronger if we applied a higher observation frequency.

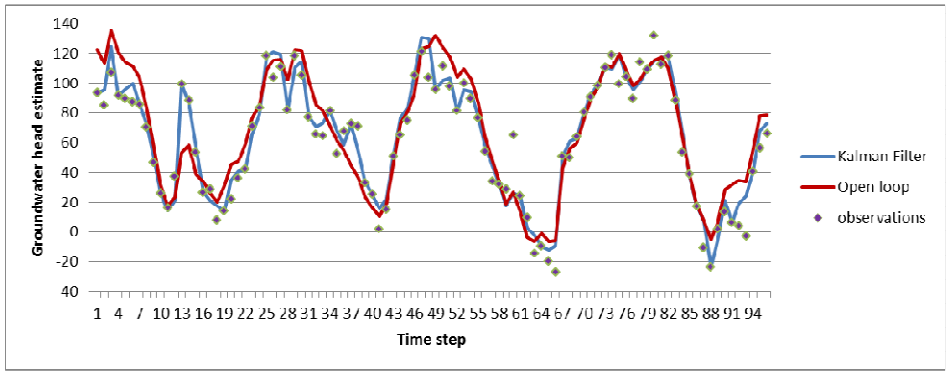



*Figure 9.5. State estimation Kalman filter and open loop estimation compared to groundwater head observations.*

The differences between the observations and respectively the Kalman filter prediction and the open loop prediction are show in Figure 9.6. From this figure it is even more clear that the Kalman filter predictions are closer to the observations than the open loop predictions. Moreover, the differences of the open loop predictions show more persistence in time. In other words, the errors of the open loop prediction show more temporal correlation than the Kalman Filter prediction errors, meaning that more information remains unused.

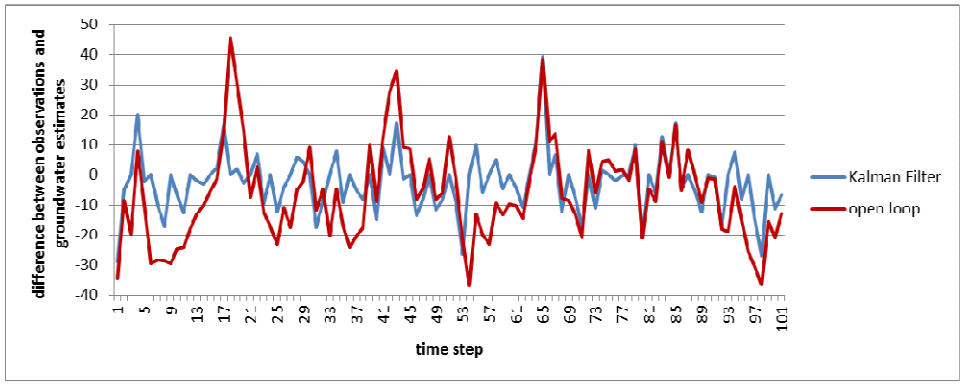



*Figure 9.6. Difference between observations and respectively the Kalman Filter estimate and the open loop estimate.*

## 9.5 Kalman Filtering and spatial distributed systems

In section 9.4 we discussed the application of the Kalman Filter to a scalar time series model. However, Kalman Filtering can be applied to any system that can be described by a linear difference equation. In this section, we show how the Kalman Filter can be applied to a simple spatially distributed groundwater system.

### 9.5.1 Example uncertainty for a one-dimensional groundwater flow.

We consider a one-dimensional groundwater system as shown in figure 9.7. The system can be described with the differential equation (9.20). The groundwater head is a function of a spatial *x*-co-ordinate and time *t*: *h(x,t).* The left hand side boundary condition is the surface water level $H_l$ and the right hand side boundary condition is $H_r$.

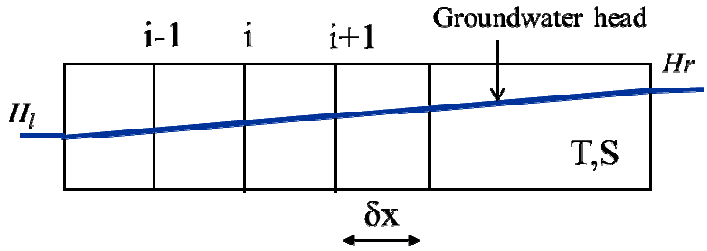



*Figure 9.7. One-dimensional groundwater system.*

The governing differential equation is:

$$S\frac{\partial H}{\partial t} = -T\frac{\partial^2 H}{\partial x^2} \tag{9.23}$$

Where: $H$ is the groundwater head (a function of space coordinate $x$ and time $t$)
$S$ is the storage coefficient
$T$ is the transmissivity

Equation (9.23) can be discretized using the following difference approximations:

$$\frac{\partial H}{\partial t} \approx \frac{H_{i,t-1} - H_{i,t}}{\delta t} \tag{9.24}$$

and

$$\frac{\partial^2 H}{\partial x^2} \approx \frac{\left(H_{i+1,t} - H_{i,t}\right) - \left(H_{i,t} - H_{i,-1t}\right)}{(\delta x)^2} \tag{9.25}$$

