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.

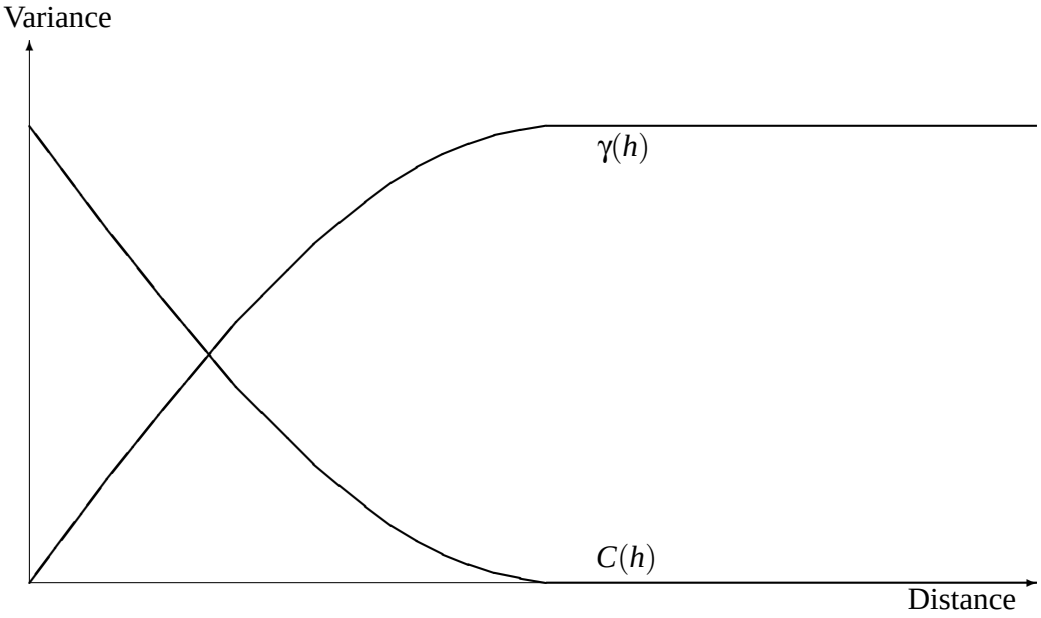

Figure 2.1: The covariance function $C(h)$ and the variogram $\gamma(h)$

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 ?"

