Introduction to Geostatistics

András Bárdossy

Institute of Hydraulic Engineering
University of Stuttgart

Contents

Chapter 1

Introduction

1.1 Variability

Most of the natural phenomena we study are variable both in space and time. Considering a topographic surface or a groundwater contamination one can observe high variability within small distances. The variability is a result of natural processes, thus deterministic. As most of these processes are sensitive and the conditions under which the they took place are not fully known, it is not possible to describe them based on physical and chemical laws completely.

The practical consequences of variability are:

- field and laboratory measurements are necessary
- upscaling is non trivial
- there is always a limited degree of explanation one can achieve

1.2 Stochastic methods

It is not uncommon to use probabilistic and statistical methods for describing partly known (or sampled) natural parameters. Philosophically the appropriate-

ness of this approach can be argued. As the realization under study already exists at the moment of the investigation, there is no randomness present in the sense of the traditional approach.

On the other side, this kind of existence does not contradict the application of probabilistic and statistical methods. For example, one faces the same kind of uncertainty before tossing a coin and after it if the coin is covered immediately, before one can see it. The more disturbing and restricting problem is the uniqueness of the realization. A coin can be tossed several times, but a natural phenomenon cannot be repeated. This means that no frequencies are available, and thus probabilities cannot be assessed this way.

Measurement values of a certain parameter, obtained from different locations are often treated as different outcomes of the same random variable. Means and variances, cumulative distributions etc. are calculated this way. However, the hypothesis of independent trials is not always applicable. Consider the following simple example.

EXAMPLE 1.1 :

No.of samples	Mean	Variance
12	279.5	4554
21	283.9	7072
46	281.6	8646

Table 1.1: Means and variances calculated from different samples

Elevation of 46 meteorological stations was considered. As a first step mean and variance corresponding to the central 12 stations were calculated. Then a bit bigger neighbourhood was considered; finally all stations were included. Table 1.1 shows the means and variances. Note that the mean does not change, while the variance increases with the sample size (proportional to the area covered by these stations). If the parameter is modeled as the realization of a random variable, then the usual assumption of independence would contradict the increase of the variance.

1.3 Geostatistics

Time series analysis is one of the first fields where variability has been considered and described with stochastic methods. These methods were extended and further developed to analyse spatial variability. These spatial methods form the discipline called geostatistics.

The word *geostatistics* is formed from the two parts geo and statistics similarly to geophysics or geochemistry. It is used with two different meanings:

1. as a collection of all statistical and probabilistic methods applied in geo sciences,

2. as an other name for the theory of regionalized variables.

The theory of regionalized variables, which is the topic of the next pages dates back to the early fifties when in South-Africa D. Krige and his colleagues started to apply statistical techniques to ore reserve estimation. In the sixties the french matematician G. Matheron gave theoretical foundations to the above methods. Geostatistics was first used by the mining industry, as high costs of drillings made the analysis of the data extremely important. Books and publications on geostatistics are mostly oriented to mining problems. As the computers got cheaper and cheaper the computationally expensive methods could also be used in other topics.

Applications of geostatistics can be found in very different disciplines ranging from the classical fields mining and geology to soil science, hydrology, meteorology, environmental sciences, agriculture, even structural engineering.

The following text does not contain a complete theory of geostatistical methods illustrated with applications. It is planned to be a practical introduction. Theoretical results and long derivations are not included, results are mostly presented without detailed proofs. However, the ideas behind these results are always presented.

1.4 Notation

Throughout the following text u will always denote a point in the d dimensional space $u = x$ if $d = 1$, $u = (x,y)$ if $d = 2$ and $u = (x,y,w)$ if $d = 3$; if time also plays a role it will be denoted by t. Measurement points are indexed, $u_i = (x_i, \ldots)$. Spaces of different dimensions $d = 1, \ldots, 3$ will be considered, and the formulation is given in a general form.

For simplicity the notation $\int f(u)\,du$ will be used for single, double, and triple

integrals, too:

$$\int f(u)du = \begin{cases} \int f(x)dx & \text{if } d = 1 \\ \int\int f(x,y)dx\,dy & \text{if } d = 2 \\ \int\int\int f(x,y,w)dx\,dy\,dw & \text{if } d = 3 \end{cases} \tag{1.1}$$

Some double integrals can also appear in the text, they are meant to be double integrals in the above (1.1) sense.

Chapter 2

Statistical hypotheses

2.1 Basic concepts

Loosely speaking a random variable is a function Z which might take different values (outcomes) with given probabilities. If the outcomes form a finite (or coutably infinite) set then one speaks of a discrete random variable.

Random variables are characterized by their distribution function:

$$F_Z(z) = P[Z \leq z] \tag{2.1}$$

Distribution functions are non-decreasing with values in $[0,1]$. The probability of Z being in the interval $[a,b]$ can be calculated using the distribution function:

$$P[a < Z \leq b] = F(b) - F(a) \tag{2.2}$$

For many non-discrete random variables the distribution function F is connected to the density function f through:

$$F_Z(z) = \int_{-\infty}^{z} f_Z(t)\, dt \tag{2.3}$$

The expected value of a random variable is its 'mean value over infinetely

many realizations'. It is:

$$E[Z] = \int_{-\infty}^{+\infty} t \, dF_Z(t) \tag{2.4}$$

For random variables with density function this can be written as:

$$E[Z] = \int_{-\infty}^{+\infty} t \, f_Z(t) \, dt \tag{2.5}$$

Moments of a random variable are defined as:

$$E[Z^m] = \int_{-\infty}^{+\infty} t^m \, dF_Z(t) \tag{2.6}$$

The central moments are:

$$E[(Z - E[Z])^m] = \int_{-\infty}^{+\infty} (t - E[Z])^m \, dF_Z(t) \tag{2.7}$$

The second central moment is called variance:

$$\mathrm{Var}[Z] = \sigma^2 = E[(Z^2 - E[Z])^2] \tag{2.8}$$

The expected value has a linear behavior:

$$E[Z_1 + Z_2] = E[Z_1] + E[Z_2] \tag{2.9}$$

and

$$E[aZ] = aE[Z] \tag{2.10}$$

This is not true for the higher moments and in general for non-linear functions g:

$$E[g(Z)] \neq g(E[Z]) \tag{2.11}$$

The joint behavior of more random variables $Z_1, \ldots, Z_n$ can be described by their joint distribution function:

$$F_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = P[Z_1 \leq z_1 \text{ and } \ldots \text{ and } Z_n \leq z_n] \tag{2.12}$$

2.2 Regionalized variables

In the theory of regionalized variables the concept of random functions plays a central role. A *random function* is a set of random variables corresponding to the points of the domain D under study. This means that for each point u in D there is a corresponding random variable $Z(u)$.

A *regionalized variable* is the realization of a random function. This means that for each point u in the d dimensional space the value of the parameter we are interested in, $z(u)$ is one realization of the random function $Z(u)$. This interpretation of the natural parameters acknowledges the fact that it is not possible to describe them completely using deterministic methods only. In most cases it is impossible to check the assumption that the parameter is the realization of a random function as we have to deal with a single realization.

One could describe a random function by its multidimensional distribution functions. This means that for each set of points $u_1,\ldots,u_n$ in the domain D, a cumulative distribution function $F_{u_1,\ldots,u_n}$ is assigned. Using these functions for each set of possible values $w_1,\ldots,w_n$ one could find the probability P :

$$P(Z(u_1)<w_1,\ldots,Z(u_n)<w_n)=F_{u_1,\ldots,u_n}(w_1,\ldots,w_n) \tag{2.13}$$

This would mean that conditional probabilities could be used for the estimation of local or global averages etc. Unfortunately there are infinitely many finite subsets in the domain D, and as for each point in D usually only one value (the realization) is available the assessment of the distribution functions based on the experimental data seems to be illusory. Even in the case of repeatedly measured parameters (for example groundwater quality) there are not enough measurements to assess the above distribution functions.

A general hypothesis which reduces the complexity of the problem is the so called *strong stationarity* . Formally it is:
The random function $Z(u)$ is stationary if for each set of points $u_1,\ldots,u_n$ in the domain D, and for each set of possible values $w_1,\ldots,w_n$, and for each vector h:

$$P(Z(u_1)<w_1,\ldots,Z(u_n)<w_n)=P(Z(u_1+h)<w_1,\ldots,Z(u_n+h)<w_n) \tag{2.14}$$

This equation means that the distribution of the random function depends on the configuration of the points and not on their locations. In other words this can be formulated that "nature" repeats itself similarly for the same configuration.

The assumption of strong stationarity is useful, but still a bit too complex to be appropriate. To deal with the problem effectively some even simpler assumptions have to be made. The two basic and very similar assumptions are the following:

2.3 Second order stationarity

Stationarity is a concept often used in time series analysis. Here the second order stationarity hypothesis is formulated for multidimensional spaces.

The assumption of second order stationarity consists of two conditions:

- The expected value of the random function $Z(u)$ is constant all over the domain D.
- The covariance of two random variables corresponding to two locations depends only on the vector h separating these two points.

These conditions can be formulated as:

$$E[Z(u)] = m \tag{2.15}$$

for all $u \in D$

$$E[(Z(u+h)-m)(Z(u)-m)] = C(h) \tag{2.16}$$

for any $u, u+h \in D$, where $C(h)$ depends only on the vector h and not on the locations u and $u+h$. The function $C(h)$ is called *covariance function*. In this case one has for $h = 0$:

$$C(0) = E[(Z(u)-m)(Z(u)-m)] = Var[Z(u)] \tag{2.17}$$

Equation (2.17) shows that the random variables corresponding to different points in the domain do not only have the same expectation, but they also have to have the same finite variance. This second condition is not always met, but weaker assumptions can be formulated.

2.4 Intrinsic hypothesis

The assumption slighly weaker than the second order stationarity is the so called intrinsic hypothesis. The first condition is the same as in the case of second order stationarity, only the second is different:

- The expected value of the random function $Z(u)$ is constant all over the domain D.
- The variance of the increment corresponding to two different locations depends only on the vector separating them.

These conditions can be formulated as:

$$E[Z(u)] = m \tag{2.18}$$

for all $u \in D$

$$\frac{1}{2}Var[Z(u+h) - Z(u)] = \frac{1}{2}E[(Z(u+h) - Z(u))^2] = \gamma(h) \tag{2.19}$$

where $\gamma(h)$ depends only on the vector h and not on the locations u and $u+h$. The function $\gamma(h)$ is called *semivariogram* . The semivariogram is often called simply *variogram* , for convenience this sloppy convention will be used throughout this text. One can see that equation (2.19) is very similar to (2.16), but the implicit assumption of the finite variance is not included. It can be demonstrated that the second order stationarity implies the intrinsic hypothesis, but the converse is not true. In the case of second order stationarity one has:

$$E[(Z(u+h) - Z(u))^2] = E[((Z(u+h) - m) - (Z(u) - m))^2] =$$
$$= Var[Z(u)] + Var[Z(u+h)] - 2E[(Z(u+h) - m)(Z(u) - m)] = 2C(0) - 2C(h) \tag{2.20}$$

So the relation:

$$\gamma(h) = C(0) - C(h) \tag{2.21}$$

Figure 2.1 shows this relationship between the covariance function and the variogram.

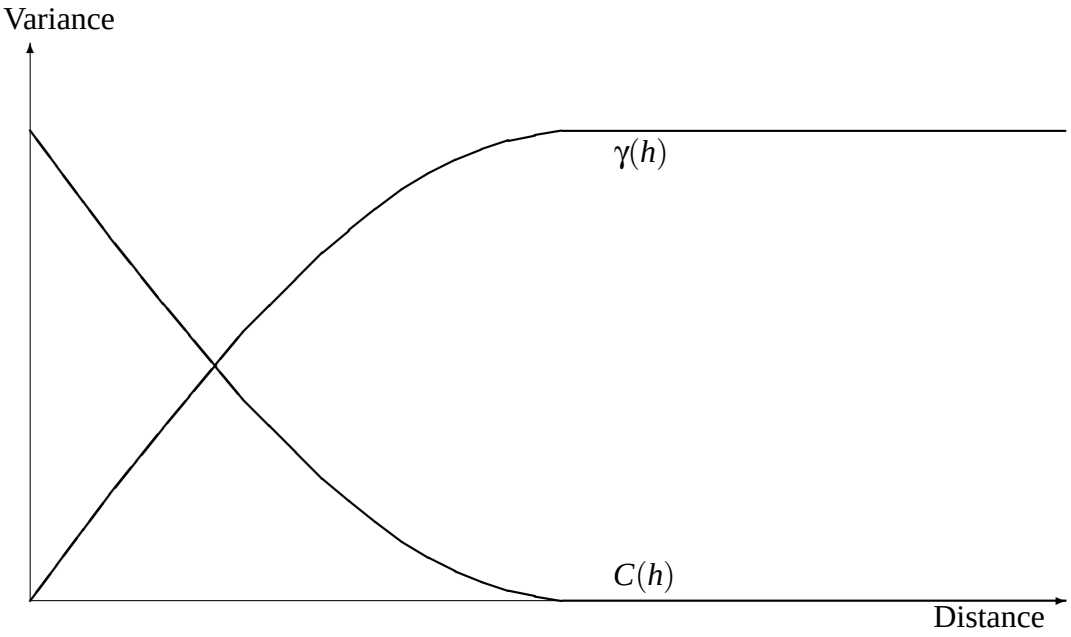


Figure 2.1: The covariance function $C(h)$ and the variogram $\gamma(h)$

The intrinsic hypothesis was first considered by pioneers of geostatistics in South Africa. The assumption of finite variances in gold deposits did not seem to be suitable, this led to the introduction of this hypothesis.

2.5 Comparison of the two hypotheses

The difference between the intrinsic hypothesis and the second order stationarity is not only the fact that the first is more general than the second. The covariance function (2.16) is defined using the value of the expectation m, while the variogram (2.19) does not depend on this value. This is an advantage because slight trends do not influence the variogram severely, in contrast to the covariance function where through the improper estimation of the mean these effects are more severe.

2.6 Selection of the regionalized variable

The regionalized variable under study has to fulfill certain conditions to apply geostatistical methods. These conditions are:

1. Data homogeneity: The data should reflect one parameter, measured by the same measurement method, and the measurements should be made on the same volume (support).

2. Additivity : The parameter should have the property that $\frac{1}{n}\sum_{i=1}^{n} Z(u_i)$ has the same meaning as $Z(u)$.

To understand the meaning of the additivity condition consider the following example:

EXAMPLE 2.1 :

Suppose that $Z(u)$ represents the thickness of a layer measured in m. If the average thickness over a certain area is needed, then the arithmetic mean of a regular sampling is a good estimator for this. If instead $Z'(u)$ is the cube of the thickness then the arithmetic mean of the individual $Z'(u_i)$ values is not the cube of the mean thickness. To see this explicitly suppose two samples are available: $Z(u_1) = 1$ and $Z(u_2) = 2$. So $Z'(u_1) = 1$ and $Z'(u_2) = 8$. Then for the mean one has

$$0.5Z(u_1) + 0.5Z(u_2) = 1.5$$

$$0.5Z'(u_1) + 0.5Z'(u_2) = 4.5$$

but using the definition of $Z'(u)$ one has:

$$(0.5Z(u_1) + 0.5Z(u_2))^3 = 3.375$$

This means that $Z'(u)$ is not additive.

Some natural parameters are clearly non additive, like hydraulic conductivity etc. In the case of non additive parameters it is possible to use transformations

which transform them to additive ones. Data homogeneity problems (like different measurement types) can sometimes be overcome, some cases are discussed later.

Chapter 3

The variogram

As the variogram is defined the variance of an increment certainly has to fulfil several conditions. The precise conditions of a variogram will be discussed in the section describing the theoretical variograms. Naturally there are also properties of the variogram which we know or suppose without any precise mathematical description.

- From the definition we have $\gamma(0) = 0$.
- From the definition $\gamma(h) \geq 0$ for all h vectors
- From the definition $\gamma(h) = \gamma(-h)$ for all h vectors
- In most cases we suppose there is some kind of continuity in the parameter we are dealing with. This means that the variance of the increments is supposed to increase with the length of the vector h.
- In several cases there is a certain limit in the continuity of the parameter. This means that taking if the vector separating two points exceeds a certain limit the variance of the increment will not increase any more.
- The variogram is often discontinuous near the origin. This means that for any $h \neq 0$ we have $\gamma(h) \geq C_0 > 0$. This phenomenon is the so called nugget

effect. The nugget effect can partly be explained by the measurement error and partly by a random component in the parameter which is not spatially dependent.

It is clear that the hypothesis about the existence of the variogram is the key point of geostatistics. The first question naturally arising is: "Can I assume that my parameter under study fulfils the intrinsic hypothesis ?"

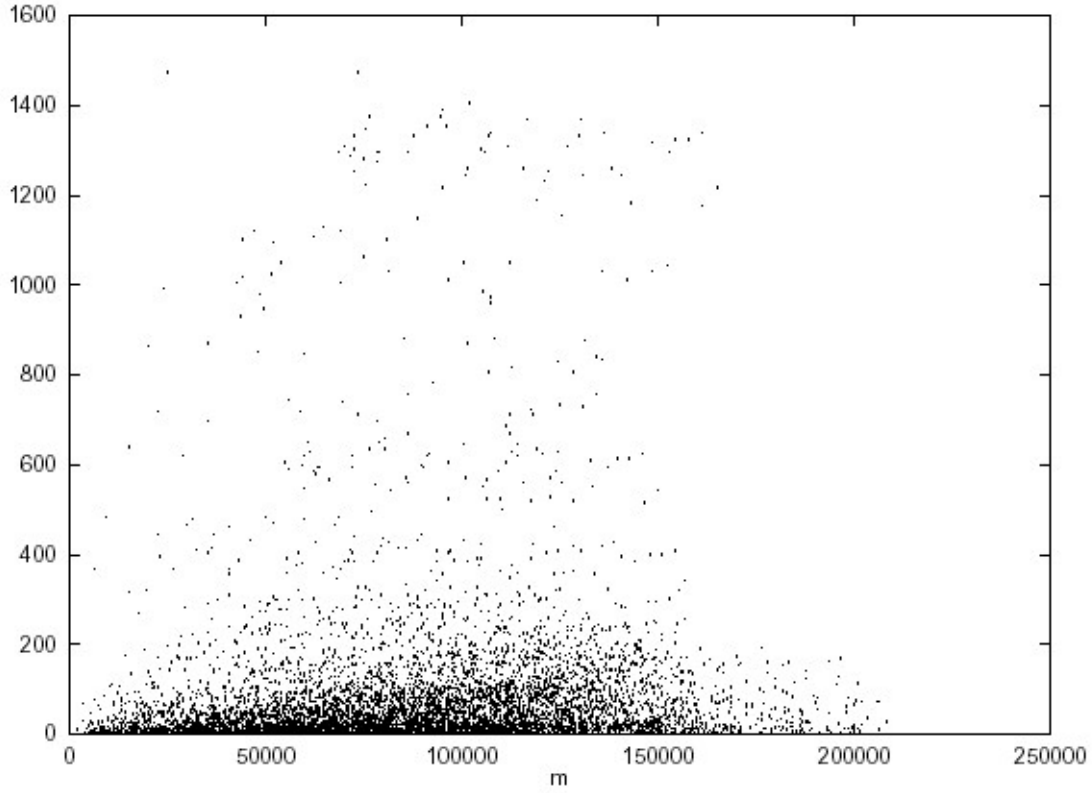


Figure 3.1: Variogram cloud $[mm^2]$ (precipitation Jan.3, 1990)

Suppose that measurements of the parameter are taken at locations u_i for $i = 1, \ldots, n$. Let $Z(u_i)$ be the measured values. As a first step the impatient reader would calculate the values $(Z(u_i) - Z(u_j))^2$ for all the pairs formed from the measurement points u_i, and would then plot them with respect to the distance (and

perhaps direction) separating the points. This way a so called *variogram cloud* is obtained. Figure 3.1 shows such a variogram cloud. It seems to be a rather discouraging result.

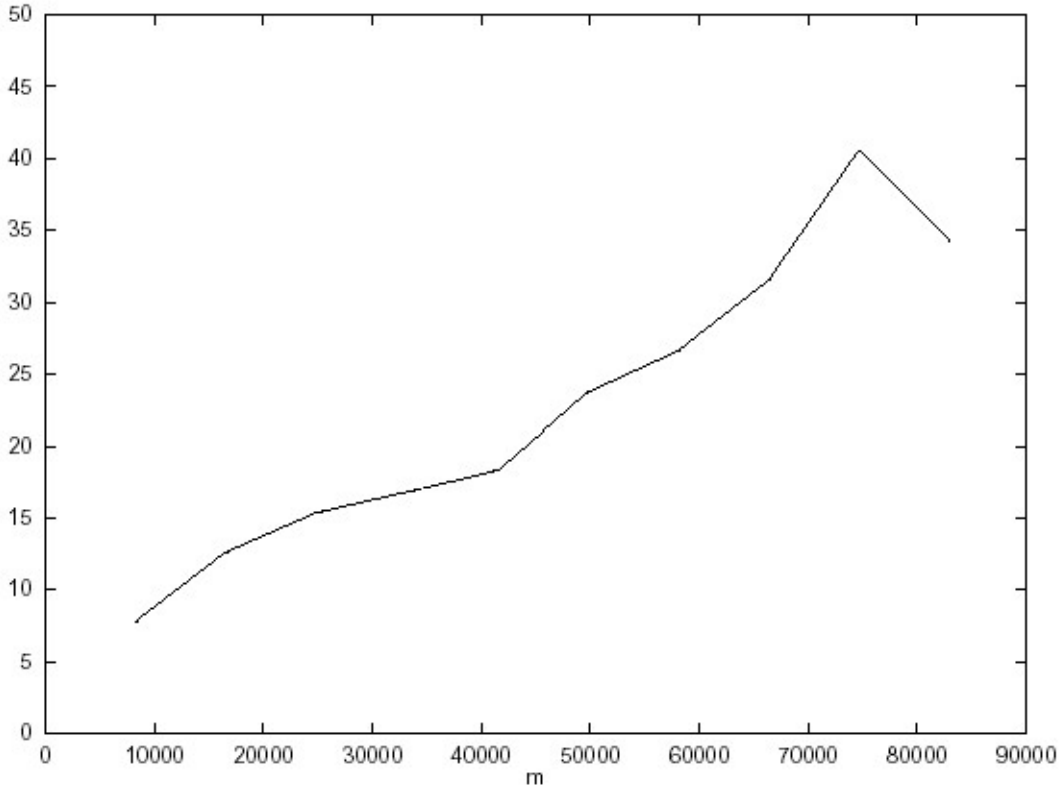


Figure 3.2: Experimental variogram $[mm^2]$ (precipitation Jan.3,1982)

However, the condition (2.19) did not promise that for all possible pairs the value of $(Z(u_i) - Z(u_j))^2$ will be close to a certain line. It is a statement on the expectation of these values. If we draw these expectations (calculated as arithmetic means) for the same case as for which the variogram cloud was obtained (figure 3.1) the result is already promising as shown on figure 3.2 .

3.1 The experimental variogram

The variogram function has to be estimated on the basis of the available data. In the case of a finite data set the estimation of the variogram can be made for a finite set of vectors only.

The variogram can be estimated with the help of the following formula:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (Z(u_i) - Z(u_j))^2 \tag{3.1}$$

Here $N(h)$ is the number of pairs of locations separated by the vector h.

The calculation of the above function, called *experimental variogram* is straightforward in the case of regularly spaced data points. Even in this case the experimental variogram is calculated for a finite number of vectors. If the points are irregularly spaced the condition for the summation $u_i - u_j = h$ has to be weakened, in order to have more pairs and not to obtain a variogram cloud as above. This can be done by allowing a certain difference in both the angle and the length of the vector. This means that the summation should be made over the pairs fulfiling:

$$\begin{aligned} |u_i - u_j| - |h| &\leq \varepsilon \\ \text{Angle}(u_i - u_j, h) &\leq \delta \end{aligned} \tag{3.2}$$

Here $|.|$ denotes the length of a vector.

EXAMPLE 3.1 :

u	1	2	3	4	5	6	7	8	9	10
$Z(u)$	41.2	40.2	39.7	39.2	40.1	38.3	39.1	40.0	41.1	40.3

Table 3.1: Data points and values for example 3.1

Suppose all measurement points are alligned along the same straight line. (For example data of the same borehole.) Also suppose that all the data points are

equally spaced - two neighbouring data points are separated by the distance of 1 m. Using the data given in Table 3.1 one has:

$$\begin{aligned}\gamma^*(1) &= \frac{1}{18}[(41.2-40.2)^2+(40.2-39.7)^2+(39.7-39.2)^2+(39.2-40.1)^2+ \\ &\quad +(40.1-38.3)^2+(38.3-39.1)^2+(39.1-40.0)^2+(40.0-41.1)^2+(41.1-40.3)^2]= \\ &= 0.4917\end{aligned}$$

and

$$\begin{aligned}\gamma^*(2) &= \frac{1}{16}[(41.2-39.7)^2+(40.2-39.2)^2+(39.7-40.1)^2+(39.2-38.3)^2+ \\ &\quad +(40.1-39.1)^2+(38.3-40.0)^2+(39.1-41.1)^2+(40.0-40.3)^2]= \\ &= 0.756\end{aligned}$$

EXAMPLE 3.2 :

In this example data of a regular grid are considered with values missing at certain locations. The configuration of the data and the values are showed on figure 10.3.