Substitution of (9.24) and (9.26) in (9.23) yields:

$$(1+2/\alpha)H_{i,t} - (1/\alpha)H_{i-1,t} - (1/\alpha)H_{i+1,t} = H_{i,k-1} \tag{9.26}$$

with:

$$\alpha = \frac{S(\delta x)^2}{T\delta t} \tag{9.27}$$

If there are $N$ grid nodes, we define the state vector as:

$$\mathbf{h_t} = \begin{bmatrix} H_{1,t} \\ \mathsf{M} \\ H_{i-1,t} \\ H_{i,t} \\ H_{i+1,t} \\ \mathsf{M} \\ H_{N,t} \end{bmatrix} \tag{9.28}$$

Using (9.26) and (9.28) the linear state equation is:

$$\mathbf{h_t} = \mathbf{A}\mathbf{h_{t-1}} + \mathbf{B}\mathbf{u_t} + \mathbf{w_t} \tag{9.29}$$

Where:

$$\mathbf{A}^{-1} = \begin{pmatrix} (2/\alpha)+1 & -1/\alpha & 0 & \cdots & \cdots & 0 \\ -1/\alpha & (2/\alpha)+1 & -1/\alpha & & & \vdots \\ 0 & -1/\alpha & (2/\alpha)+1 & & & \vdots \\ \vdots & 0 & & \ddots & -1/\alpha & 0 \\ \vdots & & & -1/\alpha & \ddots & -1/\alpha \\ 0 & \cdots & \cdots & 0 & -1/\alpha & (2/\alpha)+1 \end{pmatrix} \tag{9.30}$$

and

$$\mathbf{B}\mathbf{u}_t = \mathbf{A}^{-1} \begin{bmatrix} -(1/\alpha) & 0 \\ 0 & \mathsf{M} \\ \mathsf{M} & 0 \\ 0 & -(1/\alpha) \end{bmatrix} \begin{bmatrix} H_l \\ H_r \end{bmatrix} \tag{9.31}$$

At a point in time $t$ we observe the groundwater at $M$ grid nodes. This results in the measurement equation:

$$\mathbf{y_t} = \begin{bmatrix} 0 & 1 & 0 & \mathsf{L} & 0 \\ 0 & & 1 & & \mathsf{M} \\ & & & & \mathsf{M} \\ 0 & \mathsf{L} & & 1 & 0 \end{bmatrix} \begin{bmatrix} H_1 \\ \mathsf{M} \\ \mathsf{M} \\ H_N \end{bmatrix} + \begin{bmatrix} v_1 \\ \mathsf{M} \\ \mathsf{M} \\ v_M \end{bmatrix} \tag{9.32}$$

The elements of the measurement matrix (**H**) are all equal zero, except for the elements corresponding with the grid node where we have an observation. Those elements equal 1.

The prediction uncertainty for the groundwater system given in Figure 9.7 is quantified with the covariance matrix $\mathbf{P}_t$. In particular the diagonal elements hold the prediction variance of the grid nodes. At observation times we calculate $\mathbf{P}_t$ using the equations (9.8), (9.10) and (9.11). Analogous to (9.17) at time steps in between the observation time steps this reduces to $\mathbf{P}_t=\mathbf{M}_t$. To perform the calculations, we should know the matrices **A, H, Q** and **R**. In table 9.1 the values of the parameters which determine these matrices in this example are given.

Table 9.1 Parameter values of the example system of figure 9.7.

| Parameter | |
|---|---|
| $T$ | 600 $m^2$/day |
| $S$ | 0.1 |
| $\delta x$ | 150 m |
| $\delta t$ | 15 days |
| $q(i,j)$ | $q(i,j) = 3\left\{1-\exp(-d_{ij}/10)\right\}$ $cm^2$, with $q(i,j)$ is the element at row $i$ and column $j$ of the matrix **Q**; $d(i,j)$ is the distance between the spatial nodes $i$ and $j$. |
| $r(i,j)$ | 1 $cm^2$ $i=j$<br>0 $cm^2$ $i\neq j$<br>$r(i,j)$ is the element at row i and column j of the matrix **R** |

Note: If we only want to calculate the covariance matrix $\mathbf{P}_t$, we don't need to calculate the state and consequently we don't need the matrix **B** and the input $\mathbf{u}_t$. This can be seen from the scheme given in Figure 9.1.

In the example we use 50 grid nodes ($N$=50) and three monitoring locations ($M$=3). The monitoring locations are at the grid nodes 15, 30 and 45 and we have observations every 5 time steps. Figure 9.8 shows the prediction error variance $p(i,i)$ (i.e. the diagonal elements in matrix $\mathbf{P}_t$) for the different grid nodes at five successive time steps at and after an observation time $T$. The influence of the observations can clearly be seen. At the monitoring locations the variance is small (<1 $cm^2$). In between the monitoring locations the variance is larger. During the non-observed time steps T+1 till T+4, the variance grows. The influence of the observations is still visible, but if the time to the observations becomes larger, the influence of the observations become smaller. The time step T+5 is again an observation time step, and the variance is equal to the variance at time step T. The succession of the variance at time steps T till T+4 repeats. Also the effect of the boundary conditions, which are assumed to be exactly known at all time steps, is clear from Figure 9.8.