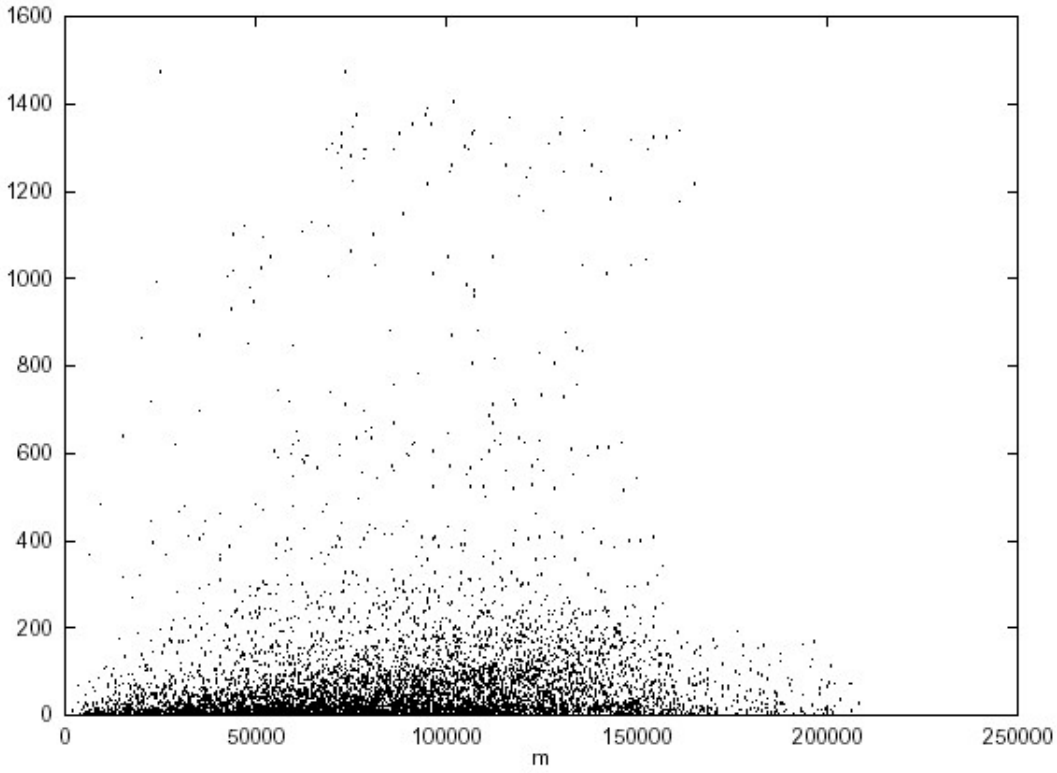

Figure 3.1: Variogram cloud $[mm^2]$ (precipitation Jan.3, 1990)

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.

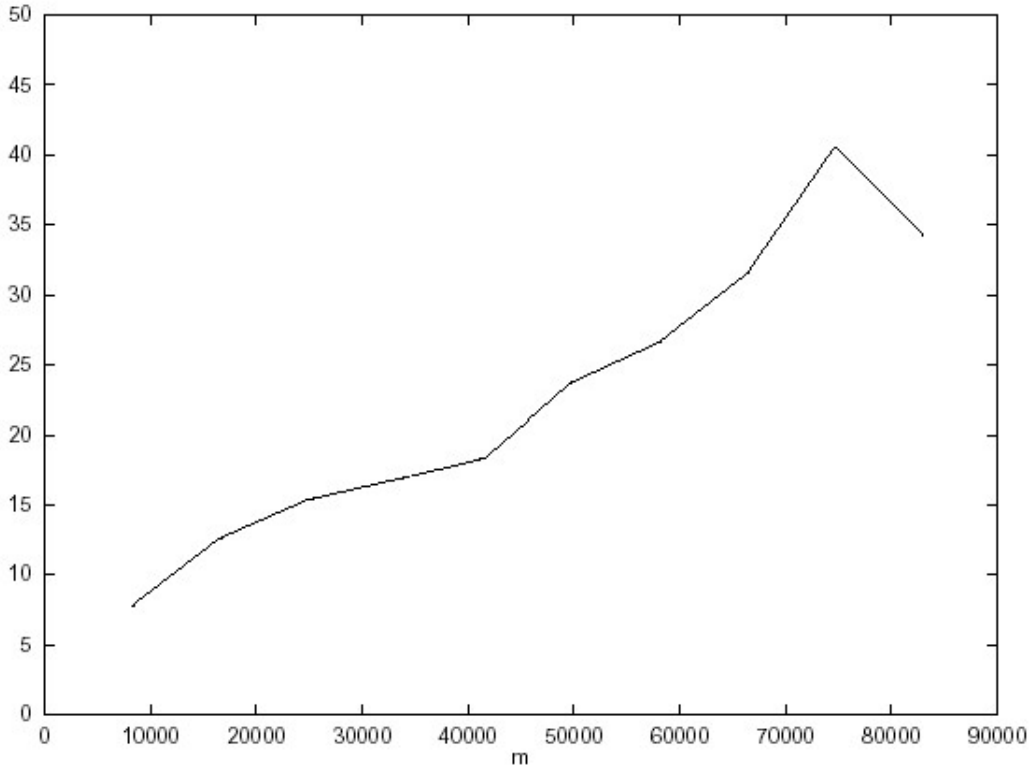

Figure 3.2: Experimental variogram $[mm^2]$ (precipitation Jan.3,1982)