The experimental variogram value corresponding to the direction of the *x* axis, with the length of 25 m can be calculated as:

$$\begin{aligned}\gamma^*(25_x) &= \frac{1}{18}[(12-11)^2+(13-12)^2+(11-10)^2+(10-11)^2+(11-11)^2 \\ &\quad +(11-12)^2+(12-10)^2+(10-14)^2+(14-13)^2]=1.4444\end{aligned}$$

From the same data in the *y* direction one obtains :

$$\begin{aligned}\gamma^*(25_y) &= \frac{1}{18}[(10-11)^2+(12-11)^2+(11-11)^2+(10-10)^2+(10-10)^2 \\ &\quad +(11-13)^2+(13-13)^2+(13-11)^2+(11-12)^2]=0.6111\end{aligned}$$

This example does not only show how the values of an experimental variogram are calculated, but also shows that the contribution of pairs with big differences is very important. Excluding the data point with the value 14 one has

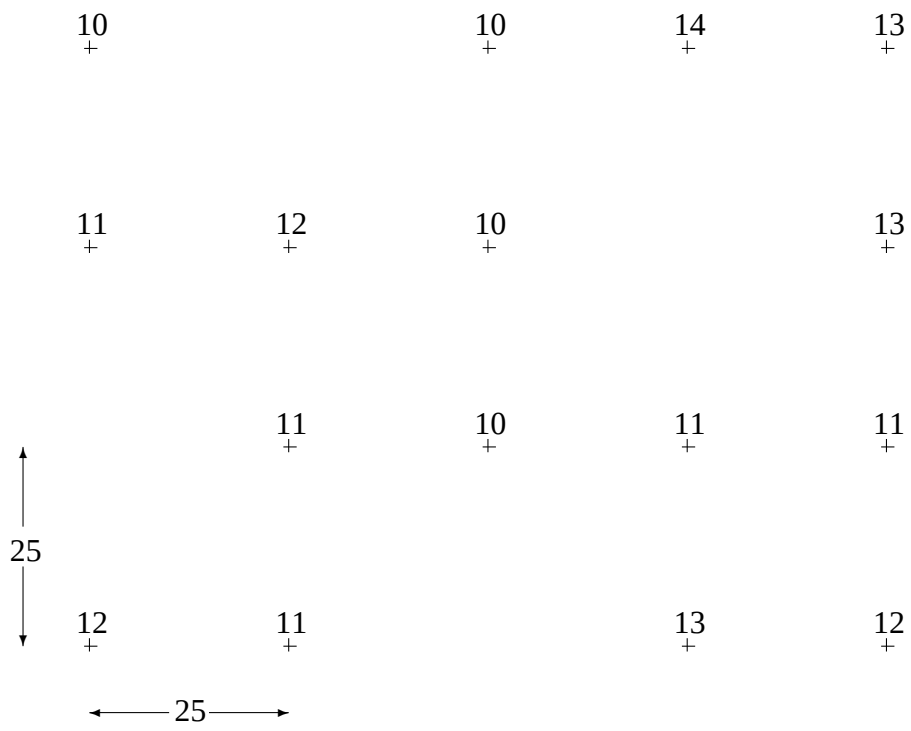


Figure 3.3: Data configuration and values for example 3.2

$\gamma^*(25_x) = 0.643$. If the number of pairs used for the calculation of the experimental variogram is large this unpleasent effect becomes less important.

3.1.1 Practice of experimental variogram calculation

Example showed that the estimation of an experimental variogram is very sensitive to extreme values (extreme differences). From this it can be concluded that in order to obtain a good estimate using (3.1) several pairs corresponding to the

vetor h are required. In general it was suggested that at least 30 pairs are required to get a more or less useful estimate.

Another practical problem is the selection of the vectors for which the experimental variogram values are calculated. It is quite common to select a few (2 to 8) directions (possibly depending on the site) and a so called lag distance. Then for each direction for multiples of the lag distance experimental variogram values are calculated (allowing a tolerance both in the direction and the distance, see equation 3.2). Of course the more directions are selected the more data are required. The calculation of the experimental variogram thus often requires several interactive steps, changing the direction tolerances and the lag distance.

Robust estimators of the experimental variogram

As example 3.2 already pointed out the experimental variogram is very sensitive to extreme values. This is because of the very skewed distribution of the squares of differences. Figure 3.4 shows the histogram of squared differences corresponding to a distance class.

It is known from statistics that in the case of skewed distributions the arithmetic mean is not the best estimator. Thus different estimators were also suggested. One of them is the formula proposed by Cressie and Hawkins (1980)

$$\gamma^*(h) = \frac{1}{2}\left(\frac{1}{N(h)}\sum_{(i,j)\in R(h)}\sqrt{|Z(x_i)-Z(x_j)|}\right)^4 (0.457+\frac{0.494}{N(h)})^{-1} \qquad (3.3)$$

This formula, based on a power transformation makes the highly skewed raw data look more similar to the normal distribution. The fourth power brings the formula back to the proper scale and the divisor adjusts it for bias.

The other concept a of robust estimator of the empirical semivariogram is the trimmed mean. The basic idea of using this estimator was to combine the advantages of expressing the central tendency via mean and via median. A mean is a good measure of central tendency if there are no extreme values in the data base. However, the mean is very sensitive to outliers. On the other hand, the median

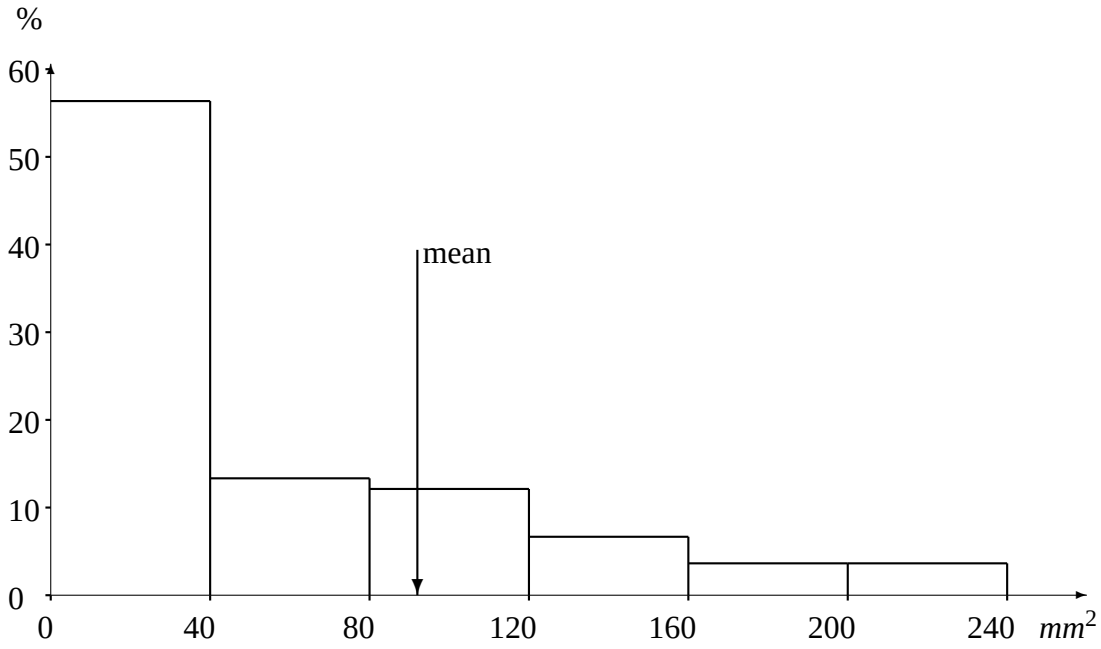


Figure 3.4: Histogram of squared differences corresponding to distance class 32 km

is a robust estimator; not contaminated by the extreme observations at all. However, when evaluating the median, one goes too far in deleting observations, as only one observed value is retained. This means that for skewed distributions, the difference between the mean and the median are unacceptably high. Trimmed mean is a natural trade-off combining the robustness of the median and the representativeness of the mean. It is calculated as a mean of the reduced data set, after elimination of some (e.g. 10 per cent) highest and some lowest observed data values. Assume, that there are n values in the sample and that the trimming is made by removing k highest and k lowest values. Then if $k/n = \alpha < 0.5$, then the trimmed mean of values $v_1, \ldots, v_n$ is:

$$M_\alpha = \frac{1}{n-2k}\left[v_{k+1} + \cdots + v_{n-k}\right] \tag{3.4}$$

where n is the number of data points, $2k$ is the number of eliminated data points

(k highest and k lowest). This formula can be used for the determination of values of experimental variogram in particular distance classes.

	Classical		Cressie Hawkins		Trimmed mean	
Distance (km)	Raw data	With one outlier	Raw data	With one outlier	Raw data	With one outlier
1.0	128.3	128.3	49.6	49.6	33.0	33.0
2.0	294.2	9903.1	152.0	220.0	120.5	120.5
3.0	405.8	405.8	298.9	298.9	196.4	196.4
4.0	484.4	6523.4	307.0	374.4	243.1	243.1
5.0	349.1	13197.7	236.8	385.7	152.8	156.8
6.0	442.5	18273.1	256.2	455.9	184.3	184.3
7.0	344.4	4674.6	255.6	295.7	165.1	165.1
8.0	435.3	22363.6	313.0	618.7	212.5	212.5
9.0	424.6	12184.0	301.0	439.7	202.4	202.4
10.0	395.6	22347.6	251.3	525.5	168.5	168.5

Table 3.2: Experimental semivariograms for chloride concentration data

Table 3.2 shows the different effect of a single extreme value on the calculated experimental variogram (calculated from 108 chloride concentration measurements). The observed value of 122 mg/l was changed to 1220 mg/l (a case which can occur quite simply). Observe the reaction of the different estimators to this single data change :

- the classical formula resulted in an unusable experimental curve

- the Cressie Hawkins formula shows some disturbances but seems still usable

- the trimmed mean shows virtually no effects at all.

3.2 The theoretical variogram

Experimental variograms are estimates of the theoretical variogram defined in equation (2.19). As experimental variograms are calculated for a finite number of vectors h, variogram values for other vectors also have to be defined. This could be done by simple linear interpolation. The disadvantage of this would be that the piecewise linear function obtained this way would not necessarily satisfy the conditions which have to hold for a variogram function defined in (2.19).

For example for any linear combination $\sum_{i=1}^{n} \theta_i Z(u_i)$, such that $\sum_{i=1}^{n} \theta_i = 0$, the variance of this combination is finite, [1] and can be calculated as:

$$Var[\sum_{i=1}^{n} \theta_i Z(u_i)] = -\sum_{j=1}^{n}\sum_{i=1}^{n} \theta_j \theta_i \gamma(u_i - u_j) \tag{3.5}$$

As the variance cannot be negative the above equation already gives a necessary condition for the variogram, i.e. that for any weights θ_i with $\sum_{i=1}^{n} \theta_i = 0$

$$-\sum_{j=1}^{n}\sum_{i=1}^{n} \theta_j \theta_i \gamma(u_i - u_j) \geq 0 \tag{3.6}$$

It can be proved that this condition is also sufficient. Unfortunately the above inequality can only be checked for a finite number of u_i and θ_i combinations. In order to relate experimental variograms to functions suitable as variograms different theoretical models were developed. These models depending whether the second order stationarity conditions hold or not form two groups.

[1] It can be proved that only linear combinations $\sum_{i=1}^{n} \theta_i Z(u_i)$ such that $\sum_{i=1}^{n} \theta_i = 0$ have a finite variance under the intrinsic hypothesis.

If the second order stationarity conditions are met then supposing that for very distant points the corresponding random variables are independent, one gets variograms which are constant after a certain distance. This is because if $Z(u)$ and $Z(u+h)$ are independent, then $C(h) = 0$ and so by (2.21) one has

$$\gamma(h) = C(0) \tag{3.7}$$

Variograms with this property are called variograms with a sill.

If the second order stationarity is not met (i.e. $C(0)$ is not finite) but the intrinsic hypothesis is true then we get variogram models without a sill.

Finally positive linear combinations of the previous variogram models also fulfil the necessary and sufficient conditions for a function to be a variogram. These are the so called complex models.

3.2.1 Variogram models with a sill

There are four commonly used elementary types of variograms with a sill. Positive linear combinations of these models are also variograms with a sill.

The pure nugget effect

The pure nugget effect corresponds to the case when there is no correlation between the random variables corresponding to different locations. This means that the value of the variogram is zero if h is zero, otherwise it is equal to the same constant which is $C(0)$ the variance of the random variable. The formula is:

$$\begin{aligned} \gamma(h) &= 0 \text{ if } h = 0 \\ \gamma(h) &= C \text{ if } h > 0 \end{aligned} \tag{3.8}$$

Figure 3.5 shows the graph of a pure nugget effect variogram.

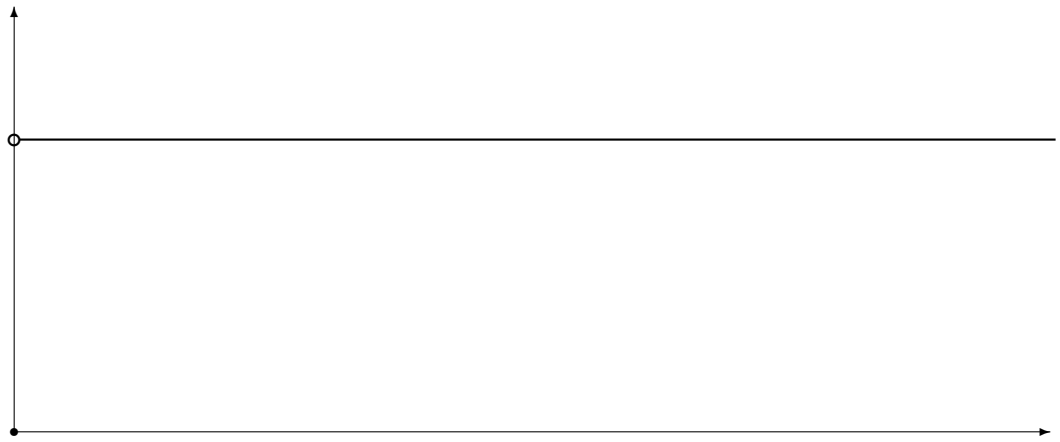


Figure 3.5: The pure nugget effect

The spherical variogram

This is the most commonly used type of variogram. It can be described by two parameters, the range and the sill. The range a is the distance which separates the correlated and the uncorrelated random variables. If two points u' and u'' are separated by a distance bigger than this range then the corresponding random variables $Z(u')$ and $Z(u'')$ are independent. Conversely if their distance is less than the range then $Z(u')$ and $Z(u'')$ are not independent. The value of the sill C is the value of the variogram for distances bigger than the range. It is equal to $C(0)$, the variance of the random variable. This implies $C > 0$. The formula is:

$$\begin{aligned} \gamma(h) &= C(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3}) \text{ if } h \leq a \\ \gamma(h) &= C \text{ if } h > a \end{aligned} \tag{3.9}$$

Figure 3.6 shows the graph of a spherical variogram.

The exponential variogram

As the spherical variogram the exponential variogram is also described with the help of two parameters. One of them is the sill, which equals $C(0)$ as for the

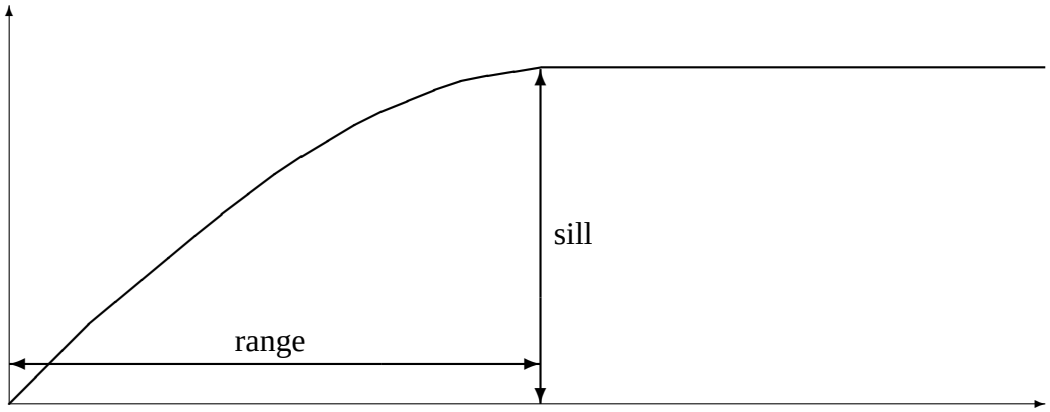


Figure 3.6: The spherical variogram

spherical variogram. The other parameter corresponds again to the change of variogram values with respect to the distance. In this case there is no special distance separating the correlated and the uncorrelated random variables as in the spherical case. All random variables are supposed to be non independent. However there is an effective range $3a$ such that random variables corresponding to points more distant than $3a$ can be considered as independent. The formula is:

$$\gamma(h) = C(1 - e^{-\frac{h}{a}}) \tag{3.10}$$

Here C is nonnegative. Figure 3.7 shows the graph of an exponential variogram.

The gaussian variogram

The gaussian variogram is also characterized by two parameters. The sill C is again equal to $C(0)$, the variance of the random variable. The parameter a is again related to the effective range of the variogram. As in the case of the exponential variogram there is no theoretical limit between correlated and non correlated random variables. The effective range in this case is $\sqrt{3}a$. The formula is:

$$\gamma(h) = C(1 - e^{-\frac{h^2}{a^2}}) \tag{3.11}$$

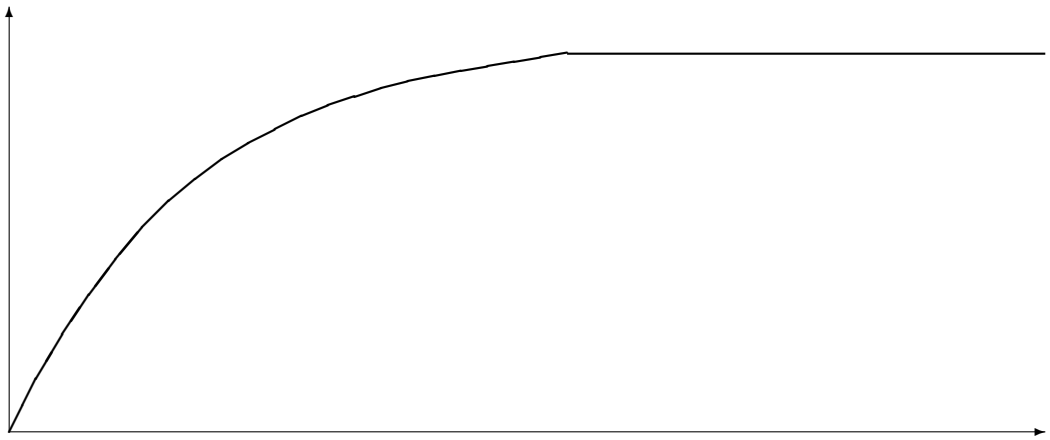


Figure 3.7: The exponential variogram

C is positive. Figure 3.8 shows the graph of a gaussian variogram.

Note the difference between the gaussian and the exponential and spherical variograms in the neighbourhood of the origin. The exponential and the spherical variograms show a linear increase, while the increase of the gaussian is much smoother - showing a quadratic type of behavoir near 0.

3.2.2 Variogram models without sill

If the regionalized variable does not fulfil the second order stationarity hypothesis but is intrinsic, then its variogram can show an unlimited increase.

Models h^λ

The function defined as:

$$\gamma(h) = Ch^\lambda \text{ for } 0 < \lambda < 2 \tag{3.12}$$

represents a valid variogram model. The case $\lambda = 1$ is the linear variogram, and it is quite often used in geostatistics. Figure 3.9 shows h^λ models for different λ values.

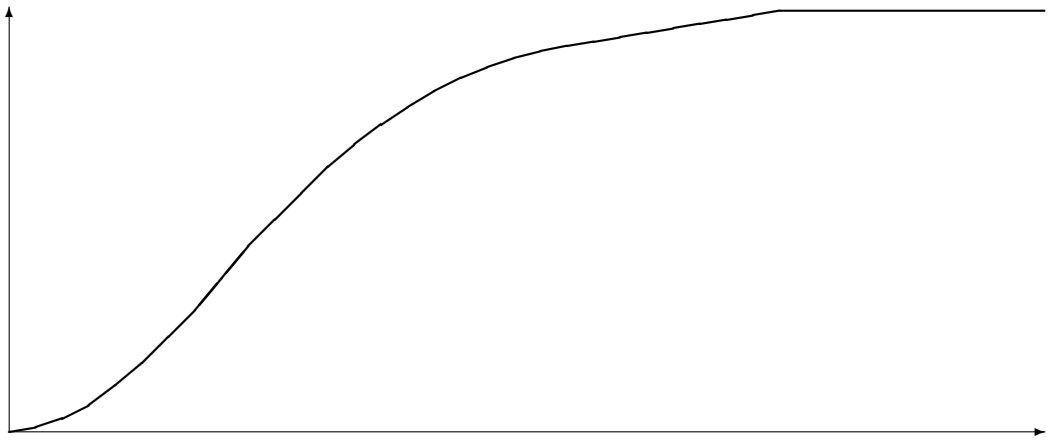


Figure 3.8: The gaussian variogram

Complex models

All previously listed variogram models satisfy (3.6). Unfortunately these models can not always describe the variability of the regionalized variable under study. Combinations of the previous models enrich the set of theoretical variograms.

It can be shown that if $\gamma_1(h), \ldots, \gamma_K(h)$ are all variogram models satisfying (3.6) and $c_1, \ldots, c_K$ are nonnegative numbers then:

$$\gamma(h) = \sum_{k=1}^{K} c_k \gamma_k(h) \tag{3.13}$$

is also a function satisfying (3.6), and thus an appropriate variogram model. Formula (3.13) makes it possible to combine models of different range describing the different types of variability of the regionalized variable. The most commonly used complex models are the combinations of a nugget effect and a simple model (like spherical).

Complex models also occur in the case when the variogram of a linear combination of regionalized variables is calculated. Suppose $Z_i(u)$ and $Z_j(v)$ are inde-

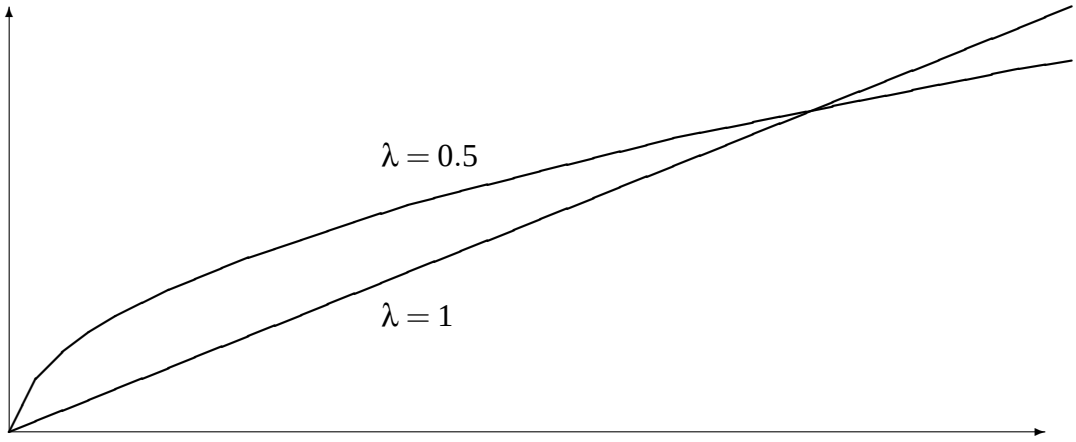


Figure 3.9: The h^λ variograms

pendent for $i \neq j$. Then for defining:

$$Z(u) = \sum_{i=1}^{I} c_i Z_i(u) \tag{3.14}$$

the variogram for Z can be calculated with the help of the variograms of the Z_i-s. Namely:

$$\gamma(h) = E\left[(\sum_{i=1}^{I} b_i Z_i(u+h) - \sum_{i=1}^{I} b_i Z_i(u))^2\right] =$$
$$= E\left[\sum_{i=1}^{I} b_i (Z_i(u+h) - Z_i(u))^2\right] = \sum_{i=1}^{I} b_i^2 \gamma_i(h) \tag{3.15}$$

Here $\gamma_i(h)$ is the variogram for Z_i.

This formula can be useful for certain non natural variables, like for example if the value of an ore is proportional to its contents of some of its components.

3.3 Variogram fitting

On the previous pages several methods and practical remarks were given for the calculation of experimental variograms. As pointed out these curves do not satisfy

the statistical properties of a variogram. Thus a theoretical curve has to be fitted to the experimental one. The previous section described several possible theoretical models, the next step is the procedure of fitting one of them to the experimental.

There are several different approaches to do this. First we have to mention that theoretical studies yielded the conclusion that the values of an experimental variogram corresponding to distant pairs are unreliable. It turned out that only the first few values can be used for finding a theoretical fit. As a rule of thumb variogram values corresponding to distances greater then the half of the greatest distance between two points in D are not considered for further use.