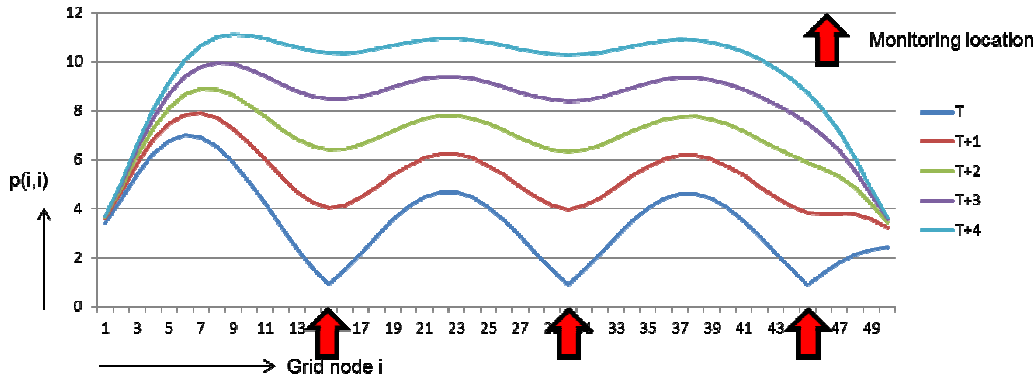



*Figure 9.8. Variance at 5 time steps as a function of the space coordinate x.*

Figure 9.9 shows the estimation variance at the grid nodes 15, 20 and 24 as a function of time. The grid node 15 is a monitoring location whereas grid nodes 20 and 24 are in between the observation locations.

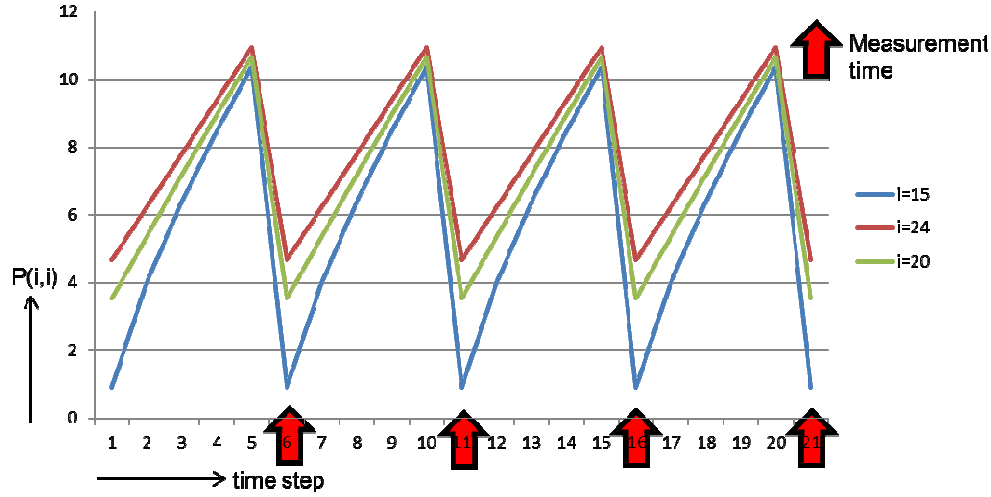



*Figure 9.9. Estimation variance at three different grind nodes as a function of time.*

### 9.5.2 Example two dimensional groundwater flow

Let's consider the hypothetical two-dimensional domain given in figure 9.10 taken from De Gruijter et al. (2006). The size of the domain is 30 x 30 km and a numerical groundwater model is built with a square calculation grid $\delta x = \delta y = 1\,\text{km}$ and the model time step $\delta t = 1\,\text{day}$. Similar to the one-dimensional example in 9.5.1, we can construct a state equation of the form (9.29) and a measurement equation of the form (9.32). We assume a monitoring network with 16 locations at regular distances, where the groundwater head is observed. The spacing between the observation screens is $\Delta x = \Delta y = 10\,\text{km}$. The groundwater head is observed at regular time intervals of 10 times the calculation time step ($\Delta t = 10\delta t$). The system is time invariant and the matrices **A, B, Q** and **R** are known.

At observation times those elements of measurement matrix **H** that correspond with the 16 locations of the observations equal 1, the other elements equal 0.