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

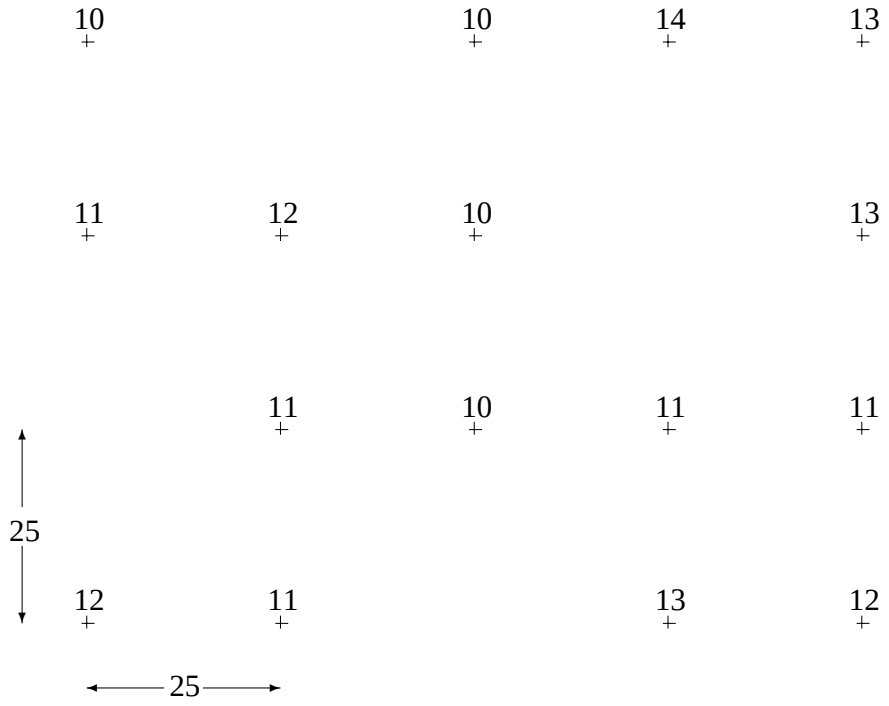

Figure 3.3: Data configuration and values for example 3.2

$\gamma^*(25_x) = 0.643$. 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

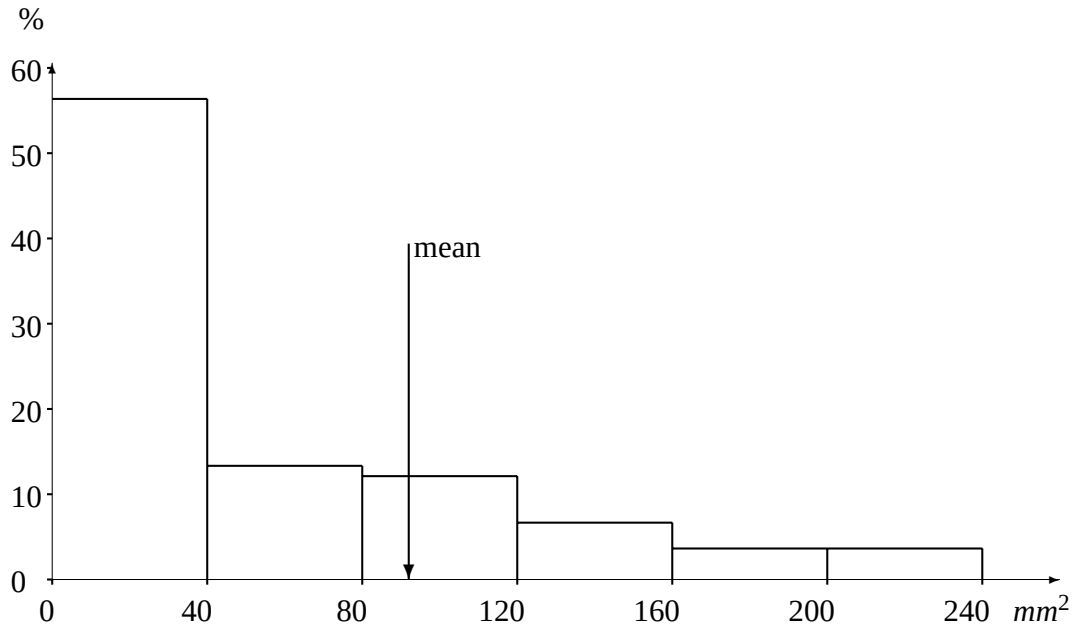

Figure 3.4: Histogram of squared differences corresponding to distance class 32 km

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.