The most common method for fitting a variogram is doing it ”by eye”. This means that one plots the useful part of the experimental variogram and then tries to find a linear combination of theoretical models (i.e. a complex model) which produces a graph close to the experimental one. The disadvantage of this method is clear - it is not statistically justified and different experts can fit different theoretical models to the same experimental variogram. However, the great advantage of this method is that plotting the experimental curve one can detect many problems of the data set and the calculations. Extremely high or low variogram values must have reasons to be so and can be traced back. Errors of the data set (e.g. mistyping) can often be detected this way. Also the intrinsic hypothesis can partly be checked by looking at the experimental variogram. Curves increasing in certain directions and steady in others often indicate the existence of trends. Inhomogeneities of the data set can also cause problems and be detected this way. Also the correct selection of the lag and the tolerance values can be checked this way. Engineering and geological information can be used in this procedure by implicit weighting of the variogram values.

There are authors who suggest that the theoretical variogram should be fitted by a standard least squares approach. There are several problems with this approach: The method is ”blind”, the previously mentioned errors cannot be found. Another disadvantage is that this method assumes that the errors (the deviation of the theoretical from the experimental) are supposed to be independent. This

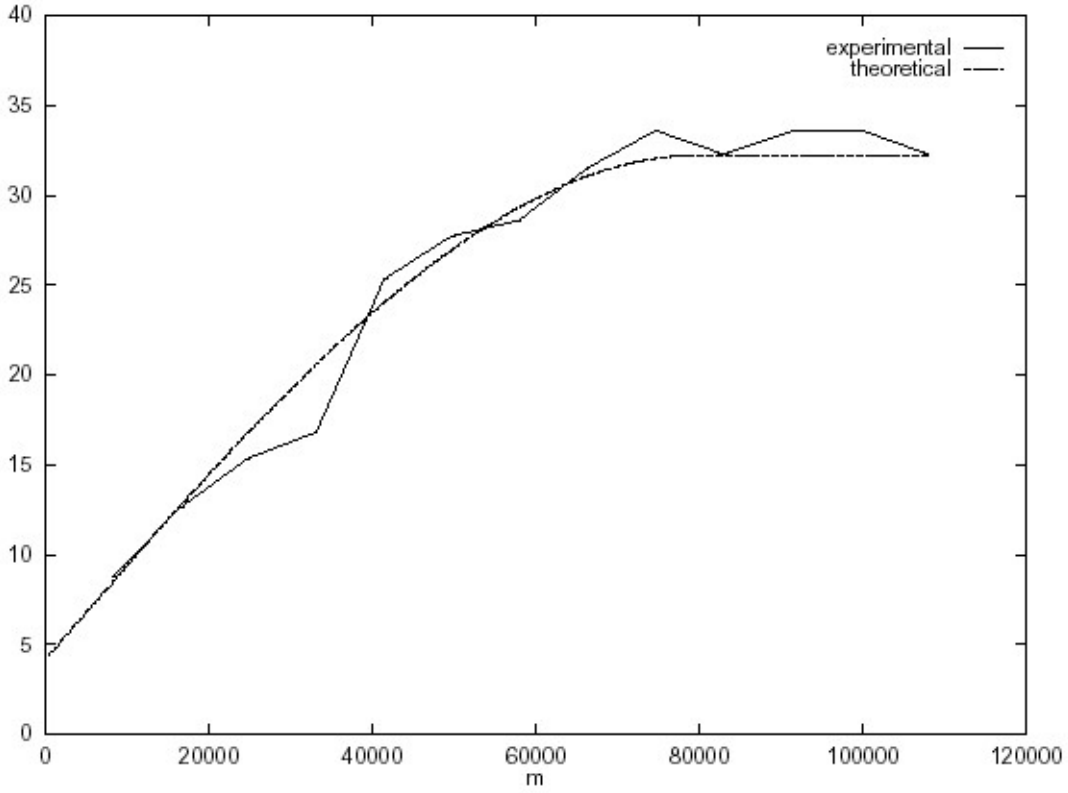


Figure 3.10: Experimental variogram with an easy fit

assumption is generally not met.

Other methods like the maximum likelihood fit were also developed. Using a maximum likelihood method one has to postulate distributions for different distance classes. These distributions are to describe the deviations of the square of the difference of two parameter values from the theoretical model. For each pair a probability depending on the parameter values can be calculated. The maximum likelihood estimator is that parameter combination which yields the highest product of these probabilities. This estimator is also ”blind” as the least squares method. It also supposes independence between the different squares corresponding to different data pairs - which is generally not met.

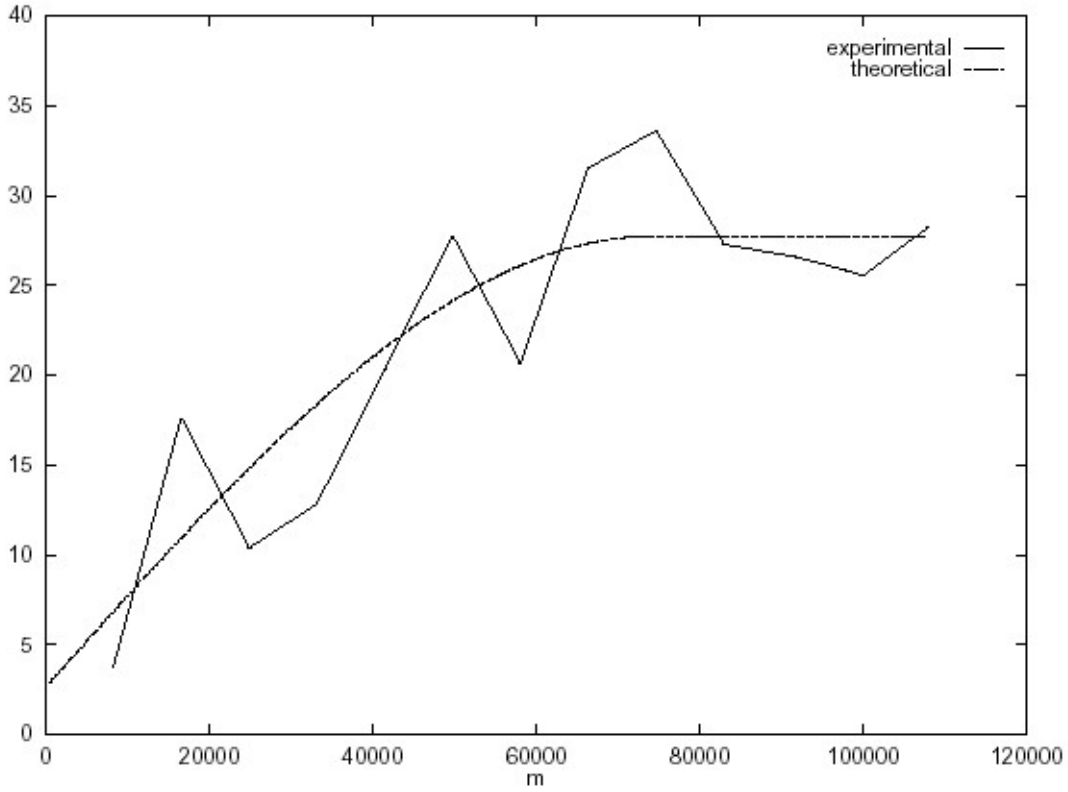


Figure 3.11: Experimental variogram with a difficult fit

Figure 3.10 shows an “easy” by eye fit, figure 3.11 shows a “difficult” case.

3.4 Isotropy — anisotropy

The random function is called *isotropic* if its variogram depends only on the length of the vector h. In this case the experimental variogram can be calculated with the only limiting condition $|u_i - u_j| = |h|$.

Isotropy of a random function can partly be checked if there is a sufficient amount of ”well spaced” (for example not alligned) data. In this case experimental variograms corresponding to different directions can be calculated and compared.

However, in many cases especially in the case of small data sets this assumption has to be made in order to have enough data for each selected class. If a random function is not isotropic, then it can show different types of anisotropy.

3.4.1 Geometric anisotropy

The regionalized variable has a geometric anisotropy if there is a coordinate transformation T such that $Z(u') = Z(Tu)$ is isotropic. This means that for geometric anisotropy a simple transformation of the coordinates leads to a case where only distances (in the new coordinate system) play a role.

The natural question arises: how does one find such a transformation? The existence of such a transformation implies that the value of the sill (if there is any) is the same for each direction. Ranges corresponding to different directions can then be plotted. If these ranges fall on an ellipse, then a rotation and a subsequent shrinking will be the appropriate transformation T. The corresponding geometric transformation is described with two parameters:

φ= the angle between the x coordinate and the main axes of the anisotropy (ellipse)

λ= the ratio of the two orthogonal ranges representing the highest and the lowest variability

The corresponding transformation has the mathematical form:

$$\begin{aligned} x' &= \lambda(x\cos\varphi + y\sin\varphi) \\ y' &= -x\sin\varphi + y\cos\varphi \end{aligned} \tag{3.16}$$

with (x,y) being the coordinates in the original and (x',y') those in the transformed system. Calculations then can be carried out in the transformed system as in the isotropic case.

In three dimensions the ellipse is replaced by an ellipsoid. In practice the variability in the vertical direction is much higher then in horizontal directions, leading to a strong anisoptropy.

3.4.2 Zonal anisotropy

If the ranges do not fall on an ellipse, or even the sill values are different then it is a *zonal anisotropy* . In the case of a zonal anisotropy a complex model has to be fitted. The individual terms of the complex model show different geometric anisotropies, and some of them might change in only one direction.

Chapter 4

Ordinary Kriging

Variograms provide a lot of information about the parameter under study, but essentially they are tools for other geostatistical calculations. One of the possible (and perhaps the most important) use of variograms is in the estimation of parameter values at unsampled locations, and/or the estimation of the average of the parameter over a certain area. The simplest geostatistical procedure doing this is ordinary kriging. Ordinary kriging is the procedure which is most widely known (and often labeled by the single word kriging).

4.1 Point kriging

One of the most common interpolation (and extrapolation) problems is the estimation of a parameter at unsampled location u. In the framework of regionalized variables this can be done with the help of the procedure labeled point kriging.

A linear estimator, i.e. a linear combination of the values of the regionalized variable at known locations, is to be found. This means that the estimator is of the form:

$$Z^*(u) = \sum_{i=1}^{n} \lambda_i Z(u_i) \tag{4.1}$$

There are infinitely many possible choices for the weights λ_i. It is desirable to

select them in order to have an unbiased estimator which also has the smallest possible estimation variance. Using the second order stationarity or the intrinsic hypothesis one has:

$$E[Z(u)] = m \text{ for all } u \in D \tag{4.2}$$

This means for the linear estimator

$$E[Z^*(u)] = \sum_{i=1}^{n} \lambda_i E[Z(u_i)] = m \tag{4.3}$$

so the weights have to fulfil:

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{4.4}$$

This is the so called unbiasedness condition. Using the second order stationarity hypothesis the estimation variance can be calculated with the help of the covariance function $C(h)$ as:

$$\begin{aligned}
\sigma^2(u) = \mathrm{Var}[Z(u) - Z^*(u)] = E\left[(Z(u) - \sum_{i=1}^{n} \lambda_i Z(u_i))^2\right] = \\
= E\left[Z(u)^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j Z(u_i) Z(u_j) - 2\sum_{i=1}^{n} \lambda_i Z(u_i) Z(u)\right] = \\
= C(0) + \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \lambda_i C(u_i - u_j) - 2\sum_{i=1}^{n} \lambda_i C(u_i - u)
\end{aligned} \tag{4.5}$$

The estimation variance is a quadratic function of the weights λ_i. The best linear unbiased estimator (BLUE) is the one which minimizes the estimation variance with respect to the unbiasedness condition. This constrained optimization problem can be solved with the help of a Lagrange multiplier μ. The function

$$\sigma^2(u) - 2\mu\left(\sum_{i=1}^{n} \lambda_i - 1\right) \tag{4.6}$$

is to be minimized. Using the partial derivatives with respect to the unknown parameters λ_i and μ one has to solve the linear equation system:

$$\sum_{j=1}^{n} \lambda_j C(u_i - u_j) - \mu \;=\; C(u_i - u) \quad i = 1, \ldots, n$$

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{4.7}$$

Solving (4.7) yields the weights λ_i for the linear estimator. The equation system (4.7) is called kriging system in terms of covariances.

If the intrinsic hypothesis is used the estimation variance can be expressed with the help of the variogram:

$$\sigma^2(u) = Var[Z(u) - Z^*(u)] = -\sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \lambda_i \gamma(u_i - u_j) + 2\sum_{i=1}^{n} \lambda_i \gamma(u_i - u) \tag{4.8}$$

The goal is to minimize $\sigma^2(u)$ under the unbiasedness conditions. This optimization problem can also be solved with the help of a linear equation system. Introducing the Lagrange multiplier μ the weights that minimize $\sigma^2(u)$ are the solution of:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \gamma(u_i - u_j) + \mu &= \gamma(u_i - u) \quad i = 1, \ldots, n \\ \sum_{j=1}^{n} \lambda_j &= 1 \end{aligned} \tag{4.9}$$

The above equation system is called kriging system, the weights λ_i are the kriging weights. The minimal estimation variance can be obtained by substituting the kriging weights into (4.8). This variance is called kriging variance $\sigma_K^2(u)$. It can be proved that :

$$\sigma_K^2(u) = \sum_{i=1}^{n} \lambda_i \gamma(u_i - u) + \mu \tag{4.10}$$

This equation is of no theoretical interest, but it simplifies the calculation of the estimation variance.

EXAMPLE 4.1 :

Suppose that using two points on a straight line the value at a third point is to be estimated. The points are $u_1 = 1$ and $u_2 = -2$. The point for which the estimation is to be done is $u = 0$. Figure 4.1 shows the configuration. Let the

measurement values be $Z(u_1) = 2$ and $Z(u_2) = 4$. Suppose the variogram is linear $\gamma(h) = h$.

Figure 4.1: Data configuration for example 4.1

The kriging equations are:

$$\begin{aligned} 0\lambda_1 + 3\lambda_2 + \mu &= 1 \\ 3\lambda_1 + 0\lambda_2 + \mu &= 2 \\ \lambda_1 + \lambda_2 &= 1 \end{aligned} \tag{4.11}$$

From this one has $\lambda_1 = 0.6667$, $\lambda_2 = 0.3333$ and $\mu = 0$. Thus $\sigma^2 = 1.3333$ and $Z^*(u) = 2.6667$. It is clear that kriging yielded the same weights as linear interpolation or inverse distance method.

Suppose the configuration is changed and u_2 is moved to the other side of the origin: $u_2 = 2$. Figure 4.2 shows the modified configuration.

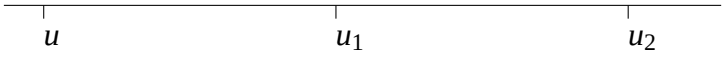


Figure 4.2: Modified data configuration for example 4.1

The kriging equations are:

$$0\lambda_1 + 1\lambda_2 + \mu = 1$$

$$\begin{aligned} 1\lambda_1 + 0\lambda_2 + \mu &= 2 \\ \lambda_1 + \lambda_2 &= 1 \end{aligned} \tag{4.12}$$

From this one has $\lambda_1 = 1.0$, $\lambda_2 = 0.0$ and $\mu = 1.0$. Thus $\sigma^2 = 2.0$ and $Z^*(u) = 2.0$. The result is different from the previous, but it would not be different in the case of the inverse distance method. This example demonstrates that the data configuration plays an important role in kriging. The increased estimation variance shows that the extrapolation in the second case is more uncertain than the interpolation in the first.

4.2 Block kriging

Quite often applications require average values of the parameter over certain areas, instead of point values. These averages could be calculated using point kriging for a great number of points in the area and taking their average. A simpler way of doing this is using block kriging.

Suppose the average of the parameter over a volume V (block) in the domain D is to be estimated.

$$Z(V) = \frac{1}{|V|} \int_{\vee} Z(u)\, du \tag{4.13}$$

Again a linear estimator of the form :

$$Z^*(V) = \sum_{i=1}^{n} \lambda_i Z(u_i) \tag{4.14}$$

is to be found. The unbiasedness condition leads again to:

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{4.15}$$

The estimation variance in this case is:

$$\sigma^2(V) = Var[Z(V) - Z^*(V)] = -\bar{\gamma}(V,V) - \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \lambda_i \gamma(u_i - u_j) + 2\sum_{i=1}^{n} \lambda_i \bar{\gamma}(u_i, V) \tag{4.16}$$

here $\bar{\gamma}$ is the average variogram value:

$$\bar{\gamma}(u_i,V) = \frac{1}{|V|}\int_{\vee} \gamma(u_i - u)\, du \tag{4.17}$$

$$\bar{\gamma}(V,V) = \frac{1}{|V|}\int_{\vee}\int_{\vee} \gamma(u-v)\, du\, dv \tag{4.18}$$

The minimization of $\sigma^2(V)$ under the unbiasedness condition leads to the linear equation system:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \gamma(u_i - u_j) + \mu &= \bar{\gamma}(u_i,V) \;\; i = 1,\ldots,n \\ \sum_{j=1}^{n} \lambda_j &= 1 \end{aligned} \tag{4.19}$$

EXAMPLE 4.2 :

Suppose that for the same configuration as in the first part of example 4.1 instead of point $u = 0$ the average over the interval $[-0.5, 0.5]$ is to be found. Block kriging is applied for the estimation. The left hand side of the equation system is identical to the point kriging case. The right hand side is:

$$\bar{\gamma}(u_1,V) = \int_{-0.5}^{+0.5} |t-1|\, dt = 1$$

$$\bar{\gamma}(u_2,V) = \int_{-0.5}^{+0.5} |t+2|\, dt = 2$$

Thus the kriging equations are again:

$$\begin{aligned} 0\lambda_1 + 3\lambda_2 + \mu &= 1 \\ 3\lambda_1 + 0\lambda_2 + \mu &= 2 \\ \lambda_1 + \lambda_2 &= 1 \end{aligned} \tag{4.20}$$

From this one has $\lambda_1 = 0.6667$, $\lambda_2 = 0.3333$ and $\mu = 0$. To calculate the estimation variance one also needs the value of $\bar{\gamma}(V,V)$. This is:

$$\bar{\gamma}(V,V) = \int_{-0.5}^{+0.5}\int_{-0.5}^{+0.5} |t-s|\, dt\, ds = 2\int_{-0.5}^{+0.5}\int_{-0.5}^{s} s - t\, dt\, ds = \frac{1}{3}$$

Thus $\sigma^2 = 1.000$ and $Z^*(V) = 2.6667$. For this case block kriging yielded the same weights as point kriging, but the estimation variance is smaller using block kriging. (The weights calculated for the center of a block using point kriging are not necessarily equal to the weights corresponding to the block !)

4.3 Properties of ordinary kriging

The kriging estimator has several interesting partly advantageous and partly disadvantageous properties. First some general properties are listed, then the relationship between kriging and the variogram is investigated.

4.3.1 Kriging as an interpolator

Kriging is an interpolation (and extrapolation) technique. Important properties of the kriging interpolator are:

1. Kriging is an exact interpolator: for each observation point u_i $Z(u_i) = Z^*(u_i)$, and the corresponding estimation variance is zero. This is because taking $\lambda_i = 1$ and $\lambda_j = 0$ if $i \neq j$ the kriging equations are satisfied.

2. Kriging weights are calculated with the help of the variogram and the locations of the measurement points and the point to be estimated. Not only distances between measurement points and the point to be estimated are considered but also the relative position of the measurement points.

3. Kriging weights sum up to 1, but they can also be negative. Thus the usual hypothesis
$$\max\{Z(u_i)\} \leq Z^*(u) \leq \min\{Z(u_i)\}$$
is not true.

4. Kriging weights are not influenced by the measurement values. If the same configuration appears at two different locations the kriging weights will be

the same, independently from the measured values. The measured values influence the variogramm which is the basis for the calculation of the kriging weights.

5. Kriging weights show a screening effect, distant points receive lower weights if closer measurements are available. This effect is demonstrated in example 4.3.

EXAMPLE 4.3 :

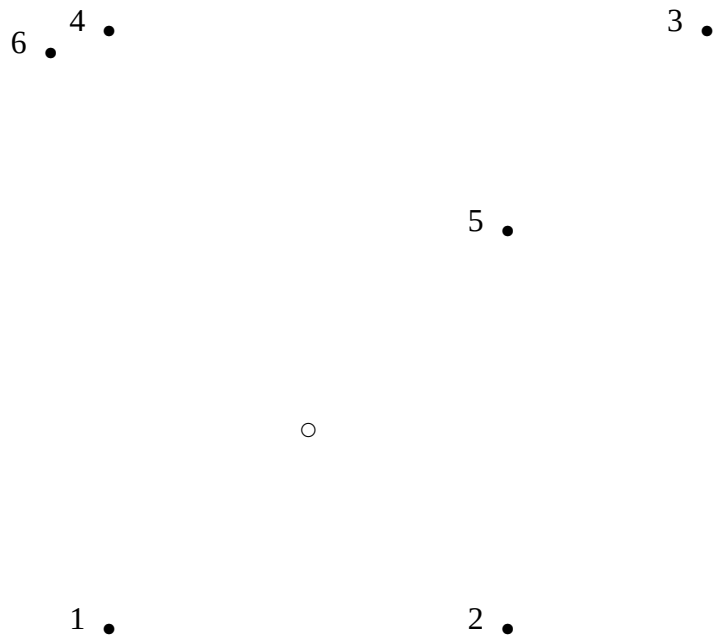


Figure 4.3: Data configuration for example 4.3

Suppose the value of the regionalized variable has to be estimated at the point (0,0) with the help of a subset of the points listed in table 4.1. The configuration is also displayed on figure 4.3. The variogram is known :

$$\gamma(h) = C_0 + C_1\gamma_S(h) \text{ for } h > 0 \tag{4.21}$$

where $\gamma_S(h)$ is a spherical model with a range $a = 10$. $C_0 = 0.05$ is the nugget effect and $C_1 = 0.20$.

Three different cases are considered:

1. kriging using points 1,2,3 and 4
2. kriging using points 1,2,3,4 and 5

No.	x	y
1	-1.00	-1.00
2	1.00	-1.00
3	2.00	2.00
4	-1.00	2.00
5	1.00	1.00
6	-1.10	1.90

Table 4.1: Different possible measurement locations

3. kriging using points 1,2,3,4 and 6.

Weights calculated for each case are shown in table 4.2.

Comparing case 1 and case 2 one can see that the weight corresponding to point 3 decreased substantially because of the inclusion of point 5. The other weights did not change drastically.

In case 3 part of the weight associated to point 4 was shifted to point 6, the other weights were much less influenced.

These two examples show that kriging filters out the useful information and assigns less weight to points which are close to other points or which are screened by other points.

4.3.2 Kriging and the variogram

As the estimation variance is calculated with the help of the variogram, and the kriging equations also contain variogram values it is obvious that the variogram plays a central role in kriging.

Using the variogram kriging delivers not only estimated values but also pro-

Point	Weights		
	Case 1	Case 2	Case 3
1	0.322	0.294	0.304
2	0.317	0.255	0.311
3	0.144	0.047	0.130
4	0.217	0.163	0.123
5	—	0.240	—
6	—	—	0.132

Table 4.2: Kriging weights for the three different cases

vides corresponding estimation variances. (Unfortunately these weights only depend on the data configuration and the variogram but not on the actual data values.) These estimation variances express the quality of the interpolation, high estimation variance means uncertain interpolation — low estimation variance shows good interpolation. Estimation variances are often used as normal error variances.

As mentioned previously the estimation variance is zero if the parameter is to be estimated at a measurement point location. In the neighbourhood the estimation variance is low (depending on the variogram) and as the distance from measurement points increases so does the estimation variance. Points (or blocks) with high estimation variances indicate areas where the estimation is uncertain.

Comparing estimation variances obtained using point and block kriging one can see that the latter are substantially smaller. This is because of the additional term $\bar{\gamma}(V,V)$ for the block variances. As $\bar{\gamma}(V,V)$ increases with the block dimensions the estimation variance decreases. This fact is in full agreement with the fact known from statistics, that a mean can be estimated with much higher accuracy than an individual value.

EXAMPLE 4.4 :

To show the role of the nugget effect consider the data of example 4.3. Three different variogram models were used to calculate the kriging weights.

$$\gamma(h) = C_0 + C_1\gamma_S(h) \text{ for } h > 0 \tag{4.22}$$

where $\gamma_S(h)$ is a spherical model with a range $a = 10$. For γ_1 $C_0 = 0.05$ is the nugget effect and $C_1 = 0.20$. For γ_2 $C_0 = 0.20$ is the nugget effect and $C_1 = 0.05$. For γ_3 $C_0 = 0.0$ is the nugget effect and $C_1 = 0.25$.

Point	Weights		
	γ_1	γ_2	γ_3
1	0.322	0.265	0.341
2	0.317	0.262	0.352
3	0.144	0.230	0.098
4	0.217	0.243	0.210

Table 4.3: Kriging weights for the three different variograms

Kriging weights for the three different models are shown in table 4.3. Note that for γ_2, where the nugget value is increased, the weights are almost equal. The highest weight differences are for the case of γ_3, where there is no nugget effect. This example shows that a high nugget effect leads to estimators around the sample mean.

If the variogram $\gamma(h)$ is replaced by its constant multiple $c\gamma(h)$ then the kriging weights do not change. This is a consequence of (4.8), as the estimation variance is also multipled by the same constant, thus the minimum variance is realized using the same weights.