Using the Kalman Filter algorithm given in 9.2.2 we calculate the prediction error covariance matrix (measurement up-date) at all time steps ($\mathbf{P}_t$). As before, the element $p(i,i)$ at the diagonal of the $\mathbf{P}_t$ is the variance of the measurement up-date in grid point $(i,i)$. Also, at non-observed time steps we calculate the matrix $\mathbf{P}_t$ as $\mathbf{P}_t=\mathbf{M}_t$. The ratio between the variances in the matrix $\mathbf{P_t}$ and the variance of the system noise (denoted as $\left(\sigma_P^2/\sigma_Q^2\right)$ ) is plotted in Figure 9.11. The figure at the left hand side is the spatial pattern of this ratio at a observation time, and the figure at the right hand side is the pattern of the ratio 9 time steps ( $\delta t$ ) after observation time.

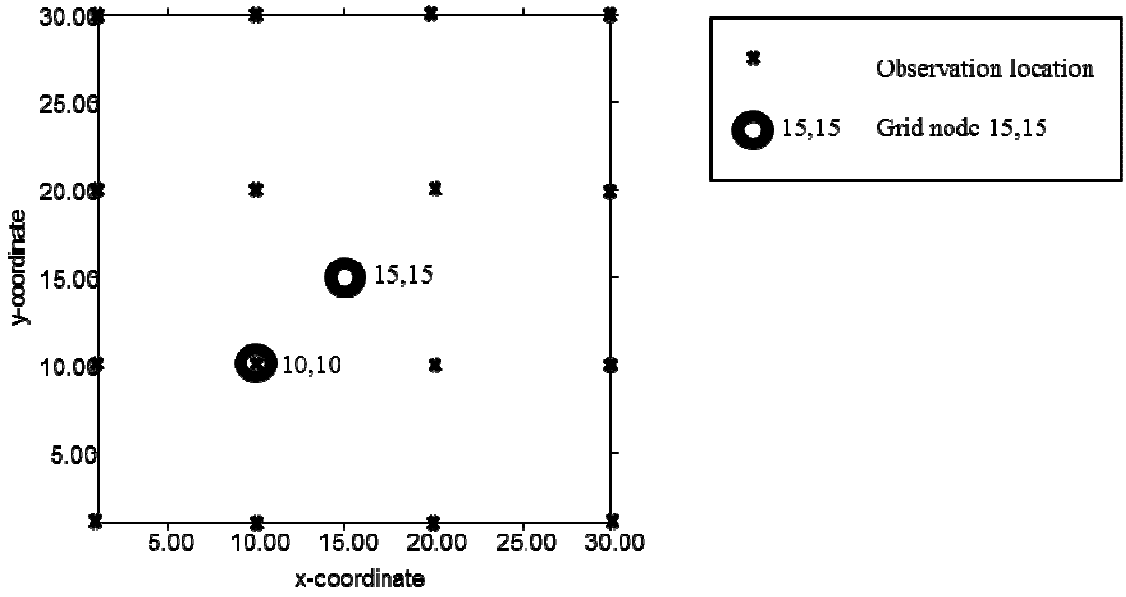



*Figure 9.10. Hypothetical two-dimensional groundwater flow domain.*

From the left hand side Figure 9.11 we can easily identify the locations of the 16 observation locations at the dark spots. At the observation locations the ratio is very small (<0.1). Thus, also the uncertainty is very small at those locations. If we move away from an observation location the ratio grows. At points far from the observation locations, such as point (15,15) the ratio exceeds 0.5. Observations provide more information in the vicinity of the observation locations than at points at some distance from the observation locations. The pattern in the left hand side figure 9.11 is comparable to the Kriging variance described in chapter 7. In the right hand side figure 9.11, the observation locations are not identifiable. Obviously, the observations carry only information about the groundwater for a limited period of time. The ratio is much larger than at observation times. In the middle (location (15,15)) the ratio is over 2.2. The low values at the boundaries are due to the fact that in this hypothetical example the boundaries conditions are known, fixed heads.

As shown in figure 9.11 the spatial pattern of the ratio at two points in time are quite different. It is clear that the estimation variance is a function of space as well of time. In figure 9.12 the temporal behavior of the ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$ is given for two locations (see figure 9.10). At the observation times, groundwater observations are available at location (10,10). The location (15,15) is the location that is the most far from the observation locations.