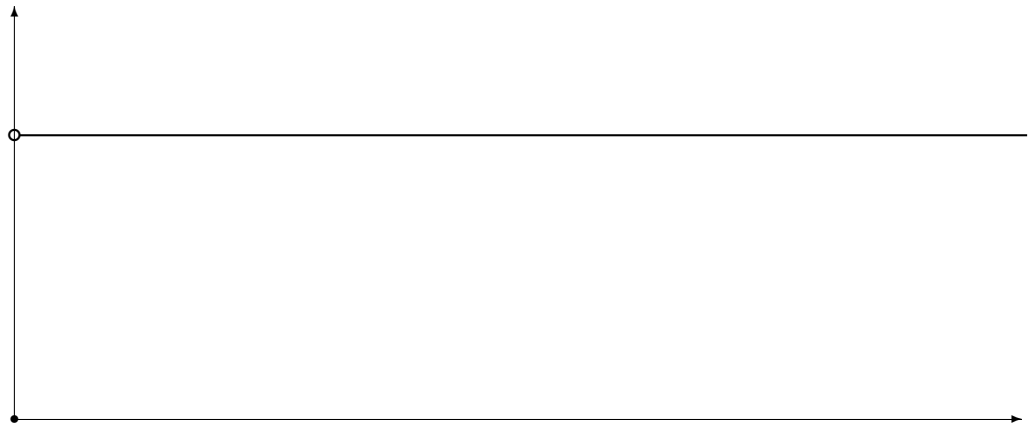

Figure 3.5: The pure nugget effect

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

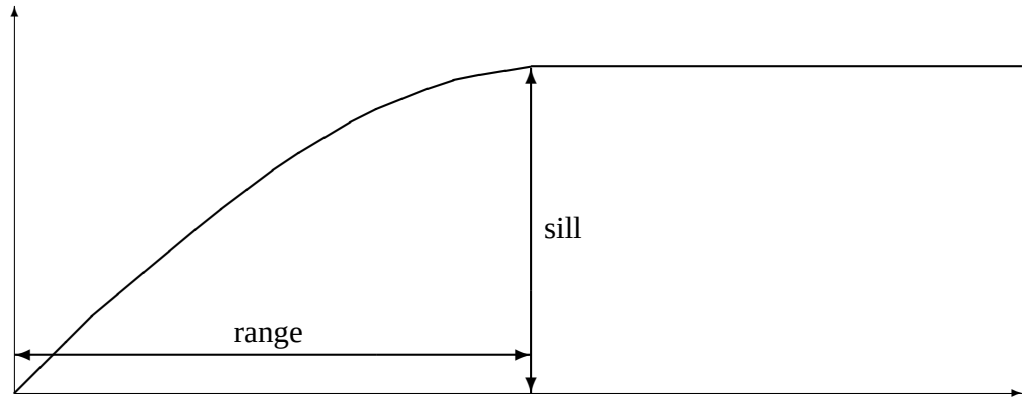

Figure 3.6: The spherical variogram

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:

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

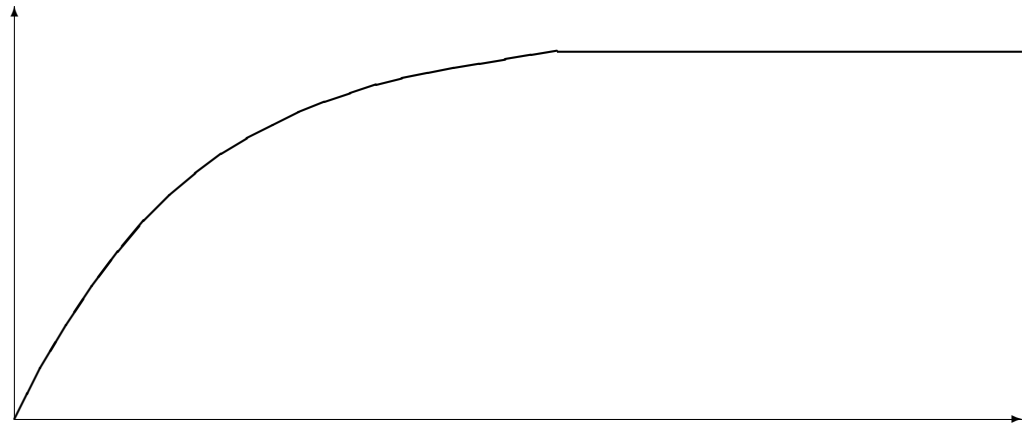

Figure 3.7: The exponential variogram

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.