If $\gamma(h)$ is replaced by another variogram which is close to it, then the kriging weights do not change substantially. Unfortunately the possible changes depend both on the configuration of the data points and the actual data values.

4.4 Practice of kriging

4.4.1 Selection of the neighbourhood

As example 4.3 already demonstrated the screening property of kriging leads to small weights for distant samples. On the other hand the intrinsic hypothesis is supposed to hold locally within a certain distance. These two facts and the numerical efficiency of the solution imply that only the closest few samples should be used in kriging.

Usually the points used for the kriging of a point or block are selected within a certain distance (usually around the range) with taking into account the anisotropy. If there are still too many points in such a neighbourhood the closest n are taken, where n is a prescribed limit.

It is important to notice that the above procedure fails to work properly if the points are very irregularly spaced. In such a case different criteria have to be given. (for example directional search)

In three dimensions when the number of points is too high a regrouping of the points into blocks and then kriging from these blocks can reduce the computations.

4.4.2 Kriging with a “false” variogram

Kriging is sometimes used also without the calculation of an experimental variogram, but only assuming a theoretical model. As mentioned above the selection of the variogram parameters can influence the kriging results. Usually a complex model of two elements a nugget effect and a simple model (spherical, exponential, gaussian or linear) is assumed. As the multiplication of the variogram by a

constant does not influence the kriging results, the most important factor in this case is the relative nugget effect (= sill divided by the nugget effect).

In any case an interpolator having the above mentioned properties is used. The estimation variances calculated without a proper variogram will be meaningless.

4.5 Cross validation

As previously mentioned the uniqueness of the realization makes the use of statistical test in geostatistics quite difficult. However, the subjective "by eye" fit of theoretical variograms should be checked somehow to reduce its effects. One possible way of doing this is the so called "cross validation". This procedure tests the variogram by a procedure where it is most often used, namely the kriging procedure.

For each measurement location u_i the values are estimated (using kriging) as if they were unknown. This estimator is now denoted by $Z^v(u_i)$ and the corresponding kriging standard deviation is $\sigma^v(u_i)$. Then the estimated values are compared with the true values $Z(u_i)$. If the kriging standard deviation can be interpreted as an estimation error with normal distribution then

$$S(u_i) = \frac{Z^v(u_i) - Z(u_i)}{\sigma^v(u_i)} \tag{4.23}$$

should be normally distributed with 0 mean and 1 as standard deviation ($N(0,1)$). The mean indicates whether the estimator is unbiased or not, the variance of S indicates the correctness of the kriging standard deviations.

The calculation of the $S(u_i)$ values with the fitted variogram is the first test of the appropriateness of the fit. If the distribution is different from $N(0,1)$ then variation of the coefficients can improve the fit.

Cross validation techniques can be used to detect outliers of the measurement values.

4.6 Kriging with uncertain data

It is quite often the case that the same parameter is measured or estimated with the help of different methods. If these methods yield different accuracies the corresponding measurement values should also be handled differently.

Suppose that for each point u_i there is an unknown error term $\varepsilon(u_i)$ having the following properties:

1. Unbiased :
$$E[\varepsilon(u_i)] = 0 \tag{4.24}$$
2. Uncorrelated :
$$E[\varepsilon(u_i)\varepsilon(u_j)] = 0 \text{ if } i \neq j \tag{4.25}$$
3. Uncorrelated with the parameter value:
$$E[\varepsilon(u_i)Z(u_i)] = 0 \tag{4.26}$$

For convenience the estimation for a block V is given here, but the same applies for point values, too. The linear estimator in this case is:

$$Z^*(V) = \sum_{i=1}^{n} \lambda_i \left(Z(u_i) + \varepsilon(u_i)\right) \tag{4.27}$$

The unbiasedness condition has to hold as in the case of ordinary kriging. So :

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{4.28}$$

The estimation variance is:

$$Var[Z(V) - Z^*(V)] = -\bar{\gamma}(V,V) - \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j\lambda_i\gamma(u_i - u_j) + 2\sum_{i=1}^{n} \lambda_i\bar{\gamma}(u_i,V) + \sum_{i=1}^{n} \lambda_i^2 E[\varepsilon(u_i)^2] \tag{4.29}$$

To minimize the estimation variance an equation system similar to the ordinary kriging system has to be solved. Namely:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \gamma(u_i - u_j) + \lambda_i E[\varepsilon(u_i)^2] + \mu &= \bar{\gamma}(u_i, V) \;\; i = 1, \ldots, n \\ \sum_{j=1}^{n} \lambda_j &= 1 \end{aligned} \tag{4.30}$$

To illustrate the above methodology consider the following example:

EXAMPLE 4.5 :

Hydraulic conductivity is measured with different methods:

1. Direct measurements
2. Gravimetric measurements
3. Nuclear measurements

In the case of gravimetric and nuclear measurements the logarithm of the hydraulic conductivity is estimated from the measured water content and the dry density with the help of a nonlinear regression. The regression error for gravimetric measurements is $D[\varepsilon_G] = 0.30997$, for nuclear measurements $D[\varepsilon_N] = 0.32828$. The measurement data are listed in table 4.4. The average log K value of the square block V with opposite corner coordinates (0,0) and (3,3) is to be estimated. Figure 4.4 shows the data configuration.

The variogram of log K was estimated on the basis of other measurement data, and a theoretical model was fitted:

$$\gamma(h) = C_0 + C_1 \gamma_S(h) \text{ for } h > 0 \tag{4.31}$$

where $\gamma_S(h)$ is a spherical model with a range $a = 6\ m$. $C_0 = 0.05$ is the nugget

No.	x	y	log K	Measurement type
1	-1.00	-1.00	-7.07	Direct
2	4.00	1.50	-7.89	Direct
3	-1.00	1.50	-6.41	Gravimetric
4	4.00	-1.00	-6.84	Gravimetric
5	4.00	4.00	-7.69	Nuclear
6	1.50	-1.00	-7.94	Nuclear

Table 4.4: Different log K measurement data

effect and $C_1 = 0.15$. The equation system (4.30) for this case is:

$$
\begin{array}{rcrcrcrcrcrcrcr}
 & + & 0.199\lambda_2 & + & 0.138\lambda_3 & + & 0.194\lambda_4 & + & 0.200\lambda_5 & + & 0.138\lambda_6 & + & \mu & = & 0.167 \\
0.199\lambda_1 & + & & + & 0.194\lambda_3 & + & 0.138\lambda_4 & + & 0.138\lambda_5 & + & 0.167\lambda_6 & + & \mu & = & 0.141 \\
0.138\lambda_1 & + & 0.194\lambda_2 & - & 0.096\lambda_3 & + & 0.199\lambda_4 & + & 0.199\lambda_5 & + & 0.167\lambda_6 & + & \mu & = & 0.141 \\
0.194\lambda_1 & + & 0.138\lambda_2 & + & 0.199\lambda_3 & - & 0.096\lambda_4 & + & 0.194\lambda_5 & + & 0.138\lambda_6 & + & \mu & = & 0.167 \\
0.200\lambda_1 & + & 0.138\lambda_2 & + & 0.199\lambda_3 & + & 0.194\lambda_4 & - & 0.108\lambda_5 & + & 0.199\lambda_6 & + & \mu & = & 0.167 \\
0.138\lambda_1 & + & 0.167\lambda_2 & + & 0.167\lambda_3 & + & 0.138\lambda_4 & + & 0.199\lambda_5 & - & 0.108\lambda_6 & + & \mu & = & 0.141 \\
\lambda_1 & + & \lambda_2 & + & \lambda_3 & + & \lambda_4 & + & \lambda_5 & + & \lambda_6 & + & & = & 1
\end{array}
$$

The solution of the equation system is shown in table 4.5. The value of $\bar{\gamma}(V,V)$ is 0.1003, the estimation variance is 0.0778 and the estimated log K value is -7.36.

In the case of ordinary kriging without error terms the kriging equations would be the same except the main diagonal being zero. The solution in this case is can also be found in table 4.5.

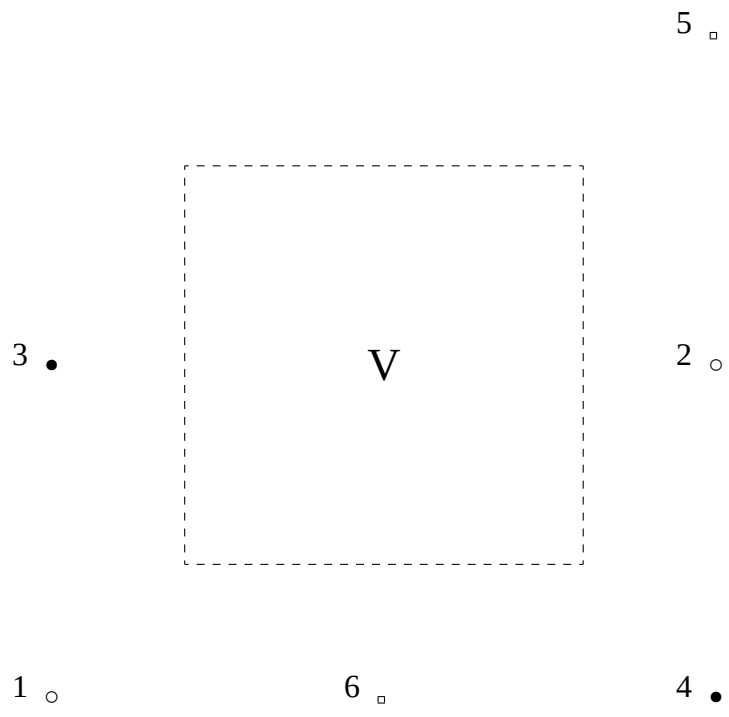


Figure 4.4: Data configuration for example 4.5

Note that observations 2,3, and 6 have similar weights as they are the closest observations to the block to be estimated. Weights for the direct measurements decreased, as all measurements are handled equally in this case.

Weights	Kriging with uncertainty	Point kriging
λ_1	0.147	0.042
λ_2	0.303	0.252
λ_3	0.210	0.294
λ_4	0.077	0.051
λ_5	0.108	0.126
λ_6	0.155	0.235
μ	0.020	0.009

Table 4.5: Weight calculated using uncertain and exact data

4.7 Simple Kriging

The Ordinary Kriging procedure is based on the assumption that the expected value of the underlying process is the same over the domain under study. The knowledge of this constant was not neccessary. Simple kriging is an alternative to OK supposing the mean $m(u)$ is known (not neccessarily constant) in the whole domain. In this case the estimator: Again a linear estimator of the form :

$$Z^*(u) = m(u) + \sum_{i=1}^{n} \lambda_i (Z(u_i) - m(u_i)) \tag{4.32}$$

is to be found. The unbiasedness condition means in this case:

$$E[Z^*(u) - Z(u)] = m(u) + \sum_{i=1}^{n} \lambda_i E[Z(u_i) - m(u_i)] - m(u) = 0 \tag{4.33}$$

This condition does not imply any additional constraints. The variance of the estimator is expressed using the covariance function C:

$$\begin{aligned} \mathrm{Var}[Z^*(u) - Z(u)] = E[Z^*(u)^2 + Z(u)^2 - 2Z^*(u)Z(u)] = \\ \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C(u_i - u_j) + C(0) - 2 \sum_{i=1}^{n} \lambda_i C(u_i - u) \end{aligned} \tag{4.34}$$

The estimation variance is minimal if:

$$\frac{\partial \mathrm{Var}[Z^*(u) - Z(u)]}{\partial \lambda_i} = 0 \tag{4.35}$$

This leads to the simple kriging equation system:

$$\sum_{j=1}^{n} \lambda_j C(u_i - u_j) = C(u_i - u) \tag{4.36}$$

Chapter 5

Non stationary methods

Unfortunately many natural parameters do not fulfil the intrinsic hypothesis, because of a known systematic change in the parameter value. For example groundwater head is a parameter which usually has a systematic drift, and thus should not be investigated with stationary methods. Systematic changes contaminate the experimental variogram and lead to unacceptable results. Some known methods are presented in this chapter to deal with this problem.

Suppose that the first assumption of the intrinsic hypothesis is not met. Namely instead of a constant expectation there is a drift present. It is supposed the the difference between the regionalized variable and the drift is intrinsic. Formally:

$$Z(u) = f(u) + Y(u) \tag{5.1}$$

where $Y(u)$ is intrinsic and $E[Y(u)] = 0$.

The most common method for estimating a drift is the use of least squares trend fitting. The assumption for fitting is that the residuals are independent. This contradicts the basic hypothesis, namely that a the regionalized variable is the sum of a deterministic drift and an intrinsic residual. (It would be true only if the residual had a pure nugget effect variogram.)

In order to deal with the drift four different methods could have been used.

- Universal kriging

- Intrinsic random functions of order k (IRF-k)
- Residual kriging
- External Drift Kriging.

A main difference between these methods is that universal kriging and residual kriging are based on a more or less explicit estimation of the drift, while IRF-k method only attempts to filter out its effect.

5.1 Universal kriging

The main problem in nonstationary cases is that the estimation of the drift would require the knowledge of the variogram, but the estimation of the variogram requires the knowledge of the drift. Universal kriging is a method where the drift paramaters are estimated in an iterative way, in order to estimate the variogram. Later in the kriging process this drift is not explicitly used, instead the effect of such a drift is filtered out.

The variogram is insensitive to constants added to the regionalized variable. Thus the drift $f(u)$ is be to found up to an additive constant. Suppose that the drift is of the form:

$$f(u) = \sum_{s=0}^{S} b_s f_s(u) \tag{5.2}$$

where $f_0(u) = 1$ and coefficients b_s are unknown, and have to be found for $s > 0$. It is supposed that 5.2 does not hold for the entire domain but only for neighbourhoods. Thus the coefficients are also "local". The next equations refer to the case of one neighbourhood. The estimators of the coefficients are taken as linear combinations of the measured values:

$$B_s = \sum_{i=1}^{n} d_{i,s} Z(u_i) \tag{5.3}$$

These estimators should be unbiased, which means that:

$$E[B_s] = b_s = \sum_{i=1}^{n} d_{i,s} E[Z(u_i)] \tag{5.4}$$

Using (5.2) one has:

$$b_s = \sum_{i=1}^{n} d_{i,s} \sum_{q=1}^{S} b_q f_q(u_i) \tag{5.5}$$

From this it follows that:

$$b_s = \sum_{q=1}^{S} b_q \sum_{i=1}^{n} d_{i,s} f_q(u_i) \tag{5.6}$$

If the functions $f_s(u)$ are linearly independent then from (5.6) it follows that:

$$\sum_{i=1}^{n} d_{i,s} f_q(u_i) \begin{cases} 1 & \text{if } q = s \\ 0 & \text{if } q \neq s \end{cases} \tag{5.7}$$

The variance of the estimator is

$$\text{Var}[B_s] = \text{Var}\left[\sum_{i=1}^{n} d_{i,s} Z(u_i)\right] \tag{5.8}$$

as only those linear combinations have a finite variance for which

$$\sum_{i=1}^{n} d_{i,s} = 0 \tag{5.9}$$

Using this one can calculate the estimation variance:

$$\text{Var}[B_s] = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{i,s} d_{j,s} \gamma(u_i - u_j) \tag{5.10}$$

This estimation variance is to be minimized under the unbiasedness conditions 5.7 . Introducing the Lagrange multipliers this leads to a set of equation systems similar to the kriging system:

$$\sum_{j=1}^{n} d_{j,s}\gamma(u_i - u_j) + \mu_{0,s} + \sum_{q=1}^{S} \mu_{q,s} f_s(u) = 0 \text{ for } i = 1,\ldots,n$$

$$\sum_{i=1}^{n} d_{i,s} = 0$$

$$\sum_{i=1}^{n} d_{i,s} f_q(u_i) \begin{cases} 1 & \text{if } q = s \\ 0 & \text{if } q \neq s \end{cases} \tag{5.11}$$

Solving the above equation systems for $s = 1,\ldots,S$ one obtains the coefficients $d_{i,s}$ and from this the b_s-s. The only problem with the above approach is that the calculation of the coefficients requires also the knowledge of the variogram. An iterative procedure can help to overcome this problem.

1. Determine the type of the drift (usually order of the polynomial).
2. Take a theoretical variogram γ, and calculate the drift coefficients.
3. Calculate the experimental variogram of the residuals.
4. Compare the theoretical variogram taken in step 2 and the calculated experimental. If the correspondance between the two curves is good then stop, else repeat from step 2 with a new theoretical variogram fitted to the experimental.

The previous procedure was ment to estimate the variogram of the nonstationary regionalized variable. If the variogram is available then the estimation of the value at a point or block can be done in a similar way to kriging. The main difference is that the drift has to be taken into account. The estimator is linear :

$$Z^*(u) = \sum_{i=1}^{n} \lambda_i Z(u_i) \tag{5.12}$$

The unbiasedness conditions in this case is:

$$E\left[\sum_{i=1}^{n} \lambda_i Z(u_i) - Z(u)\right] = 0 \tag{5.13}$$

Using equations (5.1) and (5.2) this leads to:

$$\sum_{i=1}^{n} \lambda_i \sum_{s=0}^{S} b_s f_s(u_i) - \sum_{s=0}^{S} b_s f_s(u) = 0 \tag{5.14}$$

From which:

$$\sum_{s=0}^{S} b_s \left[\sum_{i=1}^{n} \lambda_i f_s(u_i) - f_s(u) \right] = 0 \tag{5.15}$$

This equation should hold for any possible coefficients b_s. This is fulfilled if:

$$\sum_{i=1}^{n} \lambda_i f_s(u_i) - f_s(u) = 0 \text{ for } s = 0, \ldots, S \tag{5.16}$$

As the estimation variance is:

$$\sigma^2(u) = Var[Z(u) - Z^*(u)] = -\sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_j \lambda_i \gamma(u_i - u_j) + 2 \sum_{i=1}^{n} \lambda_i \gamma(u_i - u) \tag{5.17}$$

the best unbiased linear estimator is the one which minimizes $\sigma^2(u)$ under the constraints (5.16). Introducing the usual Lagrange multipliers this leads to a linear equation system:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \gamma(u_i - u_j) + \sum_{s=0}^{S} \mu_s f_s(u_i) &= \gamma(u_i - u) \;\; i = 1, \ldots, n \\ \sum_{i=1}^{n} \lambda_i f_s(u_i) &= f_s(u) \;\; s = 0, \ldots, S \end{aligned} \tag{5.18}$$

Universal kriging was the first geostatistical method dealing with non-stationary random functions. The iterative estimation of the variogram is a time consuming work, and there is no guarantee that the results will converge.

5.2 Intrinsic random functions of order k

Suppose that instead of a complete knowledge of the drift its funtional form is given:

$$f(u) = \sum_{s=0}^{S} b_s f_s(u) \tag{5.19}$$

where coefficients b_s are unknown.

The basic idea is to use increments of the sample values instead of the original values. These increments are formed in such a way that the unknown drift coefficients b_j do not influence them. $Z'(\Theta)$ is an increment if it is of the form:

$$Z'(\Theta) = \sum_{i=1}^{n} \theta_i Z(u_i) \tag{5.20}$$

here Θ stands for the vector $(\theta_1, \ldots, \theta_n)$, where θ_i-s are real numbers, such that the value of the increment is independent of the drift.

Increment $Z(\Theta)$ is independent of the unknown coefficients b_j of the drift if for all j

$$\sum_{i=1}^{n} \theta_i f_s(u_i) = 0 \tag{5.21}$$

In this case using (5.1) one has:

$$\begin{aligned} Z'(\Theta) &= \sum_{i=1}^{n} \theta_i Z(u_i) = \\ &= \sum_{i=1}^{n} \theta_i f(u_i) + \sum_{i=1}^{n} \theta_i Y(u_i) = \\ &= \sum_{i=1}^{n} \theta_i \sum_{s=0}^{S} b_s f_s(u_i) + \sum_{i=1}^{n} \theta_i Y(u_i) = \\ &= \sum_{s=0}^{S} b_s \sum_{i=1}^{n} \theta_i f_s(u_i) + \sum_{i=1}^{n} \theta_i Y(u_i) = \\ &= \sum_{i=1}^{n} \theta_i Y(u_i) \end{aligned} \tag{5.22}$$

Thus in the case of the above drift the increment is the same as for the stationary residual. This enables the calculation of a generalized covariance function. The simplest case is if the functions $f(u)$ is a polynomial of order k. In this case (5.21) can be written as:

$$\sum_{i=1}^{n} \theta_i x_i^p = 0 \tag{5.23}$$

for all $p \leq k$ in the 1 dimensional case.

$$\sum_{i=1}^{n} \theta_i x_i^p y_i^q = 0 \tag{5.24}$$

for all p,q such that $p+q \leq k$ in the two dimensional case.

$$\sum_{i=1}^{n} \theta_i x_i^p y_i^q w_i^r = 0 \tag{5.25}$$

for all p,q,r such that $p+q+r \leq k$ in the three dimensional case.
In the case of $k = 1$, which means that there is an unknown linear trend, (5.23) represents 2 equations for the 1 dimensional case, namely:

$$\sum_{i=1}^{n} \theta_i = 0 \tag{5.26}$$

$$\sum_{i=1}^{n} \theta_i x_i = 0 \tag{5.27}$$

In two dimensions (5.24) represents 3 equations, the above two and:

$$\sum_{i=1}^{n} \theta_i y_i = 0 \tag{5.28}$$

In three dimensions (5.25) represents 4 equations, the above three and:

$$\sum_{i=1}^{n} \theta_i w_i = 0 \tag{5.29}$$

If $k = 2$ (quadratic trend) $\frac{d(d+1)}{2}$ additional equations have to hold. For example in the two dimensional case:

$$\sum_{i=1}^{n} \theta_i x_i^2 = 0 \tag{5.30}$$

$$\sum_{i=1}^{n} \theta_i y_i^2 = 0 \tag{5.31}$$

$$\sum_{i=1}^{n} \theta_i x_i y_i = 0 \tag{5.32}$$

In practice higher order polynomials are rarely used.

As $Z'(\Theta)$ is supposed to be stationary, its variance can be calculated with the help of the generalized covariance function $K(h)$

$$Var[Z'(\Theta)] = \sum_{i=1}^{n}\sum_{j=1}^{n}\theta_i\theta_j K(u_i - u_j) \tag{5.33}$$

Here $h_{i,j}$ is the distance between points u_i and u_j. Matheron (1973) has shown that valid generalized covariances for IRF-k s are of the form :

$$K(h) = C\delta(h) + \sum_{r=0}^{k}(-1)^{r+1}a_r h^{2r+1} \tag{5.34}$$

Here $\delta(h) = 1$ if $h = 0$ and $\delta(h) = 0$ elsewhere. This term represents the nugget effect.

Coefficients have to fulfil some additional conditions: $C \geq 0$ $a_0 \geq 0$, $a_2 \geq 0$ and $a_1 \geq \frac{10}{3}\sqrt{a_0 a_2}$. Using a set of admissible increments and calculating the corresponding variances, according to equation (5.33) the coefficients C and a_r can be estimated. This can be done in several ways, an overview is given in Kitanidis (1983). In the present study the most traditional estimators using weighted regression techniques (Delfiner 1976), and the minimum norm estimator (Kitanidis 1983) was used.

Having identified the generalized covariance function, the minimum variance estimator of the parameter at a given location can be calculated. For weights λ_i

$$Var[Z(u) - \sum_{i=1}^{n}\lambda_i Z(u_i)] = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j K(h_{i,j}) - 2\sum_{i=1}^{n}\lambda_i K(h_i) \tag{5.35}$$

Here h_i is the distance between points u_i and u. The weights should be selected to be insensitive to the drift:

$$\sum_{i=1}^{n}\lambda_i f_j(u_i) = f_j(u) \text{ for } j = 1,\ldots,m \tag{5.36}$$

Depending on the dimension of the space and the order k (5.36) means a different number of equations. If $k = 1$ then in the one dimensional case:

$$\sum_{i=1}^{n}\lambda_i = 1 \tag{5.37}$$

and

$$\sum_{i=1}^{n} \lambda_i x_i = x \tag{5.38}$$

In two dimensions the above two equations and:

$$\sum_{i=1}^{n} \lambda_i y_i = y \tag{5.39}$$

In three dimensions the above three equations and:

$$\sum_{i=1}^{n} \lambda_i w_i = w \tag{5.40}$$

If $k = 2$ (quadratic trend) $\frac{d(d+1)}{2}$ additional equations have to hold. For example in the two dimensional case:

$$\sum_{i=1}^{n} \lambda_i x_i^2 = x^2 \tag{5.41}$$

$$\sum_{i=1}^{n} \lambda_i y_i^2 = y^2 \tag{5.42}$$

$$\sum_{i=1}^{n} \lambda_i x_i y_i = xy \tag{5.43}$$

Weights λ_i can now be calculated with the help of the following linear equation system:

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j K(h_{i,j}) + \sum_{s=0}^{S} \mu_s f_s(u_i) = K(h_i) \;\; i = 1,\ldots,n \\ \sum_{i=1}^{n} \lambda_i f_s(u_i) = f_s(u) \;\; s = 0,\ldots,S \end{aligned} \tag{5.44}$$

here the μ_s-s are the Lagrange parameters. Depending on the order k and the dimension of the space d these equations can be written with the help of the coordinates x, y, w.

As coefficients $C, a_0, a_1, a_2, \ldots$ of the generalized covariance function can be calculated with the help of regression methods. IRF-k are well suited for automatic structure identification and automatic contouring.