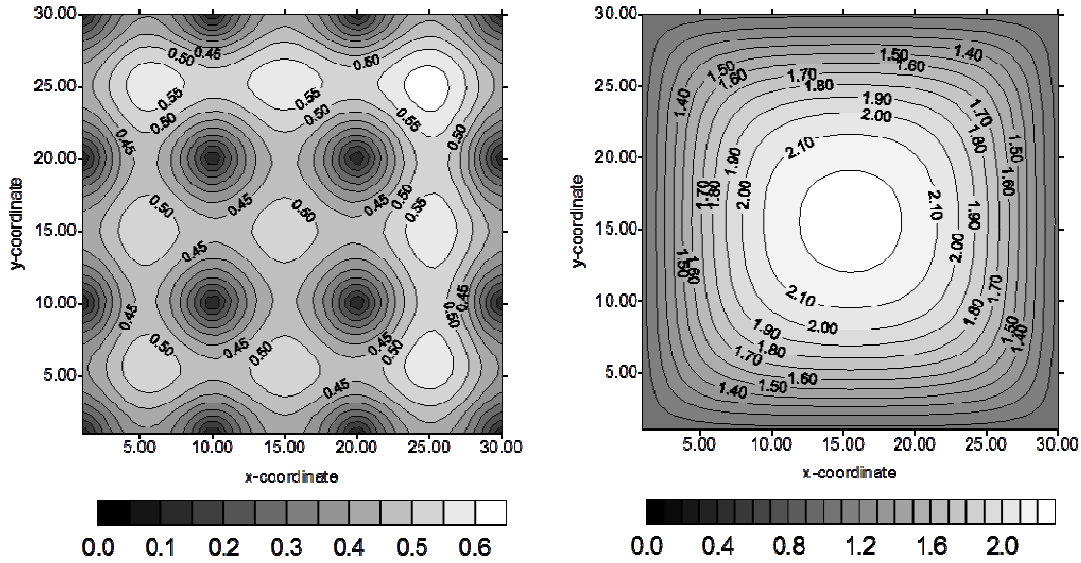



*Figure 9.11. Ratio of the estimation variance the variance of the system noise* $\left(\sigma_P^2 / \sigma_Q^2\right)$ *at an observation time (left) and 9 time steps after a observation time (right) (From de Gruijter et al., 2006).*

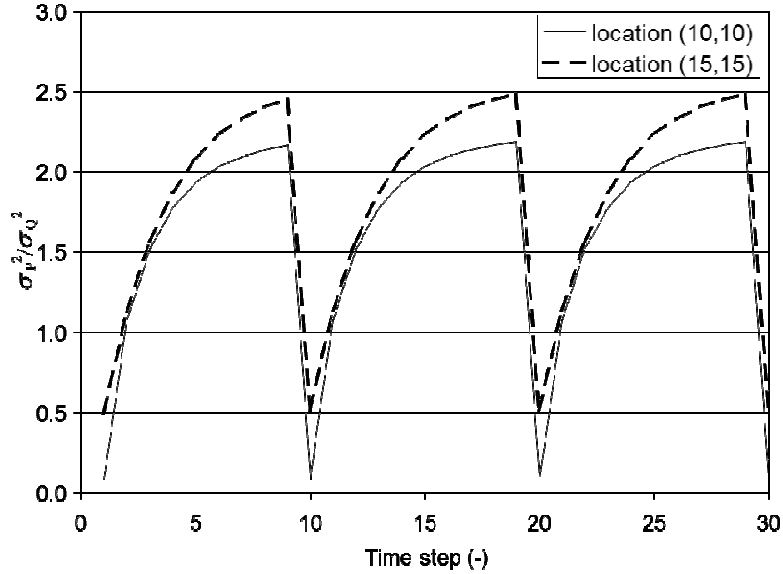



*Figure 9.12. Ratio* $\left(\sigma_P^2 / \sigma_Q^2\right)$ *as function of time for two locations (10,10) and (15,15) ) (From de Gruijter et al., 2006).*

From figure 9.12 it can be seen that the ratio grows with time after a observation time. At the next observation time, the ratio drops. The ratio at the monitoring location is always smaller than at the non-observed location.

*Evaluation of a monitoring strategy.*
When looking closer at the Kalman Filter algorithm (Equations 9.9 and 9.10), it becomes clear that for a time invariant system, the only reason for the covariance matrices $\mathbf{M}_t$ and $\mathbf{P}_t$ to be functions of time is the fact that the measurement matrix $\mathbf{H}_t$ is a function of time. The measurement matrix $\mathbf{H}_t$ holds the monitoring strategy (when and where do we have observations). So the covariance matrix of measurement up-date $\mathbf{P}_t$ is a function of the monitoring strategy only. Similar to the Kriging variance, we can calculate the covariance matrix $\mathbf{P}_t$ for a given monitoring strategy (locations and times) without having actual observations. We can use the relationship between $\mathbf{P}_t$ and $\mathbf{H}_t$ to evaluate the effectiveness of monitoring strategies. Here the domain average ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$ is adopted as evaluation criterion. The spacing between the observation locations is varied between two to ten times the grid size of the model grid ($0.1 \le (\delta x / \Delta x) \le 0.5$). The observation interval is varied between two to ten times the time step of the model ($0.1 \le (\delta t / \Delta t) \le 0.5$).

In figure 9.13 the domain average ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$ is plotted against the ratio's $(\delta x / \Delta x)$ and $(\delta t / \Delta t)$. In this figure lines of equal value of the ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$ are given. Figure 9.13 shows that an increase of the network density (smaller spacing of the observation locations) or an increase of the monitoring strategy frequency (smaller observation intervals) result in smaller values for the ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$, and therefore in smaller prediction uncertainty. Figure 9.13 also shows there is a trade off between network density and observation frequency. The equipotential lines give multiple combinations of network density and observation frequency that result in the same value of the ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$. This type of analysis can be used in optimizing the monitoring strategy. If we specify a target level of the ratio $\left(\sigma_P^2 / \sigma_Q^2\right)$ we can choose the most suitable (=cheapest) combination of network density and observation frequency to reach that target.