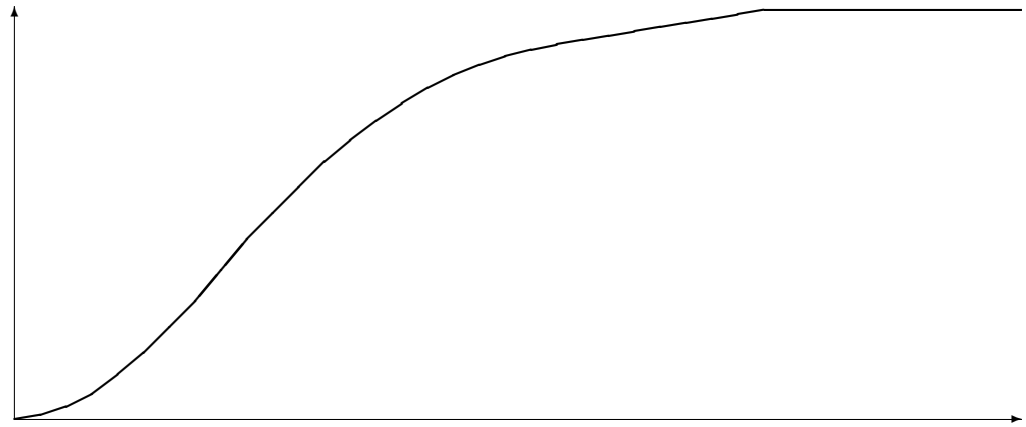

Figure 3.8: The gaussian variogram

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-

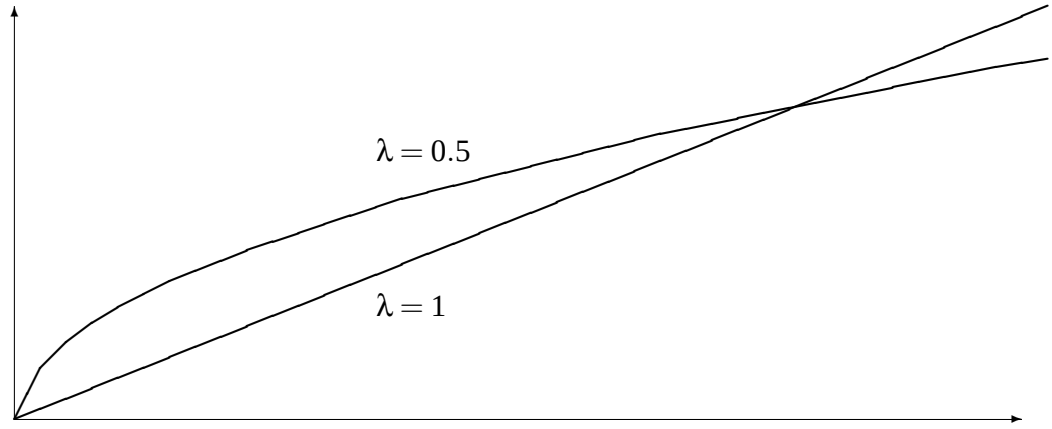

Figure 3.9: The h^{λ} variograms

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