A substantial difference between ordinary kriging and use of IRF-k lies in the different degree of structural analysis. Variograms show several properties of the parameter which cannot be recognized using IRF-k. Different methods (Cressie - Hawkins estimators, trimmed means) can be used to obtain a useful variogram but there is no analogue for IRF-k.

EXAMPLE 6.1 :

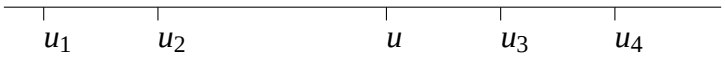


Figure 5.1: Data configuration for example 6.1

Suppose that using four points on a straight line the value at a fifth point is to be estimated. The points are $u_1 = -3$, $u_2 = -2$, $u_3 = 1$, $u_4 = 2$. The point for which the estimation is to be done is $u = 0$. Figure 5.1 shows the configuration. Suppose that Z is a non stationary random variable of the form

$$Z(u) = u^2 + u + Y(u) \tag{5.45}$$

where Y is intrinsic. Table 5.1 shows the data values.

Interpolation using IRF-0 IRF-1 and IRF-2 is investigated. The generalized covariance function is assumed to be in each case $K(h) = -h$. (Note that different constant multiples would yield the same results.)

The equation system in case of $k = 0$ is

$$\begin{aligned} 0\lambda_1 - 1\lambda_2 - 4\lambda_3 - 5\lambda_4 + \mu_1 &= -3 \\ -1\lambda_1 + 0\lambda_2 - 3\lambda_3 - 4\lambda_4 + \mu_1 &= -2 \\ -4\lambda_1 - 3\lambda_2 + 0\lambda_3 - 1\lambda_4 + \mu_1 &= -1 \\ -5\lambda_1 - 4\lambda_2 - 1\lambda_3 + 0\lambda_4 + \mu_1 &= -2 \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 &= 1 \end{aligned} \tag{5.46}$$

Point	Y	Z
u_1	-1.0	7.0
u_2	-0.5	2.5
u_3	0.0	2.0
u_4	1.0	7.0

Table 5.1: Values of Y and Z

From this one has $\lambda_2 = 0.3333$, $\lambda_3 = 0.6667$ $\lambda_1 = \lambda_4 = 0$ and $\mu_1 = 0$. Thus $\sigma^2 = 1.3333$ and $Z^*(u) = 2.1667$.

The equation system in case of $k = 1$ is

$$\begin{aligned}
0\lambda_1 - 1\lambda_2 - 4\lambda_3 - 5\lambda_4 + \mu_1 - 3\mu_2 &= -3 \\
-1\lambda_1 + 0\lambda_2 - 3\lambda_3 - 4\lambda_4 + \mu_1 - 2\mu_2 &= -2 \\
-4\lambda_1 - 3\lambda_2 + 0\lambda_3 - 1\lambda_4 + \mu_1 + 1\mu_2 &= -1 \\
-5\lambda_1 - 4\lambda_2 - 1\lambda_3 + 0\lambda_4 + \mu_1 + 2\mu_2 &= -2 \\
\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 &= 1 \\
-3\lambda_1 - 2\lambda_2 + 1\lambda_3 + 2\lambda_4 &= 0
\end{aligned} \tag{5.47}$$

For this case the solution is again $\lambda_2 = 0.3333$, $\lambda_3 = 0.6667$ $\lambda_1 = \lambda_4 = 0$ and $\mu_1 = \mu_2 = 0$. Thus $\sigma^2 = 1.3333$ and $Z^*(u) = 2.1667$. The explanation of this is that the additional equation was already fulfilled by the solution of the previous system.

The equation system in case of $k = 2$ is

$$\begin{aligned}
0\lambda_1 - 1\lambda_2 - 4\lambda_3 - 5\lambda_4 + \mu_1 - 3\mu_2 + 9\mu_3 &= -3 \\
-1\lambda_1 + 0\lambda_2 - 3\lambda_3 - 4\lambda_4 + \mu_1 - 2\mu_2 + 4\mu_3 &= -2 \\
-4\lambda_1 - 3\lambda_2 + 0\lambda_3 - 1\lambda_4 + \mu_1 + 1\mu_2 + 1\mu_3 &= -1 \\
-5\lambda_1 - 4\lambda_2 - 1\lambda_3 + 0\lambda_4 + \mu_1 + 2\mu_2 + 4\mu_3 &= -2
\end{aligned}$$

$$\begin{aligned} \lambda_1+\lambda_2+\lambda_3+\lambda_4 &= 1 \\ -3\lambda_1-2\lambda_2+1\lambda_3+2\lambda_4 &= 0 \\ 9\lambda_1+4\lambda_2+1\lambda_3+4\lambda_4 &= 0 \end{aligned} \tag{5.48}$$

For this case the solution is different, namely: $\lambda_1 = -0.25$, $\lambda_2 = 0.5833$, $\lambda_3 = 0.9167$, $\lambda_4 = -0.25$ $\mu_1 = -0.75$, $\mu_2 = 0.125$ and $\mu_3 = 0.125$. Thus $\sigma^2 = 1.5833$ and $Z^*(u) = -0.2083$. Note that the selection of $k = 2$ (the "correct" choice) yields a substantially different result. The difference between the drift $u^2 + u$ is the smallest for this case. As the number of constraints increased the estimation variance which is a constrained optimum also increased.

5.3 External-Drift-Kriging

External knowledge can be incorporated into the system with the External-Drift Kriging (EDK) (Ahmed and de Marsily 1987). Here it is supposed that an additional variable $Y(u)$ that is linearly related to the $Z(u)$ exists. The assumption of the constant expected value is thus replaced by:

$$E[Z(u) \mid Y(u)] = a + bY(u) \tag{5.49}$$

where a and b are unknown constants. The linear estimator (5.50) should be unbiased for any a and b values. The linear estimator:

$$Z(u) = \sum_{i=1}^{n} \lambda_i Z(u_i) \tag{5.50}$$

is considered. Minimizing the estimation variance under the above assumption leads to the linear equation system:

$$\begin{aligned} \sum_{j=1}^{I} \lambda_j \gamma(u_i - u_j) + \mu_1 + \mu_2 Y(u_i) &= \gamma(u_i - u) \qquad i = 1,\ldots,I \\ \sum_{j=1}^{I} \lambda_j &= 1 \end{aligned}$$

$$\sum_{j=1}^{I} \lambda_j Y(u_j) \quad = \quad Y(u) \tag{5.51}$$

where μ_1 and μ_2 are Lagrange-multipliers. The Variogramm used in (5.51) is the time invariant curve, as also used in OK. Note that the variable Y has to be known at the location x, to perform an estimation. The estimator thus depends on the additional variable $Y(u)$.

EDK is an alternative for co-kriging. EDK can be taken if the secondary information $Y(u)$ is available in a high spatial resolution, preferably regular grid. Cokriging would require the estimation of covariogramms.

Chapter 6

Indicator Kriging

6.1 Indicator Variables

There are a great number of categorical natural variables, such as lithofacies, soil types or other classes. It is often important to know the spatial extension of these variables. One possibility to do this is to use indicator variables. The indicator variable of a class C is defined as:

$$I_C(u) = \begin{cases} 1 & \text{if } u \in C \\ 0 & \text{else} \end{cases} \tag{6.1}$$

Even continuous variables can be transformed to indicators. In this case the classes are defined with the exceedence of certain selected thresholds. The indicator variable I_α for a given threshold α is defined as:

$$I_\alpha(u) = \begin{cases} 1 & \text{if } Z(u) \leq \alpha \\ 0 & \text{if } Z(u) > \alpha \end{cases} \tag{6.2}$$

The indicator variable I can also be regarded as the probability of u belonging to class C or $Z(u)$ being less than α:

Figure 6.1: Indicator transformation of $Z(u)$ for different α values

$$I_C(u) = P[u \in C] \tag{6.3}$$

If indicator coding is performed for for a numerical variable $Z(u)$) using each real value α then each observation is transformed into a step function. In practice a set of different α_k values $k = 1,\ldots,K$ is selected, and $Z(u)$ is transformed into the K dimensional vector $(I_{\alpha_1}(u),\ldots,I_{\alpha_K}(u))$.

The indicator variables $I_\alpha(u)$ are defined by dividing the measurement values into classes.

6.2 Indicator Variograms

Similarly as in the case of arbitrary numerical variables one can also calculate variograms using indicators. The experimental Indicator variogram $\gamma^*_I(h)$ is calculated as: The variogram can be estimated with the help of the following formula as in the case of the experimental variograms (Eq. 3.1):

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (I_\alpha(u_i) - I_\alpha(u_j))^2 \tag{6.4}$$

Here $N(h)$ is the number of pairs of locations separated by the vector h.

The mean of an indicator variable p equals the probability of occurence of the corresponding property. The variance of the variable is $p(1-p)$.

6.3 Indicator Kriging

$$I^*_C(u) = \sum_{i=1}^{n} \lambda_i I_C(u_i)\ . \tag{6.5}$$

This result can be interpreted as an estimator for probability of $u \in C$.

The weights λ_i are calculated from the kriging equations as in the case of OK.

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j \gamma_I(u_i - u_j) + \mu &= \gamma_I(u_i - u) \quad i = 1, \ldots, n \\ \sum_{j=1}^{n} \lambda_j &= 1 \end{aligned} \tag{6.6}$$

Indicator values can also be interpolated using SK or even EDK.

For numerical variables indicator kriging can be performed for a set of selected levels α_k. Once this is done the numerical value of $Z(u)$ can also be estimated. For this purpose the class means $\bar{Z}_k$ are introduced:

$$\bar{Z}_k = \frac{\sum_{i=1}^{n} Z(u_i)(I_{\alpha_{k+1}}(u) - I_{\alpha_k}(u))}{\sum_{i=1}^{n} Z(u_i)(I_{\alpha_{k+1}}(u) - I_{\alpha_k}(u))} \tag{6.7}$$

The estimator $Z^*(u)$ for the unknown $Z(u)$ can then be formulated as:

$$Z^*(u) = I^*_{\alpha_0}(u)\alpha_0 + \sum_{k=0}^{K} (I^*_{\alpha_{k+1}}(u) - I^*_{\alpha_k}(u))\bar{Z}_k \,, \tag{6.8}$$

with

$$\min z(u_i) = \alpha_0 < \alpha_1 < \ldots < \alpha_K = \max z(u_i)$$

Note that the indicator kriging approach has the advantages that the estimated values remain in the prescribed range:

$$[\min z(u_i), \max z(u_i)]$$

, and problems arising from highly skewed distributions are more or less overcome. On the other hand IK has disadvantages too. Even thought the observed indicator values $I_\alpha(u_i)$ are monotonic as a function of the threshold value α, this is not always true for the kriged $I^*_\alpha(u)$-s. The reason for this is that the kriging weights can also become negative. In this case first the $I^*_\alpha(u)$-s have to be altered in order to make them monotonic in α. Another problem is the discretization of the variables. The indicator coding does not distinguish between $Z(u_i) \neq Z(u_j)$ if they both belong to the same indicator class: $\alpha_k < Z(u_i), Z(u_j) < \alpha_{k+1}$ which means a loss of information and a loss of accuracy.

6.4 Applications

Indicators can very well be used for the mapping of categorical variables, such as lithofacies, soil types etc. The indicator transformation is also very useful for variable with highly skewed distributions.

6.4.1 Interpolation of a categorical variable

Consider the problem of interpolating a soil map. Observations of the soil type are available at selected locations u_i.

6.4.2 Detection limit problem

Pollutants with very low concentrations are difficult to interpolate. The main problem here is that there are measurements which are below the detection limit. If one tries to estimate the mean value of these parameters for unobserved locations assumptions of the values being below the detection limits have to be made. There are several possibilities:

1. To consider these values as zero. This is a very optimistic assumption.
2. Not to consider these data. This assumption leads to an overestimation as all small values are removed.
3. To consider all these values as a selected value between zero and the detection limit. This assumption is arbitrary and changes the uncertain data to exact values.

Instead of the above mentioned cases one can estimate the exceedence probabilities using indicator variables and indicator kriging.

EXAMPLE 6.1 :

A work on interpolation of the pollution of groundwater in Baden-Würtemberg, done in 1995, had to face the following problems dealing with the pollutant Atrazine:

- The concentrations of Atrazine were of same quantity as the detection limits, so an concrete consideration of values below the detection limits would have strong influence on the interpolation, the same problem occurs, if the data below the detection limits would have been removed (s.a.).
- Different measurements had very different detection limits, the biggest limits being multiples of the smaller ones and of same quantity as the values of most measurement points.

To deal with this data, they were coded with three indicator variables:
$\alpha_1 = 0,01\mu g/l \qquad \alpha_2 = 0,02\mu g/l \qquad \alpha_3 = 0,05\mu g/l.$

The positive effect of the indicator coding was, that only measurements with detection limits bigger than the indicator limit had to be removed from the data-set in case of being below the detection limit. In the case of a detection limit smaller than the indicator all measurements could be used: all measurements below the detection limit had value $I_\alpha = 1$ then. In this case not all low data are removed and the problem of overestimation is reduced.

In that work the data-set contained measurements of 2540 points, after removal of not usable points the data-sets of the three indicator contained:

I_3: 2437 points
I_2: 1772 points
I_1: 1449 points

On this three data-sets kriging was made with different methods as OK, Bayes-Markov-Kriging and SU. The necessary variograms were calculated for each indicator variable separately.

The figure 6.2 shows the SUK of Desethylatrazine which is a product of decomposition of herbicide Atrazine for$\alpha = 0,05\mu g/l$.

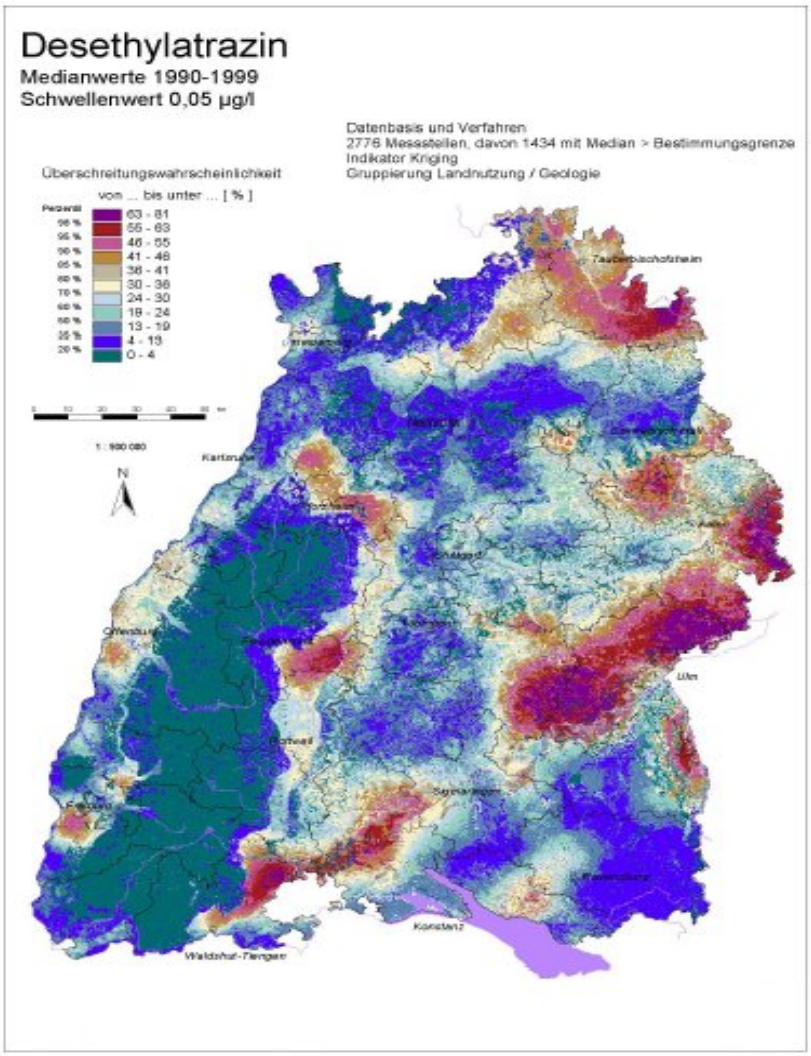


Figure 6.2: SUK of Desethylatrazine in groundwater in Baden-Württemberg

Chapter 7

Kriging with arbitrary additional information

The previously described procedures can consider additional information, but only in a numerical form, assuming a linear relationship. A non-linear relationship with the additional variable or a close connection to a censored variable is also possible. An estimation method to cope with these problems is the Markov-Bayes Kriging (MBK) described in Journel and Zhu (1990). Another possibility is to use a simple updating (SU) procedure.

7.1 Markov-Bayes-Kriging

Formally MBK uses the assumption that additional information can be taken into account for the assessment of prior distributions at selected locations.

$$U_\alpha(u) = P[Z(u) \leq \alpha | \text{additional information}] \tag{7.1}$$

If this U_α is different from the distribution $F(\alpha)$ then the additional information is useful for the estimation of $Z(u)$. The estimation is then performed by using the global prior information $F(\alpha)$, the local prior $U_\alpha(u_k)$ and the indicator coded observations $Z(u_i)$. For MBK it has to be assumed that $U_\alpha(u_k)$ is available at

considerably more locations as direct observations of Z. This way one hopes to improve the estimation of Z by using the additional information coded in U. In several applications the additional information is available for all points in the investigated domain.

The indicator values I_α are estimated using a simple cokriging approach:

$$I^*_\alpha(u) = \lambda_0 F(\alpha) + \sum_{i=1}^{n} \lambda_i I_\alpha(u_i) + \sum_{k=1}^{K} \nu_k U_\alpha(u_k) \tag{7.2}$$

In fact MBU is a "mixture" of three possible approaches:

1. assigning the same mean to the whole domain (λ_0)
2. spatial interpolation ($\lambda_i, i = 1,\ldots,n$)
3. assigning values from the additional information only. ($\nu_k, k = 1,\ldots,K$).

The spatial dependence, the configuration of the observation points and the usefulness of the additional information are influencing the role of the above factors.

For the calculation of the weights λ_i and ν one needs the covariance function of I_α and U_α and their cross-covariance function. The formulation of the equations by using variograms is also possible, but in this case the covariance based form is simpler.

According to Journel and Zhu (1990) the covariance function of U and the cross covariance function of I and U can be expressed with the help of the covariance function of I.

$$C_{IU}(h) = B(\alpha) C_I(h) \tag{7.3}$$

$$C_U(h) = \begin{cases} B^2(\alpha) C_I(h) & \text{if } h > 0 \\ B^2(\alpha) + V_f^2(\alpha) & \text{for } h = 0 \end{cases} \tag{7.4}$$

Here the quantities $B(\alpha), V_c(\alpha)$ and $V_f(\alpha)$ reflect the usefulness of the additional information. They are formally defined as

$$B(\alpha) = E[U_\alpha(x,t)|I_\alpha(x,t) = 1] - E[U_\alpha(x,t)|I_\alpha(x,t) = 0] \tag{7.5}$$

and

$$V_f^2(\alpha) = F_t(\alpha,x)\mathrm{Var}[U_\alpha(x,t)|I_\alpha(x,t)=1] + (1-F_t(\alpha,x))\mathrm{Var}[U_\alpha(x,t)|I_\alpha(x,t)=0] \tag{7.6}$$

The weights λ_i and ν_k are calculated by using simple cokriging with the single additional point x:

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i C_I(x_i - x_j) + \sum_{k=1}^{K} \nu_k C_{IU}(x_j - x'_k) &= C_I(x_j - x) \\ & \quad \text{for} \quad j = 1,\ldots,n \\ \sum_{i=1}^{n} \lambda_i C_{IU}(x_i - x'_l) + \sum_{k=1}^{K} \nu_k C_U(x'_l - x'_k) &= C_{IU}(x'_l - x) \end{aligned} \tag{7.7}$$

One can see from the equation system above that the updating of the prior functions depends on C_U and C_{IU}. The bigger the value of $V_f^2(\alpha)$ the higher is the importance of the hard information. This is reasonable as $V_f^2(\alpha)$ reflects the quality of the additional information.

The additional information $Y(x)$ available at each location x'_k is used to define the variable U_α as

$$U_\alpha(x,t) = \frac{1}{N(x)} \sum_{Y(x'_k)\approx Y(x)} I_\alpha(x_j,t) \tag{7.8}$$

$N(x)$ is here the number of observation points which have similar Y values ($Y(x'_k) \approx Y(x_j)$). In practice this is done the way that classes are defined for the variable $Y(x)$, and the mean indicator values over the classes are assigned to the unobserved locations as prior information. Note that $Y(x)$ can be any classification, both numeric values can be grouped into classes or categorical variables, such as land use. The quantities $B(\alpha)$ and $V_f(\alpha)$ can easily be calculated once U_α has been defined.

7.2 Simple Updating (SU)

Consider the situation where $Z(x)$ is complemented by a secondary variable $L(x)$ available at each point in the domain. This variable is discrete, and related to $Z(x)$

through a conditional expectation:

$$E[Z(x) \mid L(x) = l] = m_l \tag{7.9}$$

and a conditional variance:

$$\mathrm{Var}[Z(x) \mid L(x) = l] = \sigma_l^2 \tag{7.10}$$

In this case the first estimation of $Z(x)$ is that based on $L(x)$

$$Z'(x) = m_{L(x)} + \varepsilon_{L(x)}(x) \tag{7.11}$$

with $\varepsilon_{L(x)}(x)$ being a random error with 0 expectation and variance $\sigma^2_{L(x)}$. Using $Z'(x)$ for the estimation of $Z(x)$ combined with the observations one has:

$$Z^*(x) = \lambda_0 Z'(x) + \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{7.12}$$

The estimation variance for this estimator is:

$$\begin{aligned}\mathrm{Var}[Z(x) - Z^*(x)] = &- \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \lambda_i \gamma(x_i - x_j) + \\ &+ 2\sum_{i=1}^{n} \lambda_i (1-\lambda_0)\gamma(x_i - x) + \lambda_0^2 E[\varepsilon(x)^2]\end{aligned} \tag{7.13}$$

Minimizing this estimation variance with respect to the unbiasedness condition leads to the equation system:

$$\begin{aligned}\sum_{j=1}^{n} \lambda_j \gamma(x_i - x_j) + \mu &= (1-\lambda_0)\gamma(x_i - x) \;\; i = 1, \ldots, n \\ \sum_{j=1}^{n} \lambda_j \gamma(x - x_j) + \mu &= \lambda_0 \sigma_l^2 \\ \sum_{j=0}^{n} \lambda_j &= 1\end{aligned} \tag{7.14}$$

This is a linear equation system for the unknown weights λ_i, and the solution yields the estimator $Z^*(x)$. The additional variable $L(x)$ can be any discrete variable, for example soil type, landuse or geological code.

In practice we suppose that additional information in the form of a discrete index $(1,\ldots,M)$ $L(x)$ is available for every location in the field. For each class l the mean ant the variance can be calculated as

$$m_l = \frac{\sum_{i=1}^{n} Z(x_i) 1\{L(x_i) = l\}}{\sum_{i=1}^{n} 1\{L(x_i) = l\}} \tag{7.15}$$

$$s_l^2 = \frac{\sum_{i=1}^{n} (Z(x_i) - m_l)^2 1\{L(x_i) = l\}}{(\sum_{i=1}^{n} 1\{L(x_i) = l\}) - 1} \tag{7.16}$$

EXAMPLE 7.1 :

A work on interpolation of the pollution of groundwater in Baden-Württemberg, done in 1995, intended to optimize the interpolation by use of additional information, available for the whole area of interest. Available information was a classification of landuse in 16 classes, and a classification of the groundwater geology in 19 classes.

The pH-value was interpolated with both classifications using MBK. For this purpose the classifications had to be revised by combining smaller classes, to achieve classes with at least 18 points of measurement. This was necessary to get statistical values of this classes of sufficient quality. This grouping was done manually and reduced the number of classes to 13 classes of landuse and 15 of geology.

The data were coded to 9 indicator variables which correlations were calculated. The distribution of data of the classes of landuse and geology were calculated and also their correlations and the cross-correlations to the indicator functions. With these prior informations a BMK was calculated. This contains a big amount of calculations, especially 9 variograms for the indicator variables.

EXAMPLE 7.2 :

The same data have been optimised with the SU-algorithm in a second work on the pollution of groundwater in Baden-Württemberg in 1996. Because SU affords a significant less amount of calculation and especially only one variogram has to be fitted, it has been possible to spend more work on the optimisation of the classification.

For this purpose a combined classification of landuse and geology has been made, containing 304 classes. These had to be combined to classes with at least 18 points of measurement. For that work algorithms had been designed, combining classes under aspects of statistical similarities in the parameter of interest and under the aspect of similarities in their character (combining classes with forest-type landuse to each other etc.). This lead to classifications of a quantity of fifty to sixty classes for a data-set containing 2540 points of measurement.

These classifications were used for a SU-algorithm. In table 7.1 the Mean-Square-Error (MSE) of cross-validations of some of that calculations are presented (for two different groupings of the classes), compared to results of example 7.1 and of an OK of the same data-set. The ranking of interpolation of this examples are not general. Calculations on other data-sets (e.x. NO_3) presented different rankings, at least SU was the most powerfull algorithm for all seven data-sets in that investigation.