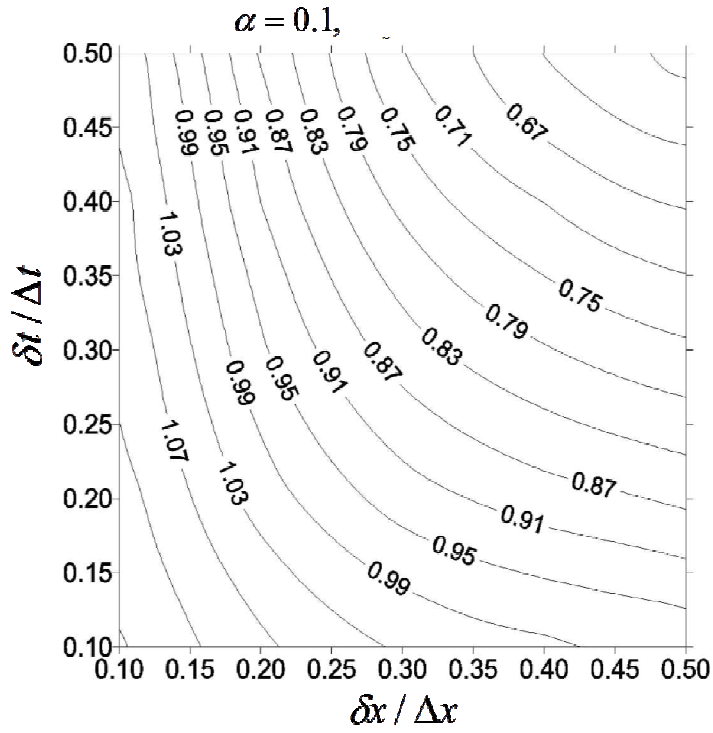



*Figure 9.13 Domain average ratio* $\left(\sigma_P^2 / \sigma_Q^2\right)$ *as function of the ratio's* $\left(\delta x / \Delta x\right)$ *and* $\left(\delta t / \Delta t\right)$ *( from de Gruijter et al. (2006)).*

## References


Alabert, F.G.,1987. The practice of fast conditional simulations through the LU-decomposition of the covariance matrix. *Mathematical Geology* 19(5), 369-386.

Armstrong, M. and P. Dowd (Eds), 1994. *Geostatistical Simulations*. Proceedings of the Geostatistical Simulation Workshop, Fontainebleau, France, 27-28 May 1993, Series on Quantitative Geology and Geostatistics, volume 7. Kluwer Academic Publishers, Dordrecht, Holland.

Bear, J., 1972. *Dynamics of Fluids in Porous Media*. Elsevier, New York.

Bierkens, M.F.P., P.A. Finke and P. de Willigen, 2000. *Upscaling and Downscaling Methods for Environmental ResearchSeries*. Developments in Plant and Soil Sciences 88, Springer, Berlin.

Bierkens, M.F.P., 2006. Designing a monitoring network for detecting groundwater pollution with stochastic simulation and a cost model. *Stochastic Environmental Research and Risk Assessment* 20, 335-351.

Box, G.E P. and D.R.Cox, 1964. An analysis of transformations. *Journal of the Royal Statistical Society*, Series B 26 (2): 211–252.

Box, G.E.P. and G.M. Jenkins, 1976. *Time series analysis, forecasting and control. Revised Edition*. Holden-Day, San Francisco.

Chatfield, C., 1989. *The analysis of time series. An introduction. Fourth edition*. Chapman and Hall, London.

Chilès, J.-P. and P. Delfiner, 1999. *Geostatistics. Modelling Spatial Uncertainty*. John Wiley & Sons, New York.

Christakos, G., 1992, *Random Field Models in the Earth Sciences*, Academic Press, San Diego, CA, USA

Christakos, G., 2000. *Modern Spatiotemporal Geostatistics*. Oxford University Press, Oxford.

Cressie, N.A.C., 1993. *Statistics for Spatial Data*. Revised Edition. John Wiley & Sons, New York.

David, M., 1977. *Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam, 364 p.

Dagan, G., 1989. *Flow and Transport in Porous Formations*. Springer Verlag, Berlin.

De Gruijter, J.J., D.J. Brus, M.F.P. Bierkens and M. Knotters, 2006. *Sampling for Natural Resources Monitoring*. Spinger, Berlin.

De Marsily, G., 1986. *Quantitative Hydrogeology*; *Groundwater Hydrology for Engineers*. Academic Press, Inc, Orlando, 440 p.

Deutsch, C.V. and A.G Journel, 1998. *GSLIB, Geostatistical Software Library and User's Guide*. 2nd Edition. Oxford University Press, New York, 369 pp.

Gardiner, C.W., 1983. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Second edition. Springer Verlag, Berlin.

Gelhar, L.W., 1993. *Stochastic Subsurface Hydrology*. Prentice Hall, Englewood Cliffs, New Jersey.