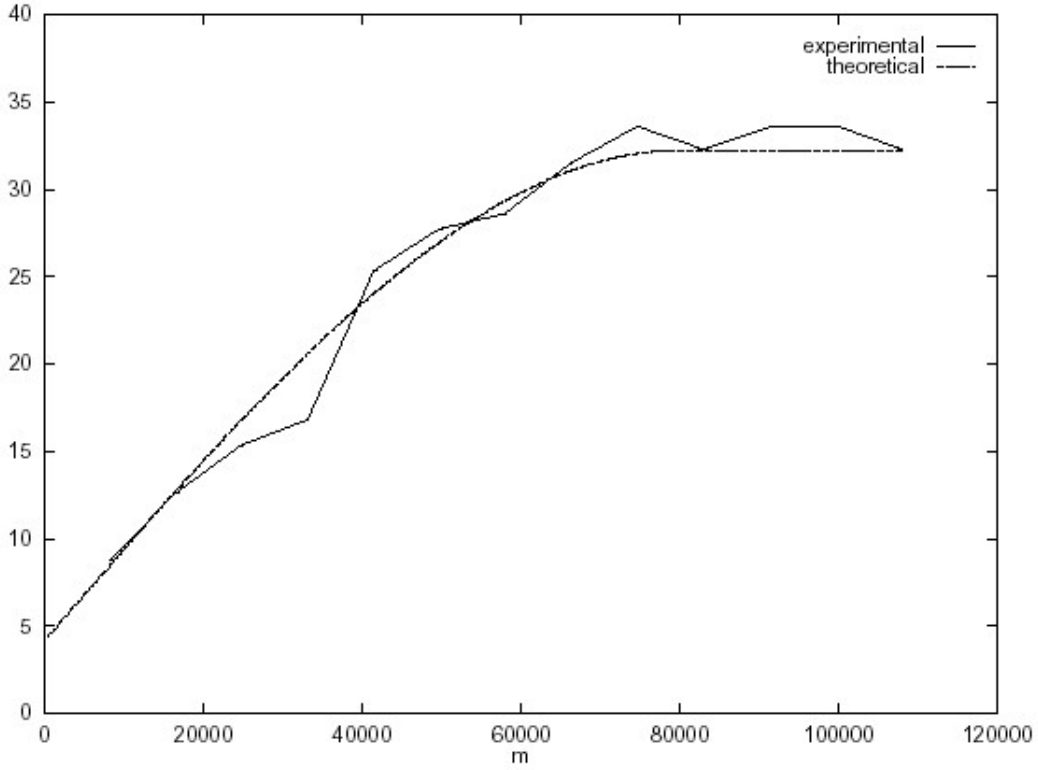

Figure 3.10: Experimental variogram with an easy fit

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.

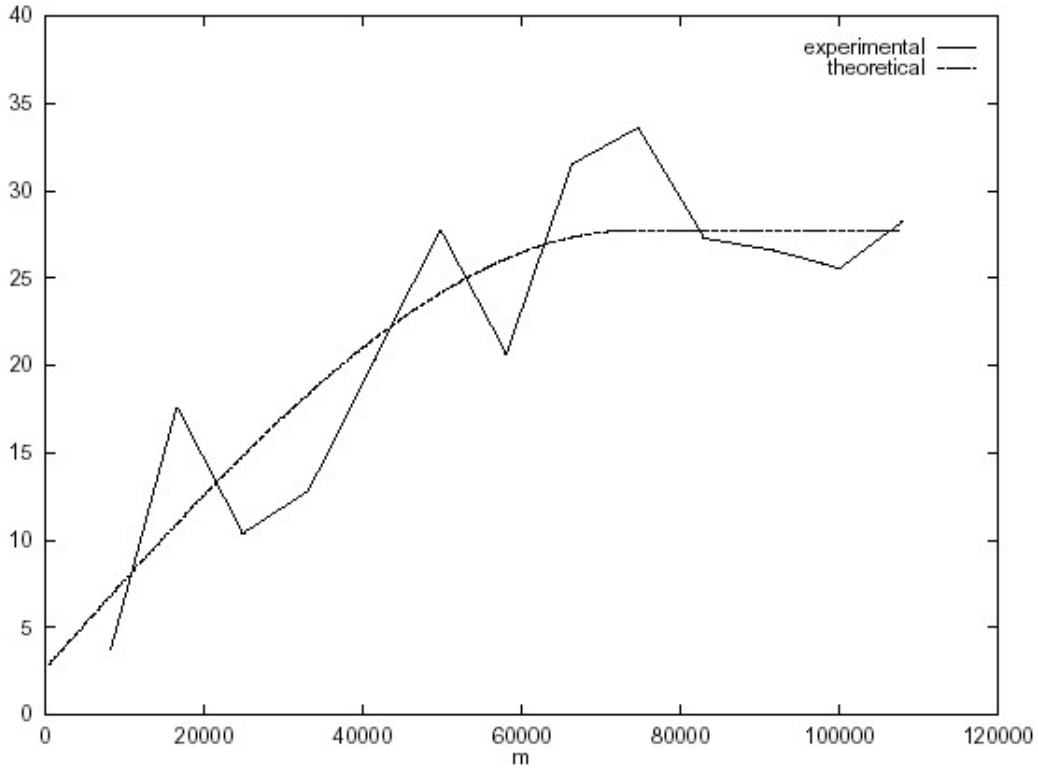

Figure 3.11: Experimental variogram with a difficult fit

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.