Table 7.1: Cross-Validation for pH-value

algorithm	OK	SU(Grouping 1)	SU(Grouping 2)	MBK(Geology)	MBK(Landuse)
MSE	0,089	0,0821	0,0822	0,086	0,097

EXAMPLE 7.3 :

Another parameter investigated in the work on the pollution of groundwater in Baden-Württemberg was Nitrate. The methods have been the same as in the

previous examples. To give some impression about the differences in interpolation some results are presented in the following.

The measurements of Nitrate have been interpolated with different methods, the results of OK, MBK and SU are presented in the figures 7.3, 7.4 and 7.5. To get an impression about the data-base figure 7.1 presents a map of the points of measurement used for the interpolations. The additional informations used for the interpolations are the landuse for the MBK and a combination of landuse (see figure 7.1) and geology (see figure 7.2) for the SU. Figure 7.6 shows another SUK (Simple Updating Kriging) of Nitrate in groundwater in Baden-Württemberg.

Nitrate is one of the seldom examples, where the OK had better results than the MBK when comparing the Mean-Error of the cross validation. Nevertheless the MBK seems to be more trusting because OK gives an overestimation in the surroundings of extreme big values and is not capable to work out spatial structures caused by the character of the regions like the Black Forest. SU is here assumed to be the best method, because it has the lowest error in cross validation combined with clearly outworked structures in the interpolation. The results of cross validation are shown in table 7.2.

Table 7.2: Cross-Validation for Nitrate $[(mg/l)^2]$

method	OK	SU(Grouping 3)	BMU(landuse)
MSE	374	337	385

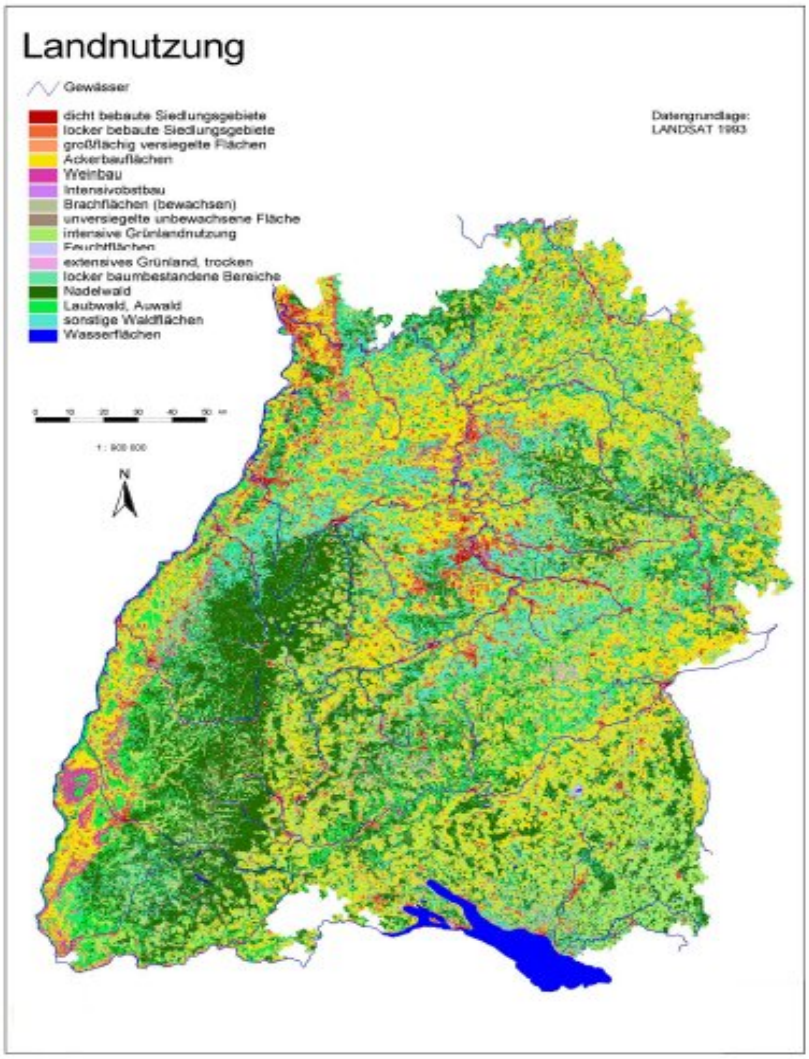


Figure 7.1: Points of groundwater-measurement and landuse in Baden-Württemberg

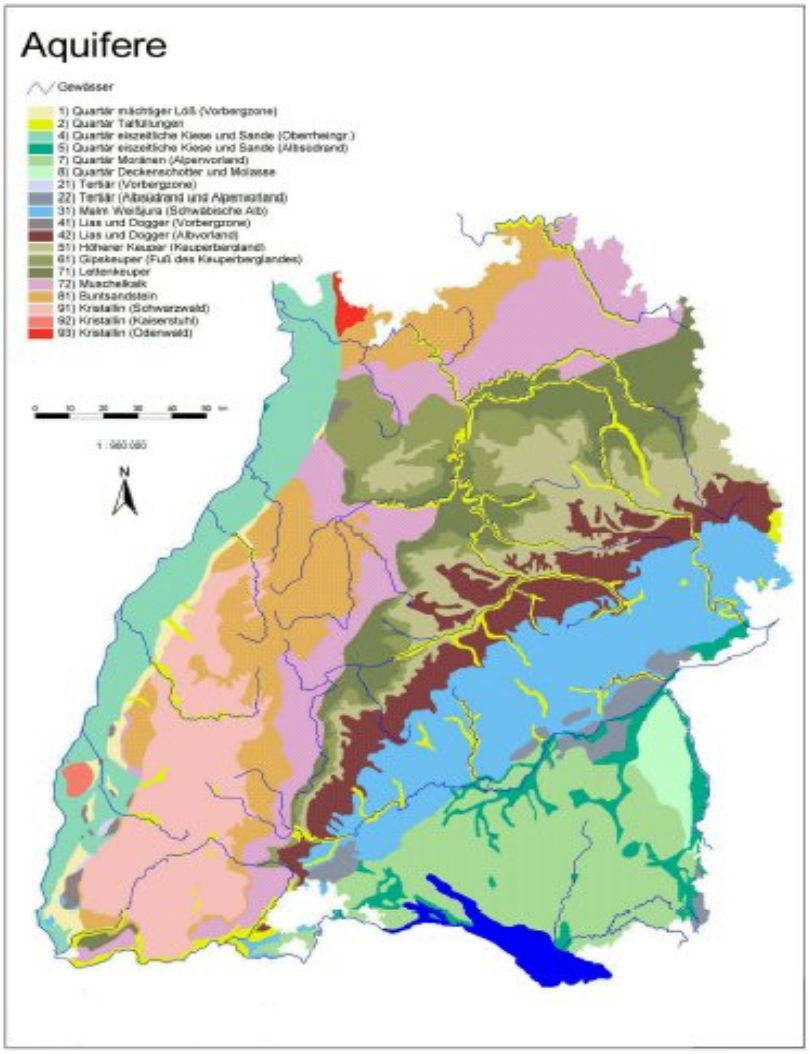


Figure 7.2: Aquifers in Baden-Württemberg as additional hydrogeological information to SU

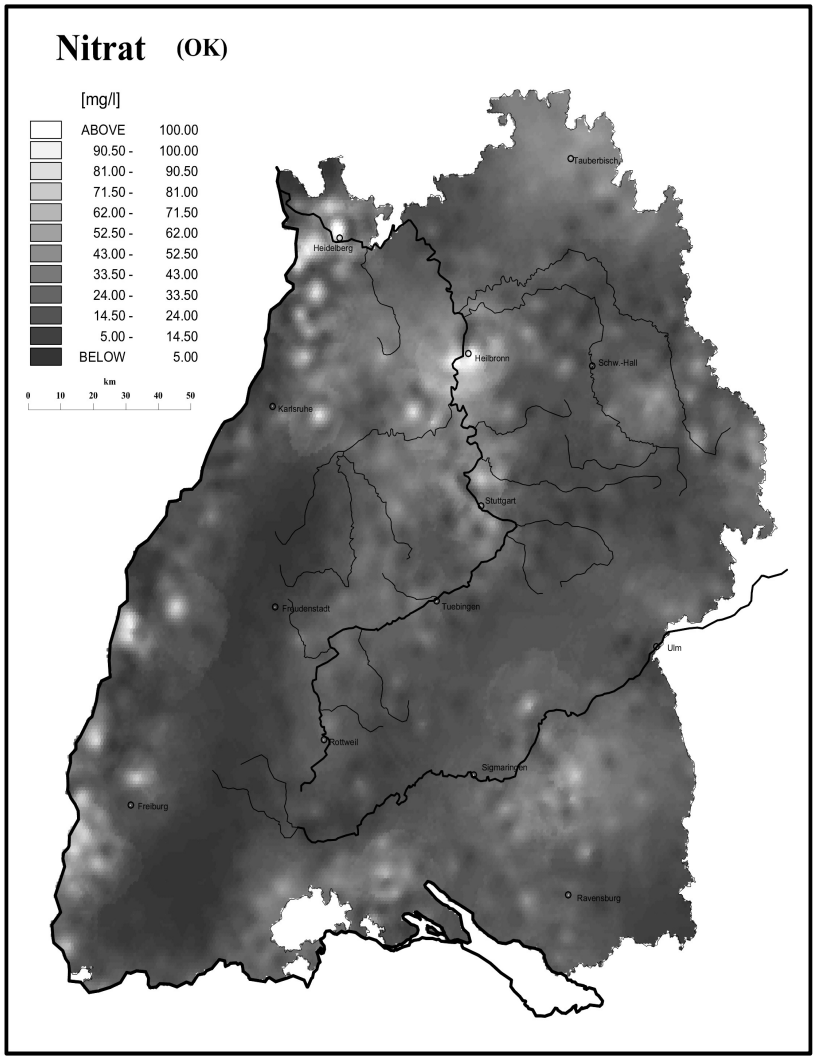


Figure 7.3: OK of Nitrate in groundwater in Baden-Württemberg

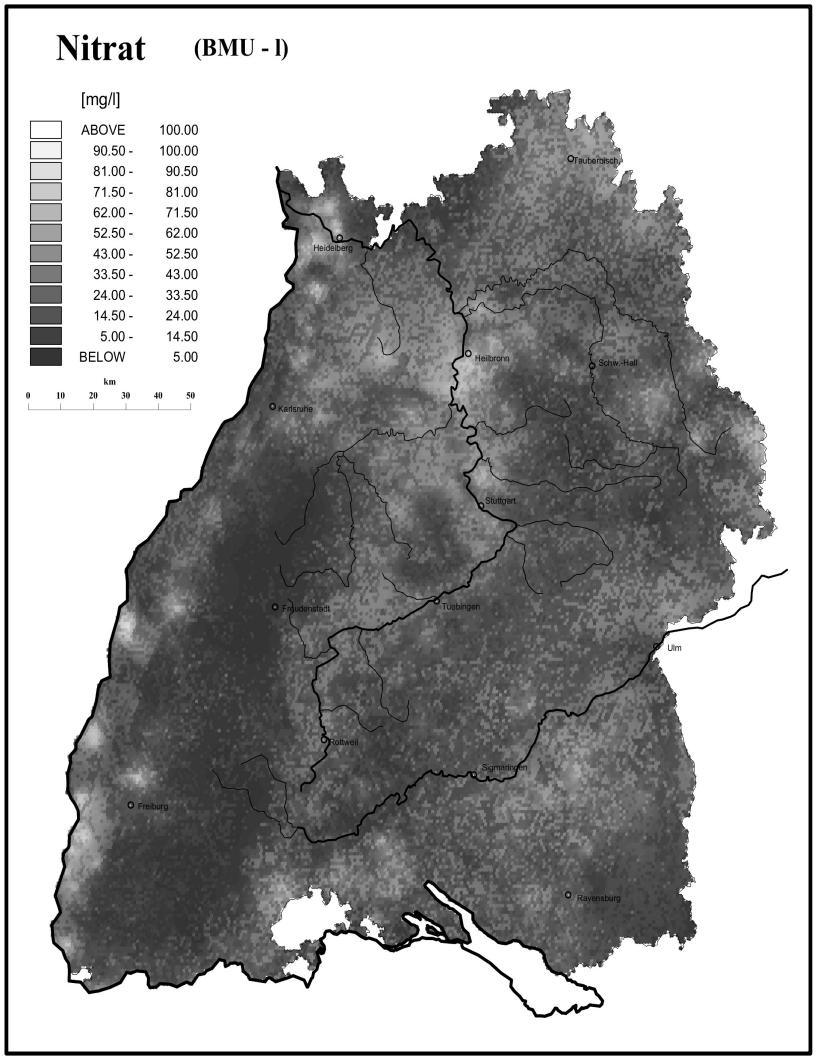


Figure 7.4: MBK of Nitrate in groundwater in Baden-Württemberg

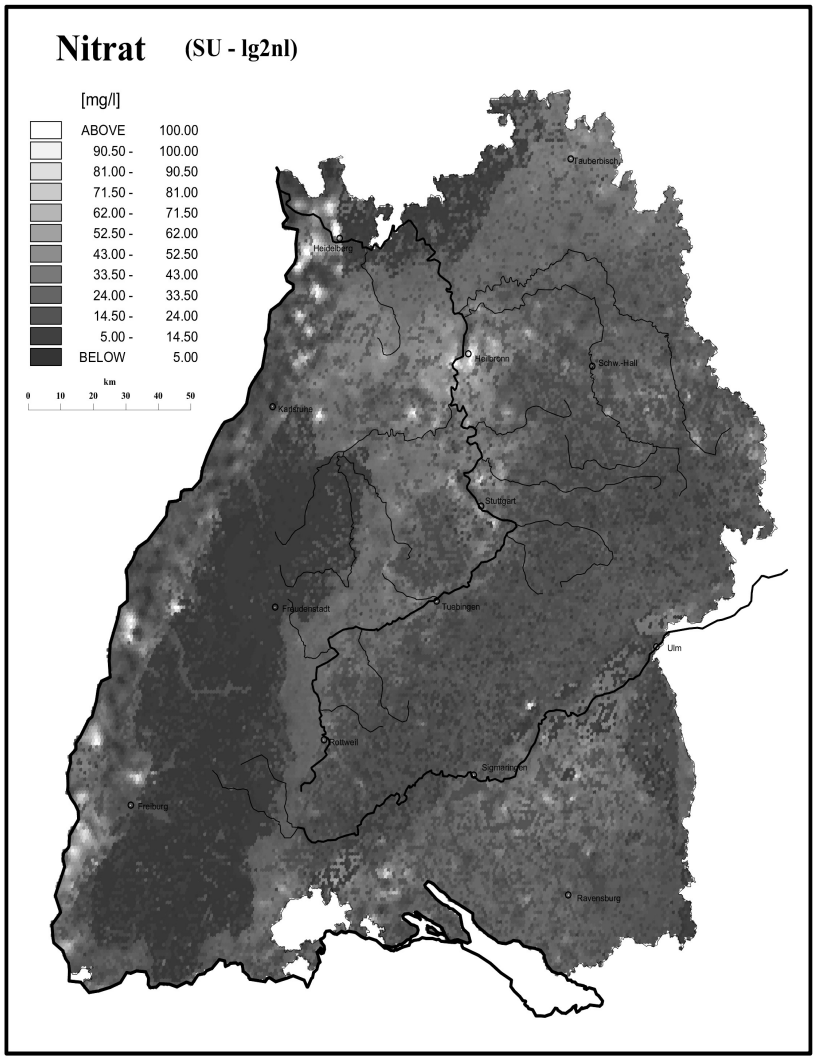


Figure 7.5: SU of Nitrate in groundwater in Baden-Württemberg

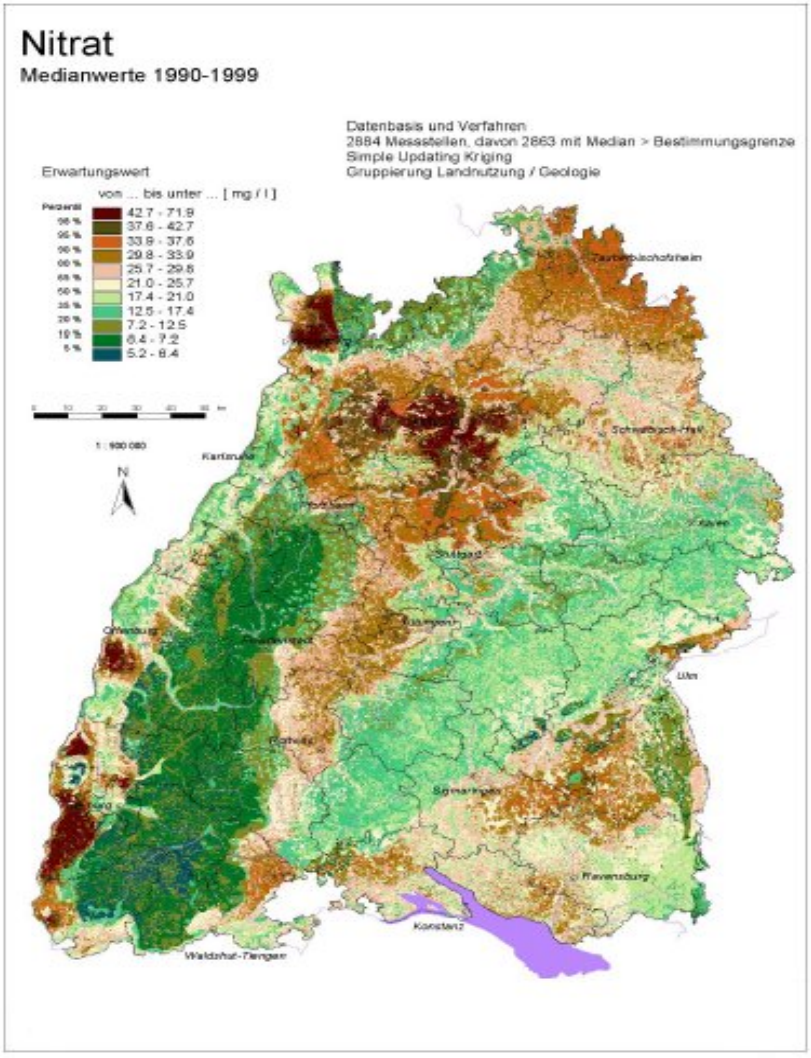


Figure 7.6: SUK of Nitrate in groundwater in Baden-Württemberg

Chapter 8

Time dependent variables

Geostatistical methods were originally thought for mining and geological problems, where at a certain location only one measurement (borehole) can be made, and a single realization is observed. However in many other applications the same location can be used for several measurements. For example groundwater quality or quantity parameters or precipitation is often measured regularly in time. The question is how to model and how to use this in geostatistical evaluations. Unfortunately because of the interest of geostatisticians in single realizations there is not much in the literature about this problem.

A possible method to include time is extending the intrinsic hypothesis to the time dimension. This means that measurement locations consist of two parts: a spatial (1,2 or 3 dimensions), and a temporal. This approach is more or less reasonable for timewise continuous variables like groundwater quality parameters. It is unreasonable for precipitation (one cannot use the precipitation of June 1 and June 30 to calculate the precipitation for June 15) and other event-based parameters.

Another possible extension is to use the data corresponding to the same time as a realization, and to suppose that the different realisations correspond to the same or at least to a similar process. (The notion of similar process will be explained later). This second method does not exclude the first, “time cuts” of an intrinsic

space-time process are also intrinsic in space, and the spatial variograms are the same.

8.1 Space — time intrinsic variables

The random function $Z(u,t)$ is space time intrinsic if:

$$E[Z(u,t)] = m \tag{8.1}$$

The space time variogram:

$$\gamma(h,\Delta t) = \frac{1}{2}\mathrm{Var}[Z(u+h,t+\Delta t) - Z(u,t)] \tag{8.2}$$

is independent of the location u and the time t.

For example slowly changing groundwater quality parameters often show such a space time intrinsic behavior.

A serious problem in calculating space time variograms is that there is no common distance measure. Spatial distances can be calculated, and time differences, too. It is important to find the spatial equivalent of a time difference. This can be done by calculating the experimental variograms for time and for the spatial stucture separately.

$$\gamma_T^*(\Delta t) = \frac{1}{2N_T(h)} \sum_{(i,j)\in R_T(h)} (Z(u_i,t_i) - Z(u_j,t_j))^2 \tag{8.3}$$

Here

$$R_T(h) = \{(i,j); \Delta t - \varepsilon \le |t_i - t_j| \le \Delta t + \varepsilon \text{ and } (u_i = u_j)\} \tag{8.4}$$

and $N_T(h) =$ the number of elements in $R_T(h)$. For the spatial structure:

$$\gamma_S^*(h) = \frac{1}{2N_S(h)} \sum_{(i,j)\in R_S(h)} (Z(u_i,t_i) - Z(u_j,t_j))^2 \tag{8.5}$$

where

$$R_S(h) = \{(i,j); h - \varepsilon \le |u_i - u_j| \le h + \varepsilon \text{ and } |t_i - t_j| \le \delta\} \tag{8.6}$$

and $N_S(h) =$ the number of elements in $R_S(h)$ There two possibilities:

1. The two kinds of experimental variograms are similar, have the same nugget effect and the same sill. This means that a geometric anisotropy can be considered between the time and spatial scale. In this case a linear transformation of the time scale leads to a space—time isotropic model. The new "length" of a vector $(h, \Delta t)$ is defined as:

$$|(h, \Delta t)| = \sqrt{|h|^2 + k_t |\Delta t|^2} \tag{8.7}$$

2. The two kinds of experimental variograms are different, having different shape and/or sill etc. In this case a zonal anisotropy type of modelling can be used. Namely the space—time variogram $\gamma_{ST}(h, \Delta t)$ can be written in the form:

$$\gamma_{ST}(h, \Delta t) = \gamma_S(h) + \gamma_T(\Delta t) \tag{8.8}$$

In both cases space—time kriging and space time simulation can be done similarly as in the spatial case.

8.2 Spatially intrinsic variables with time independent variograms

The random function $Z(u,t)$ is spatially time intrinsic with time independent variograms if:

$$E[Z(u,t)] = m \tag{8.9}$$

The spatial variogram :

$$\gamma(h) = \frac{1}{2}\mathrm{Var}[Z(u+h, t+\Delta t) - Z(u,t)] \tag{8.10}$$

is independent of the location u and the time t if $\Delta t \leq \delta$.

8.3 Spatially intrinsic variables with time dependent variograms

The random function $Z(u,t)$ is spatially time intrinsic with time dependent similar variograms if:

$$E[Z(u,t)] = m(t) \tag{8.11}$$

and the spatial variogram at the given time t:

$$\gamma(h,t) = k(t)\frac{1}{2}\mathrm{Var}[Z(u+h,t+\Delta t) - Z(u,t)] \tag{8.12}$$

is independent of the location u if $\Delta t \leq \delta$ and $k(t)$ is a time dependent function.

There are several possible choices for the function $k(t)$. For example

- Mean proportional variograms :

$$k(t) = m(t)^2 \tag{8.13}$$

  This means that $\frac{Z(u,t)}{m(t)}$ is spatially intrinsic with a time independent variogram.

- Variance proportional variograms :

$$k(t) = \mathrm{Var}[Z(u,t)] \text{ with fixed } t \tag{8.14}$$

  This means that the correlation structure is preserved over the time.

8.4 Time series interpreted as different realisations

In the case of fast changing or event-based parameters, time series can be used for a far-reaching analysis of the spatial correlation structure of the obtained data. This requires that we can assume the observed process as similar during time, but similarity is only necessary in the correlation of the events in spatial distributed

points of measurements. If this assumption is possible, can be detected by calculating the *correlation coefficient* ρ for time series of pairs of the locations of measurement u_i and u_j:

$$\rho_{ij} = \frac{C[Z(u_i,t),Z(u_j,t)]}{\sqrt{Var[Z(u_i,t)] \cdot Var[Z(u_j,t)]}} \tag{8.15}$$

with *temporal covariance C*:

$$C[Z(u_i,t),Z(u_j,t)] = E[\{Z(u_i,t) - E[Z(u_i,t)]\}\{Z(u_j,t) - E[Z(u_j,t)]\}] \tag{8.16}$$

The correlation coefficient is the standardized temporal covariance, the value 1 shows a strict positive linear relation between the two time-series, a 0 indicates two time-series without any (linear) relation, negative values indicate negative linear relations. If calculated for a number of pairs, the correlation coefficient denoted as a function of the distance between those pairs should show a similar figure then a spatial covariance function (c. chapter 2.3).

If the assumption of similarity in time is met, the correlation coefficients can be used mainly in two ways:

1. The covariances according to the correlation coefficients are a quite better base for calculating a spatial variogram than the measured values for a single time. For this they can be used directly as a covariance cloud to estimate a spatial covariance, similar to a variogram cloud used for the estimation of a variogram, and then used for kriging calculation. For this purpose it is easily possible to express the Kriging conditions in terms of correlation instead of variogram (c. chapter 4.1).

2. The information contained in the spatial correlation structure can be used for a further optimisation of a theoretical correlation function. For this purpose spatial transformations can be calculated on this base.

EXAMPLE 8.1 :

For example precipitation is known to depend on the height and on the slope of the terrain observed. In actual research an optimisation of the spatial correlation structure is aimed by a transformation in a high dimensional space. In this transformed space height and the slope are additional dimensions and the transformation includes several scaling parameters which are fitted to a theoretical spatial correlation function. The calculation fits seven spatial scaling parameters plus two free parameters for the theoretical correlation function which is quite a big amount (which can only be done numerically) but the results show a significant lesser discrepancy between theoretical and experimentally observed correlation.