Gómez-Hernández, J.J. and A.G. Journel, 1993. Joint sequential simulation of MultiGaussian fields. In: A. Soares (Ed.), *Geostatistics Tróia* '92, Volume 1, pp. 85-94. Kluwer Academic Publishers, Dordrecht.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 483 p.

Grimmet, G.R. and D.R. Stirzaker, 1998. *Probability and random processes*. Oxford University Press, Oxford.

Heuvelink, G.B.M., 1998. *Error Propagation in Environmental Modelling with GIS*. Taylor & Francis, London.

Hipel, K.W. and A.I. McLeod, 1994. *Time series modelling of water resources and environmental systems*. Elsevier, New York.

Hipel, K.W., W.C. Lennox, T.E. Unny and A.I. McLeod, 1975. *Intervention analysis in water resources.* Water Resources Research 11: 855-861.

Hosking, 1985. *Journal of Hydrology* 78, 393-396.

Isaaks, E.H. and R.M. Srivastava, 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 561 p.

Journel, A.G. and C.J. Huijbregts, 1978. *Mining Geostatistics*. Academic Press, New York, 600 p.

Knotters, M. and M.F.P. Bierkens, 1999. Physical basis of time series models for water table depths. *Water Resources Research* 36(1), 181-188.

Krige, D.G., 1951. *A statistical Approach to Some Mine Evaluations an Allied Problems at the Witwatersrand*. Master's thesis, University of Witwatersrand.

Koutsoyiannis, D., 2010. A random walk on water. *Hydrology and Earth System Sciences* 14, 585–601, 2010.

Landwehr e.a. *Water Resources Research* 1997(10) 1055-1063.

Mantoglou, A. and J.L. Wilson, 1982. The turning bands method for simulation of random fields using line generation by the spectral method, *Water Resources Research* 18(5): 1379-1394.

Matheron, G., 1970. *La Théorie des variables Régionaliséés et ses Applications*. Fascicule 5, Les cahiers du Centre de Morphologie Mathematique, Ecoles des Mines, Paris, Fontainebleau, 212 p.

Papoulis, A., 1991. *Probability, Random Variables and Stochastic processes*. Third Edition. McGraw-Hill.

Payne, R.W. (Edt.), 2000. *The Guide to GenStat*. Lawes Agricultural Trust, Harpenden.

Priestley, M.B., 1981. *Spectral analysis and time series*. Academic Press, London.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univ. Press, New York.

Rivoirard, J., 1994. *Disjunctive Kriging and Non-linear Geostatistics*. Clarendon Press, Oxford.

Stedinger, J.R., R.M. Vogel and E. Foufoula-Georgiou, 1993. Frequency analysis of extreme events. In: D.R. Maidment (editor), *Handbook of Hydrology*, chapter 18, McGraw-Hill, New York.

Snedecor, George W. and Cochran, William G. (1989), *Statistical Methods*, Eighth Edition, Iowa State University Press.

Te Stroet, C.B.M., 1995. *Calibration of stochastic groundwater flow models; estimation of system noise statistics and model parameters*. Ph.D. Thesis Delft University of Technology.

Tong, H., 1990. *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, New York.

VanMarcke, E., 1983. *Random Fields*. MIT Press, Cambridge MA.

Van Montfort, M.A.J., 1969. *Statistica Neerlandica* 23, 107.

## Appendix: Exam Stochastic Hydrology 2008

1. (5 points) The multivariate probability density function of a random function is Gaussian. The mean of the random function is constant, the variance is finite and constant and the covariance between the values at two locations only depends on the distance between these locations. Is this random function:
   a. Intrinsic?
   b. Wide sense stationary?
   c. Second order stationary?
   d. Strict stationary?
   e. Isotropic?

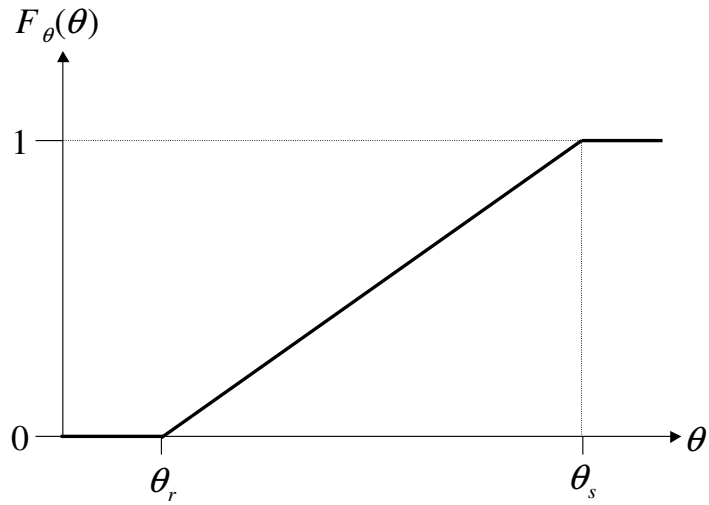
   Briefly explain your answers?

2. (2 points) Provide two advantages of using ordinary kriging over simple kriging?

3. (6 points) The cumulative probability distribution of root zone soil moisture content $\theta$ at a certain location is given by (see Figure):

$$F_\theta(\theta) = \begin{cases} 0 & \text{if } \theta \le \theta_r \\ \dfrac{\theta - \theta_r}{\theta_s - \theta_r} & \text{if } \theta_r < \theta \le \theta_s \\ 1 & \text{if } \theta > \theta_s \end{cases} \qquad (1)$$

with $\theta_s$ soil moisture at saturation and $\theta_r$ residual soil moisture content.



a) Give the expression for the probability density function of root zone soil moisture content.
b) Derive expressions for the mean and the variance of root zone soil moisture content.
c) If $\theta_s = 0.45$ and $\theta_r = 0.04$ and the root zone depth is 30 cm. On a given day we have a precipitation event of 30 mm. If we assume that all precipitation infiltrates and that percolation during the rain event can be neglected: What is the probability that this precipitation event generates surface runoff (i.e. that it causes the root zone to become saturated)?

4. (6 points) Consider the following isotropic covariance function of a wide-sense stationary random function with mean $\mu_Z = 10$:

$$C_Z(h) = 20\exp(-h/10) \tag{2}$$

Observations have been made at locations $z(x,y) = z(1,1) = 6$ and $z(x,y) = z(4,5) = 13$. Use simple Kriging to predict the value of $Z(\mathbf{x})$ at location $(x,y) = (2,4)$ and estimate the prediction error variance.

5. (8 points) Discharge in an open channel can be predicted using Manning's equation:

$$Q = \frac{A}{m} R^{2/3} S^{1/2} \tag{3}$$

with $Q$ the discharge ($m^3/s$), $A$ the wetted perimeter ($m^2$), $R$ hydraulic radius (m), $m$ Manning's coefficient and $S$ the slope of the water surface. In a hydraulic laboratory this equation is used to determine Manning's coefficients of different types of channel bottom material by running water over this material in a experimental flume with a known discharge and measuring the water height at two locations along the channel at a distance $L$ from each other. From the measurements $H_1$ and $H_2$ first the slope calculated as:

$$S = \frac{H_2 - H_1}{L} \tag{4}$$

after which Manning's coefficient is calculating by Manning'equation as:

$$m = \frac{A}{Q} R^{2/3} S^{1/2} \tag{5}$$

a. The water levels $H_1$ and $H_2$ are measured with a random measurement error with mean zero and the same variance $\sigma_H^2$. Furthermore, the measurement errors at the two locations are independent. Give an expression for the variance of the error of the slope $\sigma_S^2$ as a function of $\sigma_H^2$.

b. Use the first order Taylor approximation to derive an expression for the variance of estimated Manning coefficient $\sigma_m^2$ as a function of the variance $\sigma_S^2$; next use the results of 5a to derive the expression of $\sigma_m^2$ as a function of $\sigma_H^2$.

c. Suppose we have $A$ = 2 m$^2$, $R$ = 1 m, L = 10 m and $Q$ = 0.20 m$^3$/s and we estimate from observations that the expected value of the slope $\mu_S$ is 0.0005 m/m. The standard deviation $\sigma_H$ of the observation error that occurs when measuring the water level is $\sigma_H = 0.001$ m. Give a first order estimate of the Manning-coefficient and a first order estimate of the estimation variance. What is the relative error $\sigma_m / \mu_m$?

d. If the flume has a length of 20 m. What would be an easy way to improve the accuracy of the estimate of Manning's coefficient?

6. (7 points) Consider the following stochastic model describing monthly concentration of nitrogen $c_k$ (in mg/l) in a lake with time steps of one month (index $k$ is the month number):

$$C_k = 0.6C_{k-1} + 12q_k + W_k \tag{6}$$

with $q_k$ the monthly nitrogen input into the lake ($10^3$ kg /month) and $W_k$ a zero-mean model error. Nitrogen $q_k$ input is given in the following Table:

| Time (month numbers) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| N input ($10^3$ kg /month) | 4 | 12 | 9 | 8 | 2 | 3 | 4 | 2 | 1 | 0 |

The initial concentration is 10 mg/l with an initial error with variance $\sigma_{0|0}^2 = 9\ mg^2/l^2$. The variance of the model error is $\sigma_W^2 = 36\ mg^2/l^2$. At $k$=3 and $k$=9 observations $y_k$ are taken. We have $y_3$ = 173 mg/l and $y_9$= 83 mg/l. The variance of the observation error is the same for both times: $\sigma_V^2 = 9\ mg^2/l^2$.

a) Apply the Kalman filter to obtain the optimal prediction $\hat{c}_k$ and prediction variance $\sigma_{k|k}^2 = E[(\hat{c}_k - c_k)^2]$ for all time steps $k$=1,..,10.

b) Make a plot of $\hat{c}_k$ and $\sigma_{k|k}$ versus time $k$. In the plot with $\hat{c}_k$ also plot the results of applying the deterministic model ($c_k = 0.6c_{k-1} + 12q_k$) without using the Kalman filter. How many months can we see the influence of a model update in our model predictions?