Chapter 9

Simulation

Kriging as most interpolation techniques delivers idealized smooth results. This is because the minimum estimation variance as optimality criterion necessarily yields less variable estimators. If an experimental variogram is calculated from the kriged values, then it is different from the one obtained from measurements. Variances corresponding to different distances are usually much smaller for the kriged values. In many cases the variability of the regionalized variable plays a central role for decisions (for example reliability aspects). Therefore a procedure to obtain interpolation reproducing the variogram of the original variable is needed. Simulation is the method for doing this.

Simulation should reproduce the variability of the regionalized variable. The simulated values should have the same mean, variance and variogram as the measured ones. Quite often the histogram of the measured values should also be reproduced. Simulation should deliver one possible reality.

Simulation is very useful in the case of parameters which are themselves not a final product of the analysis. For example in groundwater modeling hydraulic conductivity is an essential parameter, but it is an input of a subsequent model. The pupose of modeling is to deliver accurate head values, which are dependent on the hydraulic conductivities. Expected head values are not the same as the head values calculated with expected conductivities.

9.1 Basic definitions

Simulation methods form two different groups:

1. methods generating realizations with given spatial variablity and distribution
2. methods gegenerating realizations with given spatial variablity under consideration of additional information

In both groups depending whether the simulated realizations honor the observations or not two cases are distinguished:

1. unconditional simulations: delivering realizations with prescribed variablity without honoring the observation data
2. conditional simulations: delivering realizations with prescribed variablity honoring the observation data

There are at least four different ways simulating realizations of $Z(u)$ under the above constraints:

- Monte Carlo simulations
- turning band simulations
- sequential simulations
- simulated annealing

9.2 Monte Carlo Simulations

Monte Carlo simulations generate realizations multidimensional (generally normal) distributions. The simulation uses Y_j independent standard normally (N(0,1))

distributed random variables Y_j. Let the variables Z_i be linear combinations of the Y_j-s:

$$Z_i = \sum_{j=1}^{N} c_{ij} Y_j$$

The covariance of Z_i and Z_k is:

$$\mathrm{Cov}(Z_i, Z_k) = \mathrm{Cov}(\sum_{j=1}^{N} c_{ij} Y_j, \sum_{l=1}^{N} c_{kl} Y_l) = E(\sum_{j=1}^{N} c_{ij} Y_j \sum_{l=1}^{N} c_{kl} Y_l) =$$

$$= E(\sum_{j=1}^{N} c_{ij} c_{kj} Y_j Y_j) = \sum_{j=1}^{N} c_{ij} c_{kj} \tag{9.1}$$

Here we used the independence of the Y_j-s.

In this case the covariance matrix Γ is decomposed as the product of a matrix and its transpose:

$$\Gamma = \mathbf{C}\mathbf{C}^{\mathbf{T}} \tag{9.2}$$

This latter matrix is then used to transform a vector of independent $N(0,1)$ random variables to the required multidimensional distribution. The disadvantage of this method is that the dimension of the covariance matrix equals the number of points for which simulated values should be calculated, thus the computational requirements of the method increase drastically with the number of points. The 'square root' of the covariance matrix (9.2) can be found using the Jacobi method. An alternative is to use an LR decomposition of the covariance matrix, which can be achieved using a Cholevski decomposition.

9.3 Turning Band Simulation

In turning band simulations sets of one dimensional simulations are merged to a two (or three) dimensional one. One dimensional simulations are performed for different possible directions "turning" around a center point. Depending on the variogram different covariance structures have to be used for the one dimensional simulations. The advantage of the method is that it is nearly independent from the

number of points. A disadvantage is that the one dimensional covariance structure corresponding to the variogram has to be calculated (or given analytically).

The general idea of turning band simulations is given on the next few pages, without full computational details.

9.3.1 Unconditional simulation

The basic idea of turning band simulations is to use a set of one dimensional simulations to construct the multidimensional one. Projecting a point in the 2 or 3 dimensional space onto these lines, and taking the sum of the values corresponding to the projected points, yields the simulated value.

Suppose that for a set of lines $l = 1,\ldots,L$, all going through the origin of the coordinate system, random functions with zero mean and $C_1(r)$ covariance functions are simulated independently. Let $Z_l(u)$ for $l = 1,\ldots,L$ be these functions. Then for a point u the random function $Z(u)$ can be defined as

$$Z(u) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} Z_l(\langle u, v_l \rangle) \tag{9.3}$$

here $\langle .,. \rangle$ denotes the scalar product of the vectors, and v_l is the unit vector on line l. Figure 9.1 explains the definition of $Z(u)$ in the two dimensional case.

Using the fact that $E[Z_i(r)] = 0$ the covariance function of the above defined random $Z(u)$ is:

$$C(u_1,u_2) = \frac{1}{L} \sum_{l=1}^{L} \sum_{k=1}^{L} Z_l(\langle u_1, v_l \rangle) Z_l(\langle u_2, v_k \rangle) \tag{9.4}$$

As Z_l and Z_k are independent if $l \neq k$ the above sum can also be written as:

$$\begin{aligned} C(u_1,u_2) &= \frac{1}{L} \sum_{l=1}^{L} Z_l(\langle u_1, v_l \rangle) Z_l(\langle u_2, v_l \rangle) \\ &= \frac{1}{L} \sum_{l=1}^{L} C_1(|\langle u_1, v_l \rangle - \langle u_2, v_l \rangle|) \\ &= \frac{1}{L} \sum_{l=1}^{L} C_1(|\langle u_1 - u_2, v_l \rangle|) \end{aligned} \tag{9.5}$$

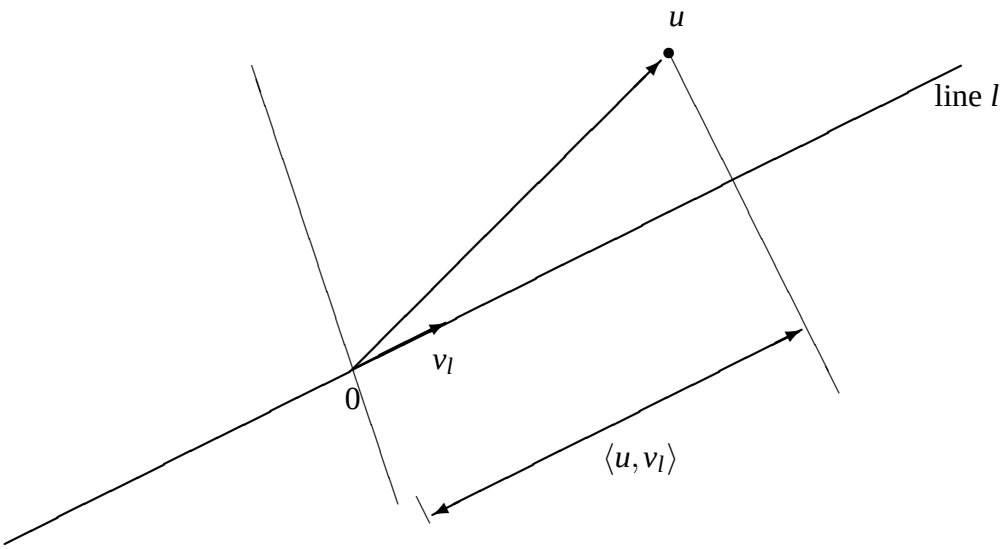


Figure 9.1: Turning bands lines and the projection

This equation shows that $Z(u)$ is also stationary. If the unit vectors v_l are uniformly distributed on unit sphere or on the unit circle, then taking the limit of the above expression as $L \to \infty$

$$C(h) = \frac{1}{S} \int_{|v|=1} C_1(|\langle h, v \rangle|) \tag{9.6}$$

here S is the surface of the unitsphere (4π) or the length of the perimeter of the unit circle (2π) depending on the dimension. As the covariance is isotropic the value of $C(r)$ for a given distance r can be calculated with the help of any vector h of length r. The two and the three dimensional cases have to be handled differently.

Two dimensional simulation

Selecting $h = (r, 0)$ and introducing the polar coordinates (9.6) can be rewritten in the form:

$$C(r) = \frac{1}{2\pi} \int_0^{2\pi} C_1(r\cos\varphi)\, d\varphi = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} C_1(r\cos\varphi)\, d\varphi \tag{9.7}$$

Substituting $\tau = r\cos\varphi$

$$C(r) = \frac{2}{\pi} \int_0^r \frac{C_1(\tau)}{\sqrt{r^2 - \tau^2}}\, d\tau \tag{9.8}$$

The integral equation (9.8) has to be solved for different possible covariance functions — corresponding to variograms with a sill.

Three dimensional simulation

Selecting $h = (r, 0, 0)$ and introducing the spherical coordinates (9.6) can be rewritten in the form:

$$C(r) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} C_1(r\cos\varphi)\, d\varphi\, d\theta \tag{9.9}$$

Introducing $\tau = r\cos\varphi$ after some calculations one finally has:

$$C(r) = \frac{1}{r} \int_0^r C_1(\tau)\, d\tau \tag{9.10}$$

Note that (9.10) is much simpler than its two dimensional equivalent (9.8). It can even be rewritten as

$$C_1(\tau) = \frac{d}{d\tau} [\tau C(\tau)] \tag{9.11}$$

This equation makes it possible to derive for each theoretical model the corresponding one dimensional covariance function. Then using one dimensional simulation methods (many of them can be found in the time series literature) Z_l functions are generated, and using (9.3) $Z(u)$ is constructed.

Simulation of complex models

The solution of the equations relating the three dimensional covariance function to the one dimensional is quite difficult. For complex models it is much simpler to use the solution for the particular models (spherical, exponential etc.) and to simulate random functions for these models. The simulation for the complex model can then be constructed with the help of (3.14) and (3.15).

The same idea can be used to simulate random functions with zonal anisotropy. Turning band simulations generate realisations of multidimensional normal distributions. The reason for this is that the random function is the sum of independent random variables, and by the central limit theorem these sums converge to the normal distribution. In order to obtain simulations for non normal distributions transformations have to be done. The transformation which transforms the required distribution to the normal has to be found. Then using this transformation variograms of the transformed variable have to be calculated. Simulation is then performed for the transformed variable, and finally the results are transformed back to the original scale. This procedure often helps but there is no guarantee. Sometimes even the conditioning transforms the distribution to the required one.

9.3.2 Conditional simulation

The measurement data are not really used in unconditional simulation. It is only through the variogram they influence simulation results. The knowledge of the value of a selected parameter at a given point restricts the possible values in a neighbourhood. Those realizations are specially interesting for which at the measurement points the simulated values equal the measurement values. Unconditional simulations are conditioned with the help of a simple transformation :

$$Z_C(u) = Z^*(u) + (Z_S(u) - Z_S^*(u)) \tag{9.12}$$

here
$Z_S(u)$ is the simulated value at point u

$Z_S^*(u)$ is the kriging estimator of Z_S based on the simulated values at the measurement points
$Z_C(u)$ is the conditionally simulated value at u
$Z^*(u)$ is the kriging estimator of Z based on the measurement data

Because of the exactness property of kriging, for measurement points u_i $Z^*(u_i) = Z(u_i)$ and $Z_S^*(u_i) = Z_S(u_i)$. Thus by definition $Z_C(u_i) = Z(u_i)$. This means that the above modification of the unconditional simulation reproduces the measurement values. It can also be shown, that (9.12) does not influence the variability, $Z_C(u)$ and $Z_S(u)$ have the same variogram.

9.4 Sequential Simulation

Another possibility to obtain simulated random fields is the sequential simulation. The basic idea of this method is, that conditional distributions of the observed variable can be assessed and used for the simulation of subsequent points.

9.5 Simulation using Markov Chains

Another possibility to simulate realizations of a random field is simulated annealing. The basic idea of simulated annealing is to generate realizations of a random field for which the distributional assumptions are not convenient to sample them directly. For this purpose a Markov-chain is defined which has the prescribed limit distribution. The theoretical basis for this is the Hastings algorithm (Hastings 1970).

9.5.1 The Hastings Algorithm

The essential part of this is the Hastings algorithm: A given probability distribution π_i on the finite Ω can be simulated using a Markov chain. For this purpose

take an arbitrary transition matrix

$$\mathbf{Q} = (q_{ij})$$

of an irreducible Markov chain. Define the matrix **P** as:

$$p_{ij} = \alpha_{ij} q_{ij} \tag{9.13}$$

and

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij} \tag{9.14}$$

Let α_{ij} be given by:

$$\alpha_{ij} = \frac{s_{ij}}{1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}} \tag{9.15}$$

where the s_{ij}-s have to be symmetric:

$$s_{ij} = s_{ji}$$

and

$$0 < s_{ij} < 1 + \frac{\pi_i q_{ij}}{\pi_j q_{ji}}$$

For this Markov chain one can see that the limiting distribution of it is π. In the Metropolis algorithm the **Q** matrix is chosen symmetric and:

$$\alpha_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} \tag{9.16}$$

Note that in all previous equations the probabilities π_i had to be known up to a constant multiplier as only the ratios $\frac{\pi_i}{\pi_j}$ were required.

The method can also be applied in the continuous case (see Chib and Greenberg 1995).

9.5.2 Simulated annealing

In simulated annealing the probability distribution π is defined with the help of an objective function O depending on the state i:

$$\pi_i = K \exp(-O(i)) \tag{9.17}$$

In simulated annealing in order to find the limiting distribution on the realizations corresponding to the minima of O the temperature T is introduced. This means the Hastings algorithm is performed for the limiting distribution:

$$\pi(T) = K(T)\exp(-\frac{O(i)}{T}) \tag{9.18}$$

If the temperature T is decreased to 0 then for each state i with no global minimum the probability $\pi_i(T) \to 0$. Constraints on the rate of decrease of the temperature T to ensure a convergence to the uniform distribution on the set of global minima can be found in Geman and Geman (1984).

The simulation is performed on a predefined grid.

The algorithm can be described as follows:

1. Assign every grid cell corresponding to a measurement location the correct (measured) value.

2. Assign the remaining cells randomly values from the overall distribution.

3. Select an "energy function" O, which measures the difference between the statistical properties of the actual image and the prescribed properties.

4. Select a starting temperature t, and a number of swaps N_S to be tried before decreasing the temperature. Select a rate in which the temperature decreases.

5. Select two cells at random which do not correspond to measurement locations. Swap the values and calculate the value of the energy function O_n and compare it with the value for the unchanged case O_o. If $O_n < O_o$ then keep the swap. Else calculate the probability of acceptance p_a:

$$p_a = \exp\left(-\frac{O_n - O_o}{t}\right) \tag{9.19}$$

   Accept the swap with probability p_a and undo it with probability $1 - p_a$.

6. Repeat the previous step N_S times.

7. Reduce the temperature slightly, according the rate of 4. Repeat steps 5-6.

8. Repeat the previous step until the energy function value O is close to zero.

The most important part of this procedure is the selection of the energy function O. Here both general assumptions - like the reproduction of a variogram, and other properties like local distributions can be taken into account. A possible form for a variogram dependent objective function is:

$$O_\gamma = \sum_{h_i} [\gamma_r(h_i) - \gamma_m(h_i)]^2 \tag{9.20}$$

where γ_r is the variogram value corresponding to the generated image and γ_m is the variogram to be achieved (model value). The function γ_m can in this case both be a theoretical variogram or even an experimental value. The advantage here is that difficult cases, such as zonal anisotropy can be simulated with this approach. Frequently the variogram values corresponding to small distances play a much more important role, thus the variogram objective function is formulated with relative deviations as:

$$O_\gamma = \sum_{h_i} \left[\frac{\gamma_r(h_i) - \gamma_m(h_i)}{\gamma_m(h_i)} \right]^2 \tag{9.21}$$

The objective function for the neighbouring points can be formulated either directly or from a training image.

$$O_\mathcal{N} = \sum_{x_i} \sum_{\mathcal{N}} |f_r - f_m| \tag{9.22}$$

The final energy function is then formulated as a linear combination of the variogram and the neighbourhood objective function:

$$O = \alpha_\gamma O_\gamma + \alpha_\mathcal{N} O_\mathcal{N} \tag{9.23}$$

with α_γ and $\alpha_\mathcal{N}$ being positive weights.

Further properties can also be formulated as objective functions. However it is very important that the objective function has a form which allows a fast update of the energy function in case of a swap.

The initial temperature should be selected in a way that quite a great number of non-improving swaps are performed at the beginning of the simulation. This way the independence of the final result from the initial picture is guaranteed. The number of swaps to be tried at a given temperature should be high enough, in general values around the number of points in the simulation grid are recomended.

The advantages of the simulated annealing algorithm include:

1. Arbitrary marginal distributions can be simulated.

2. The objective function can include different information, and can be formulated in a flexible way.

3. The space containing the realizations with the prescribed statistical properties is sampled uniformly.

4. The algorithm can easily be used for 2 and 3 dimensional realizations.

Besides these important advantages there are a few disadvantages of simulated annealing.

1. As an optimization method it does not allow deviations from the prescribed statistical properties. If these are estimated then their possible error hast to be incorporated.

2. The algorithm is slow. The numerical realization of the simulation requires fast computing possibilities and an effective formulation of the energy function changes.

9.6 Indicator Simulation

There are two main methods for simulating indicator variables. In one case a continuous variable is simulated and the truncated into discrete classes according to prescribed limits. In the other case indicators are simulated directly.

9.6.1 Truncated-Gaussian Simulation

The idea of the truncated gaussian approach is to simulate indicator values with the help of a gaussian variable. Suppose $Y(x)$ is a multigaussian random variable with 0 mean and unit variance. In this case:

$$C_i = \{x \quad ; \quad s_{i-1} < Y(x) \leq s_i\} \tag{9.24}$$

defines a random set in the space. The probability of an x belonging to C_i can be calculated as:

$$P(x \in C_i) = P(s_{i-1} < Y(x) \leq s_i) = \Phi(s_i) - \Phi(s_{i-1}) \tag{9.25}$$

The thresholds s_i have to satisfy:

$$s_0 = -\infty \quad s_n = +\infty$$

The indicator variograms and cross-variograms can be calculated form the covariance matrix of the multigaussian variable $Y(x)$

Conditional simulation of the indicators I_{C_i} requires a further step as the corresponding Y values are unknown.

9.7 Application of simulations

There are several possible applications for unconditional and conditional simulations. The classical mining applications were forecasting quality fluctuations for blending and processing.

It can also be successfully used in calibrating non linear complex systems (like groundwater movement). In this case as the output of the system is not linearly dependent on the regionalized variable usage of the kriged values does not necessarily lead to expectations of the output. Thus using a set of simulations can yield better estimates.

Another application is sensitivity analysis. Conditional simulations can help to calculate realistic sensitivities, as both the dependence and the measurement locations are considered.

9.7.1 Examples

Consider the problem of estimating areal rainfall for hydrological modelling and extreme value statistics. Areal rainfall cannot be measured directly, and thus has to be estimated on the basis of point values. The estimation can be done using OK or EDK or IRF-k. In general this is done by interpolating precipitation values on a dense regular grid, and then calculating the mean of the grid values corresponding to the selected area (subcatchment). It is obvious that the probability of measuring the maximum precipitation amounts at the raingauges is very low.

Chapter 10

Exercises

10.1 The Variogram

10.1.1 Question: 1

1. The matrix in figure 10.1 is considered to be a set of measurements (ex. relative soil humidity in percent) on a regular grid with vertical (y) and horizontal (x) distance of 1 between neighbouring points.

   Calculate the experimental variogram for horizontal (x,$\approx 0^\circ$), vertical (y,$\approx 90^\circ$) and for the diagonal (45°) direction for the shortest possible nonzero distance each.

10.1.2 Solution: 1

The general formula for the variogram - for a specific distance or vector h is:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (Z(u_i) - Z(u_j))^2$$

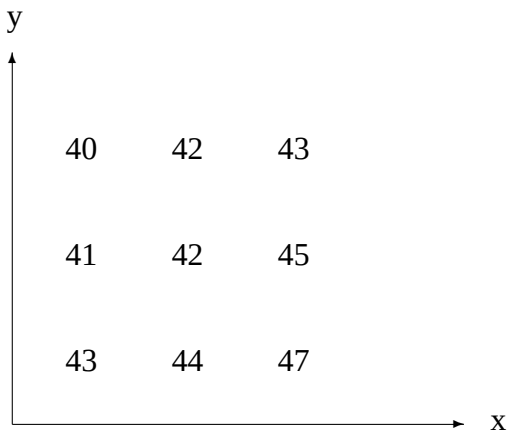


Figure 10.1: A set of measurements

where $N(h)$ is the number of pairs separated by the vector (or distance) h. Applying the formula for this exercise gives:

1. for horizontal direction, minimal nonzero distance is 1, 6 pairs of the 3x3 matrix:

$$\gamma(1_{hor}) = \frac{1}{2 \cdot 6}((40-42)^2 + (42-43)^2 + (41-42)^2 + (42-45)^2 + (43-44)^2 + (44-47)^2)$$

$$\gamma(1_{hor}) = \frac{1}{2 \cdot 6}(2^2 + 1^2 + 1^2 + 3^2 + 1^2 + 3^2)$$

$$\gamma(1_{hor}) = \frac{25}{12} = 2.08\overline{3}$$

2. for vertical direction, minimal nonzero distance is 1, 6 pairs of the 3x3 matrix:

$$\gamma(1_{ver}) = \frac{1}{2 \cdot 6}((40-41)^2 + (41-43)^2 + (42-42)^2 + (42-44)^2 + (43-45)^2 + (45-47)^2)$$

$$\gamma(1_{ver}) = \frac{1}{2 \cdot 6}(1^2 + 2^2 + 0^2 + 2^2 + 2^2 + 2^2)$$

$$\gamma(1_{ver}) = \frac{17}{12} = 1.41\overline{6}$$

3. for diagonal (45°) direction, minimal nonzero distance is 1.414 = $\sqrt{2}$, 4 pairs of the 3x3 matrix:

$$\gamma(1.414_{45}) = \frac{1}{2 \cdot 4}((41-42)^2 + (43-42)^2 + (42-43)^2 + (44-45)^2)$$

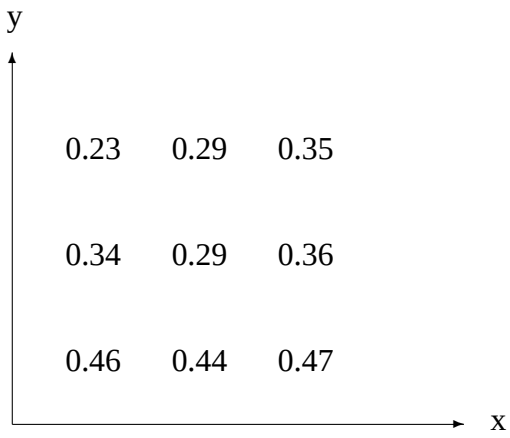
$$\gamma(1.414_{45}) = \frac{1}{2 \cdot 4}(1^2 + 1^2 + 1^2 + 1^2)$$

$$\gamma(1.414_{45}) = \frac{4}{8} = 0.5$$

10.1.3 Question: 2

The following matrix is considered to be a set of measurements (ex. relative soil humidity) on a regular grid with vertical (y) and horizontal (x) distance of 1 between neighbouring points:

Calculate the experimental indicator variograms for horizontal (x,$\approx 0°$), vertical (y,$\approx 90°$) direction for distance 1 each, using the indicator thresholds 0.3 and 0.4.

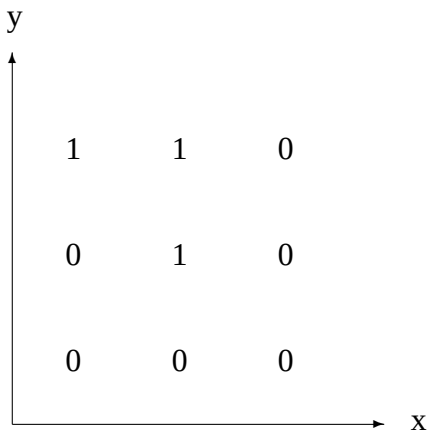


10.1.4 Solution: 2

The general formula for the variogram - for a specific distance or vector h is:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (Z(u_i) - Z(u_j))^2$$

where $N(h)$ is the number of pairs separated by the vector (or distance) h. The application of an indicator variable means to transform the matrix of measurements to 0 and 1 values (0, if above threshold; 1, if below or equal). Applying this for the threshold 0.3 gives:



Applying the formula of the variogram on this gives:

Horizontal:

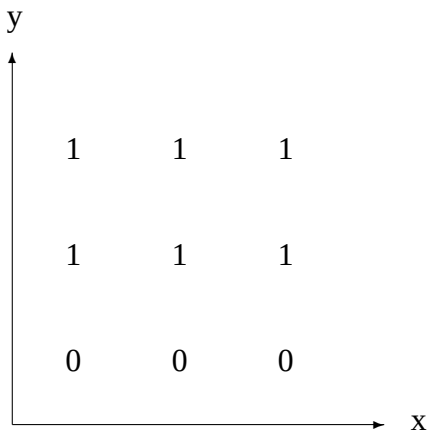
$$\gamma^*(1_{hor}) = \frac{1}{2 \cdot 6}((1-1)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2)$$

$$\gamma^*(1_{hor}) = \frac{1}{12}(0+1+1+1+0+0) = \frac{3}{12} = 0.25$$

Vertical:

$$\gamma^*(1_{ver}) = \frac{1}{12}(1+0+0+1+0+0) = \frac{2}{12} = 0.1667$$

Applying this for the threshold 0.4 gives:



Horizontal:

$$\gamma^*(1_{hor}) = \frac{1}{12}(0+0+0+0+0+0) = \frac{0}{12} = 0.0$$

Vertical:

$$\gamma^*(1_{ver}) = \frac{1}{12}(0+1+0+1+0+1) = \frac{3}{12} = 0.25$$

10.1.5 Question: 3

The following figure 10.2 presents a set of measurements of a parameter (ex. nitrate in mg/kg in the soil of a small field) on a regular grid of 1 m cell size. Fields marked with "*x*" are points without measurements:

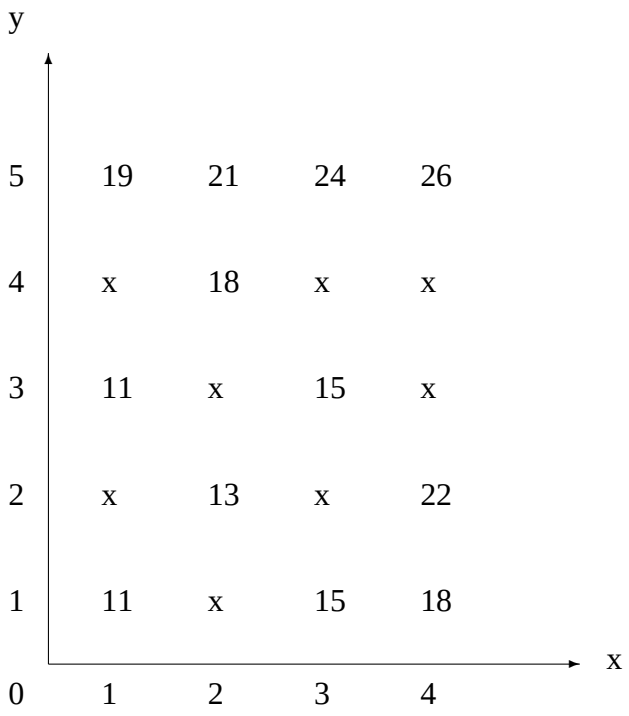


Figure 10.2: A set of measurements

Calculate the experimental variogram distances 1m and 2m in *x* and *y* direction.

10.1.6 Solution: 3

The general formula for the variogram - for a specific distance or vector h is:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (Z(u_i) - Z(u_j))^2$$

where $N(h)$ is the number of pairs separated by the vector (or distance) h. Applying the formula gives:

1. for the horizontal direction (x) and distance 1m, 4 pairs of the matrix:

$$\gamma(1_{hor}) = \frac{1}{2 \cdot 4}((19-21)^2 + (21-24)^2 + (24-26)^2 + (15-18)^2)$$

$$\gamma(1_{hor}) = \frac{1}{2 \cdot 4}(2^2 + 3^2 + 2^2 + 3^2)$$

$$\gamma(1_{hor}) = \frac{26}{8} = 3.25$$

2. for the vertical direction (y), distance 1m, 2 pairs of the matrix:

$$\gamma(1_{ver}) = \frac{1}{2 \cdot 2}((21-18)^2 + (22-18)^2)$$

$$\gamma(1_{ver}) = \frac{1}{2 \cdot 2}(3^2 + 4^2)$$

$$\gamma(1_{ver}) = \frac{25}{4} = 6.25$$

3. for the horizontal direction (x) and distance 2 m, 5 pairs of the matrix:

$$\gamma(2_{hor}) = \frac{1}{2 \cdot 5}((19-24)^2 + (21-26)^2 + (11-15)^2 + (13-22)^2 + (11-15)^2)$$

$$\gamma(2_{hor}) = \frac{1}{2 \cdot 5}(5^2 + 5^2 + 4^2 + 9^2 + 4^2)$$

$$\gamma(2_{hor}) = \frac{163}{10} = 16.3$$

4. for vertical direction (y), distance 2m, 5 pairs of the matrix:

$$\gamma(2_{ver}) = \frac{1}{2 \cdot 5}((19-11)^2 + (11-11)^2 + (18-13)^2 + (24-15)^2 + (15-15)^2)$$

$$\gamma(1_{ver}) = \frac{1}{2 \cdot 5}(8^2 + 0^2 + 5^2 + 9^2 + 0^2)$$

$$\gamma(2_{ver}) = \frac{170}{10} = 17.0$$

10.1.7 Question: 4

Measurement of a certain physical quantity which varies spatially was carried out at different locations. The measured values of the quantity and their spatial locations are shown in figure 10.3.

Calculate the lag 1 indicator variogram for the quantity for a cut-off value of 31

1. in the x direction
2. in the y direction
3. without considering the direction.

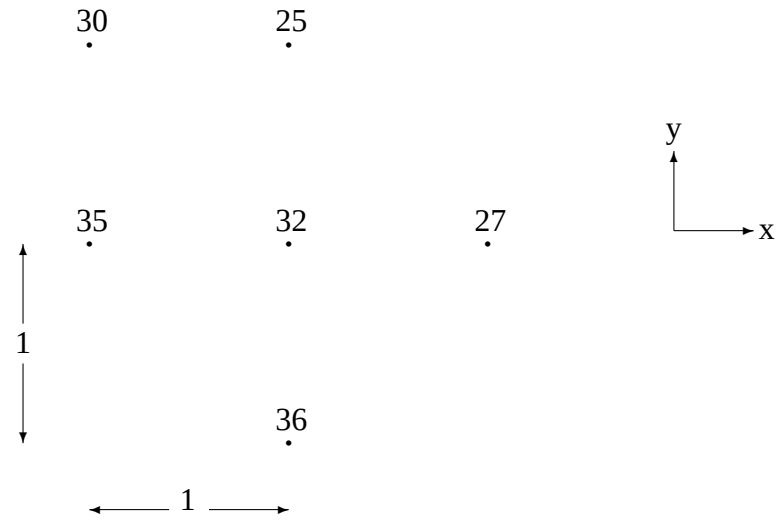


Figure 10.3: Spatial locations of measured values of the quantity

10.1.8 Solution: 4

$$I_\alpha(u) = 1 \text{ if } z(u) \leq 31$$
$$I_\alpha(u) = 0 \text{ if } z(u) > 31$$

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (I_\alpha(u_i) - I_\alpha(u_j))^2$$

1. The variogram $h = 1$ and $h = 2$ in the x direction:

$$\gamma_x^*(1) = \frac{1}{2 \cdot 3}(0 + 0 + 1^2) = \frac{1}{6}$$

$$\gamma_x^*(2) = \frac{1}{2}(1^2) = \frac{1}{2}$$

2. in the y direction

$$\gamma^*(1) = \frac{1}{2 \cdot 3}(1^2 + 1^2 + 0) = \frac{1}{3}$$

$$\gamma_y^*(2) = \frac{1}{2}(1) = \frac{1}{2}$$

3. without taking the direction into account:

$$\gamma^*(1) = \frac{1}{2 \cdot 6}(0 + 0 + 1^2 + 1^2 + 0 + 1^2) = \frac{1}{4}$$

$$\gamma^*(2) = \frac{1}{2 \cdot 2}(1^2 + 1^2) = \frac{1}{2}$$

10.2 Ordinary Kriging

10.2.1 Question: 1

The sketch in figure 10.4 shows 4 measurements $Z(u)$ for points $u = -2, -1, 1$ and 2, ordered in a straight line. Write down the equations of an ordinary kriging system for an estimation Z^* at the point $u = 0$. For the variogram take the function $\gamma(h) = h$ for a distance h. (Hint: it is not necessary to solve the equations!)

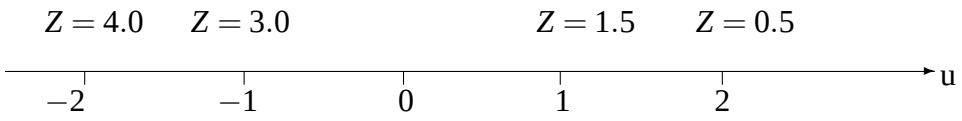


Figure 10.4: A sketch showing 4 measurements $Z(u)$ for points $u = -2, -1, 1,$ and 2 ordered in a straight line.

10.2.2 Solution: 1

The kriging - system for ordinary kriging is

$$\sum_{j=1}^{n} \lambda_j \cdot \gamma(u_i - u_j) + \mu = \gamma(u_i - u), \quad i = 1, \ldots, n$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

with kriging weights λ_i for the measurement at u_i, $i = 1, ..., n$. In this case, with $u_1 = -2$, $u_2 = -1$, $u_3 = 1$, $u_4 = 2$ and the estimation of $u = 0$, the symmetry of the locations leads to the symmetry of the weights so that:

$$\lambda_1 = \lambda_4, \quad \lambda_2 = \lambda_3$$

Therefore the equation system is reduced to 3 equations ($i = 1, 2$), because the equations for $i = 1, 4$ and $i = 2, 3$ are identical:

$$\text{i = 1: } \lambda_1 \cdot \gamma(0) + \lambda_2 \cdot \gamma(1) + \lambda_2 \cdot \gamma(3) + \lambda_1 \cdot \gamma(4) + \mu = \gamma(2)$$

$$\text{i = 2: } \lambda_1 \cdot \gamma(1) + \lambda_2 \cdot \gamma(0) + \lambda_2 \cdot \gamma(2) + \lambda_1 \cdot \gamma(3) + \mu = \gamma(1)$$

$$2 \cdot \lambda_1 + 2 \cdot \lambda_2 = 1$$

And by using the variogram function:

$$\text{i =1: } \lambda_1 \cdot 0 + \lambda_2 \cdot 1 + \lambda_2 \cdot 3 + \lambda_1 \cdot 4 + \mu = 2$$

$$\text{i = 2: } \lambda_1 \cdot 1 + \lambda_2 \cdot 0 + \lambda_2 \cdot 2 + \lambda_1 \cdot 3 + \mu = 1$$

$$2 \cdot \lambda_1 + 2 \cdot \lambda_2 = 1$$

$$\text{i = 1: } \lambda_1 \cdot 4 + \lambda_2 \cdot 4 + \mu = 2$$

$$\text{i = 2: } \lambda_1 \cdot 4 + \lambda_2 \cdot 2 + \mu = 1$$

$$2 \cdot \lambda_1 + 2 \cdot \lambda_2 = 1$$

The kriging system without using the symmetry is:

$$\text{i = 1: } \lambda_1 \cdot 0 + \lambda_2 \cdot 1 + \lambda_3 \cdot 3 + \lambda_4 \cdot 4 + \mu = 2$$

$$\text{i = 2: } \lambda_1 \cdot 1 + \lambda_2 \cdot 0 + \lambda_3 \cdot 2 + \lambda_4 \cdot 3 + \mu = 1$$

$$\text{i = 3: } \lambda_1 \cdot 3 + \lambda_2 \cdot 2 + \lambda_3 \cdot 0 + \lambda_4 \cdot 1 + \mu = 1$$

$$\text{i = 4: } \lambda_1 \cdot 4 + \lambda_2 \cdot 3 + \lambda_3 \cdot 1 + \lambda_4 \cdot 0 + \mu = 2$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$$

The last equation leads to $\lambda_2 = 0.5 - \lambda_1$, adding this to the previous equations leads to:

$$\text{i = 1: } \lambda_1 \cdot 4 + (0.5 - \lambda_1) \cdot 4 + \mu = 2 + \mu = 2$$

$$\text{i = 2: } \lambda_1 \cdot 4 + (0.5 - \lambda_1) \cdot 2 + \mu = \lambda_1 \cdot 2 + 1 + \mu = 1$$

$$\mu = 0 \text{ , } \lambda_1 = 0 \text{ , } \lambda_2 = 0.5$$

The kriging weights are 0.5 for u_2 and u_3 and 0 for u_1 and u_4. The estimation therefore is $Z(u)$=2.25.

10.2.3 Question: 2

The figure 10.5 presents a set of measurements of a parameter (ex. nitrate in mg/kg in the soil of a small field) on a regular grid of 1m cell size. Fields marked with "x" are points without measurements:
Use the data given in the figure for the following calculations:

1. Write down the equations of an ordinary kriging system for the point (2,3) using only the directly neighbouring points. For the variogram function take $\gamma(h) = 4h$. (Hint: it is not necessary to solve the equations!)

2. When using simple kriging instead of ordinary kriging, what kind of additional information do you need?

10.2.4 Solution: 2

1. The kriging - system for ordinary kriging is

$$\sum_{j=1}^{n} \lambda_j \cdot \gamma(u_i - u_j) + \mu = \gamma(u_i - u), \quad i = 1, \ldots, n$$

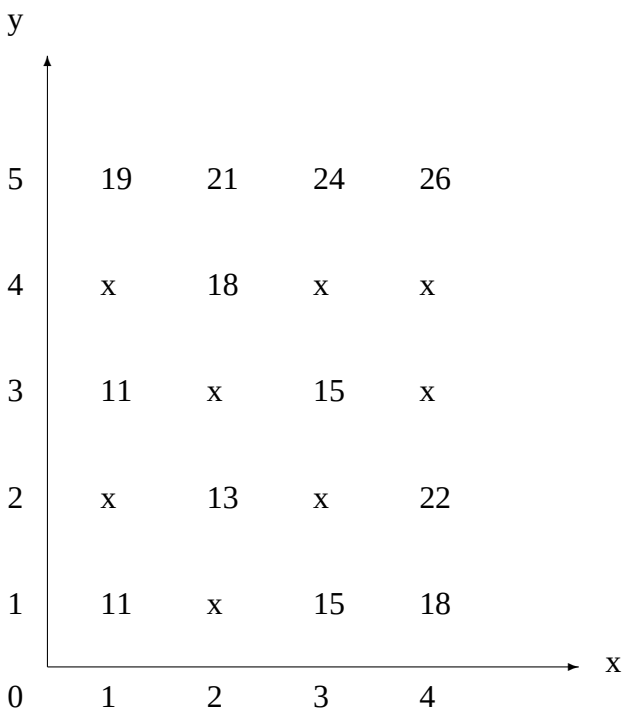


Figure 10.5: A set of measurements

$$\sum_{j=1}^{n} \lambda_j = 1$$

with kriging weights λ_i for the measurement at u_i, $i = 1,..,n$. In this case the neighbouring points of $(2,3) = u$ are: $(1,3) = u_1$,$(2,2) = u_2$, $(2,4) = u_3$, $(3,3) = u_4$

The distances in this configuration are:
1m (x and y direction) with $\gamma(1) = 4$,
2m (x and y direction) with $\gamma(2) = 8$,
1.41m (diagonal, ex. (2,2) to (3,3)) with $\gamma(1.41) = 5.64$

The kriging system without using the symmetry is:

$$\text{i = 1: } \lambda_1 \cdot 0 + \lambda_2 \cdot \gamma(1.41) + \lambda_3 \cdot \gamma(1.41) + \lambda_4 \cdot \gamma(2) + \mu = \gamma(1)$$

$$\text{i = 2: } \lambda_1 \cdot \gamma(1.41) + \lambda_2 \cdot 0 + \lambda_3 \cdot \gamma(2) + \lambda_4 \cdot \gamma(1.41) + \mu = \gamma(1)$$

$$\text{i = 3: } \lambda_1 \cdot \gamma(1.41) + \lambda_2 \cdot \gamma(2) + \lambda_3 \cdot 0 + \lambda_4 \cdot \gamma(1.41) + \mu = \gamma(1)$$

$$\text{i = 4: } \lambda_1 \cdot \gamma(2) + \lambda_2 \cdot \gamma(1.41) + \lambda_3 \cdot \gamma(1.41) + \lambda_4 \cdot 0 + \mu = \gamma(1)$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$$

and by evaluating the variogram $\gamma(h)$:

$$\text{i = 1: } \lambda_1 \cdot 0 + \lambda_2 \cdot 5.64 + \lambda_3 \cdot 5.64 + \lambda_4 \cdot 8 + \mu = 1 \quad (1)$$

$$\text{i = 2: } \lambda_1 \cdot 5.64 + \lambda_2 \cdot 0 + \lambda_3 \cdot 8 + \lambda_4 \cdot 5.64 + \mu = 1 \quad (2)$$

$$\text{i = 3: } \lambda_1 \cdot 5.64 + \lambda_2 \cdot 8 + \lambda_3 \cdot 0 + \lambda_4 \cdot 5.64 + \mu = 1 \quad (3)$$

$$\text{i = 4: } \lambda_1 \cdot 8 + \lambda_2 \cdot 5.64 + \lambda_3 \cdot 5.64 + \lambda_4 \cdot 0 + \mu = 1 \quad (4)$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \quad = 1 (5)$$

Calculating the differences between equation (2)-(3) and (1)-(4) leads to:
(2)-(3) : $\lambda_2 = \lambda_3$
(1)-(4): $\lambda_1 = \lambda_4$
evaluating these results in (1) and (2) and calculating the difference:
(1)-(2): $\lambda_2 = \lambda_1$

with (5) : $\lambda_4 = \lambda_3 = \lambda_2 = \lambda_1 = 0.25$

The kriging weights are 0.25 for all points. The estimation is therefore $Z(u) = 14.25$.

The fast way to solution (but not to the kriging equations): By looking at the symmetry of the points, it can be concluded that the system is invariant for rotations of 90°. Because the variogram contains no anisotropy, the solution must also be invariant for rotation. Therefore the solution is: equal weights for all points!

2. For the application of simple kriging, the mean value of underlying process $m(u)$ at location u hast to be known. If $m(u)$ is assumed as constant over the domain, $m(u)$ can be calculated as mean of the data.

10.2.5 Question: 3

The table 10.1 displays measurements of groundwater parameters $Z(u)$ for points u ordered in a straight line.

(a) Write down the equations of an external drift kriging system for an estimation of Z^* of chloride at the point $u = 2$ assuming that chloride is linearly related to the electric conductivity. For the variogram take the function $\gamma(h) = \frac{3}{2}h$ for a distance h. (hint: it is not necessary to solve the equations!)

(b) What other methods can be used for the solution, if you consider the dependency of chloride on electric conductivity? Please mention the steps do you have to follow for these procedures?

Location (u)	Chloride[mg/l]	Electric Conductivity[mS/m]
0	19.4	20
1	22.2	22
2	–	21
3	23.6	24
4	24.8	25

Table 10.1: Measurements of groundwater parameters $Z(u)$ for points u ordered in a straight line

10.2.6 Solution: 3

(a) The equations of external drift kriging are as follows:

$$\sum_{j=1}^{I} \lambda_j \gamma(u_i - u_j) + \mu_1 + \mu_2 Y(u_i) = \gamma(u_i - u) \qquad i = 1, \ldots, I$$

$$\sum_{j=1}^{I} \lambda_j = 1$$

$$\sum_{j=1}^{I} \lambda_j Y(u_j) = Y(u)$$

Applying these formulas for this specific case leads to 6 equations, the locations are: u_1=0, u_2=1, u_3=3, u_4=4 and u=2:

i = 1: $\lambda_1\gamma(0) + \lambda_2\gamma(1) + \lambda_3\gamma(3) + \lambda_4\gamma(4) + \mu_1 + \mu_2 Y(0) = \gamma(2)$

$$\text{i = 2: } \lambda_1\gamma(1)+\lambda_2\gamma(0)+\lambda_3\gamma(2)+\lambda_4\gamma(3)+\mu_1+\mu_2 Y(1)=\gamma(1)$$

$$\text{i = 3: } \lambda_1\gamma(3)+\lambda_2\gamma(2)+\lambda_3\gamma(0)+\lambda_4\gamma(1)+\mu_1+\mu_2 Y(3)=\gamma(1)$$

$$\text{i = 4: } \lambda_1\gamma(4)+\lambda_2\gamma(3)+\lambda_3\gamma(1)+\lambda_4\gamma(0)+\mu_1+\mu_2 Y(4)=\gamma(2)$$

$$\lambda_1+\lambda_2+\lambda_3+\lambda_4=1$$

$$\lambda_1 Y(0)+\lambda_2 Y(1)+\lambda_3 Y(3)+\lambda_4 Y(4)=Y(2)$$

Inserting the variogram $\gamma(h)=\frac{3}{2}h$, and the values of the function Y (electric conductivity) gives:

$$\text{i = 1: } 1.5\cdot\lambda_2+4.5\cdot\lambda_3+6\cdot\lambda_4+\mu_1+20\cdot\mu_2=3$$

$$\text{i = 2: } 1.5\cdot\lambda_1+3\cdot\lambda_3+4.5\cdot\lambda_4+\mu_1+22\cdot\mu_2=1.5$$

$$\text{i = 3: } 4.5\cdot\lambda_1+3\cdot\lambda_2+1.5\cdot\lambda_4+\mu_1+24\cdot\mu_2=1.5$$

$$\text{i = 4: } 6\cdot\lambda_1+4.5\cdot\lambda_2+1.5\cdot\lambda_3+\mu_1+25\cdot\mu_2=3$$

$$\lambda_1+\lambda_2+\lambda_3+\lambda_4 \quad = \quad 1$$

$$20\cdot\lambda_1+22\cdot\lambda_2+24\cdot\lambda_3+25\cdot\lambda_4=21$$

(b) Other methods to use for this problem:

Cokriging:

When applying cokriging not only variograms of both parameters have to be estimated, but also covariograms defining the relation between both parameters. The advantage of cokriging is that it is not necessary to have a measurement of the second parameter at the location of estimation.

Universal Kriging:

Universal kriging includes an explicit calculation of the external drift (which is not necessarily a linear function!). Usually the variogram has to be calculated iteratively using the residuals of the external drift calculated before.

Instrinsic Random Functions of Order k:

For IRF-k a generalized covariance has to be calculated. Assuming a linear trend IRF-1 (first order) is a suitable choice.

10.3 Short Questions

Please answer the following questions:

10.3.1 Question: 1

(a) Explain the role of the block-size on the estimation variance, when utilising block kriging.

(b) Two different measurement methods are used for a certain parameter of groundwater quality making more than one measurement at

each point. Method A delivers locally constant values at all locations, whereas Method B has an internal error producing local variances in the measured data at each point. Local mean values of Method B are the same of Method A. How do the variograms calculated for Method A and Method B differ from each other?

10.3.2 Solution: 1

(a) The estimation variance of block kriging is given by the formula:

$$\sigma^2(V) = -\bar{\gamma}(V,V) - \sum_{j=1}^{n}\sum_{i=1}^{n}\lambda_j\lambda_i\gamma(u_i - u_j) + 2\sum_{i=1}^{n}\lambda_i\bar{\gamma}(u_i,V)$$

The last term is the internal variance of the block. By increasing the size of the block, the internal variance increases as well, and this decreases the estimation variance. Therefore the bigger the size of the blocks, the smaller the estimation variance will be.

(b) The variogram of Method B is the one of Method A with an added term c_s corresponding to the internal error variance s (depending on the formulas used for the estimation of variance and variogram, these terms are not necessarily equal.) In case of Variogram A, there is no nugget effect, the Variogram B will have a nugget effect of c_s.

10.3.3 Question: 2

(a) Consider Simple Updating Kriging (SUK) and External Drift Kriging (EDK) for interpolation. What kind of data do you need to apply these methods?

(b) Based on the measured point-data, an interpolation and a simulation are done to plot a map for each. What do you expect to observe, when

you consider the variograms calculated with these maps (interpolation vs. simulation)? What kind of relations can you distinguish between the variogram calculated and the variogram obtained from the original dataset?

(c) When do you apply indicator kriging? Please define the cases where indicator kriging can be applied and such where indicator kriging must be applied!

10.3.4 Solution: 2

(a) Simple Updating Kriging (SUK) and External Drift Kriging (EDK) both need additional information which has to be known at the points of data and for the points of interpolation. For SUK, this information has to be categorical (ex. discrete class-number), for EDK a linearly related second variable is needed.

(b) The interpolation-variogram should have smaller values than the simulation-variogram. The simulation-variogram should be identical to the variogram of the original dataset.

(c) Categorical variables: Indicator Kriging (IK) must be applied.
Highly skewed distribution of parameter: Applying IK to different thresholds can deliver more plausible results than the conventional kriging methods.
Detection limit problems: If a high number of measurements is below the detection limit, IK should be applied. If the percentage is significantly high (appr. 10% or higher) only IK can deliver plausible results.
Below 10%: A substitution of measurements below detection limit is

possible.

Below 5%: IK is not necessary.

Chapter 11

References

Journel, A. G. (1983): Non parametric estimation of spatial distributions. Mathematical Geology, 15, 445-468.

Journel, A. G. and Zhu, H. (1990): Integrating Soft Seismic Data: Markov-Bayes updating, an alternative to cokriging and traditional regression. In Stanford Center for Reservoir Forecasting, Report 3, Stanford, 1-62.

Lehmann, W. (1995): Anwendung geostatistischer Verfahren auf die Bodenfeuchte in ländlichen Einzugsgebieten. Mitteilungen des Instituts für Hydrologie und Wasserwirtschaft, Nr.52, Universität Karlsruhe.

Matheron, G. (1971): The Theory of Regionalized Variables and itsApplications. Les Cahiers du Centre de Morphologie Mathématique, Fasc. 5.