9

Variograms

9.1 Introduction to Methods for Variograms

In model-based methods of spatial prediction (kriging) the values of the target variable at the sampling locations are considered to be realizations of a Stochastic Function (see Appendix B). In a spatial context a Stochastic Function (SF) is a field of spatially dependent random variables (RVs), and therefore is also referred to as a Random Field (RF). In many cases it is assumed that the mean difference of the RVs at two locations is zero (constant mean), and the variance of this difference depends on the spatial separation vector (lag) $\mathbf{h}$ only, not on the locations themselves**:**

$$E\{Z(\mathbf{s}_2) - Z(\mathbf{s}_1)\} = 0 \tag{9.1}$$

$$E\{Z(\mathbf{s}_2) - Z(\mathbf{s}_1)\}^2 = 2\gamma(\mathbf{s}_2 - \mathbf{s}_1) = 2\gamma(\mathbf{h}) \; . \tag{9.2}$$

A SF that meets these requirements is an intrinsic SF. The function $\gamma(\mathbf{h})$ is referred to as the (semi-)variogram. If this variogram is known, then one can obtain Best Linear Unbiased Predictions (BLUP) of the values at points or the means of blocks from sample data. One may also use the variogram for geostatistical simulation of fields (realizations) that are used, for instance, as input in a process-based simulation model. In this chapter we present sampling designs appropriate for estimating the variogram. In many situations the collected sample data are used both for estimating the variogram and for geostatistical interpolation or simulation. In Sect. 8.3 appropriate designs for sampling in one-phase are described. If the available time and budget allow for sampling in two phases, then we recommend to focus the first phase sample on estimating the variogram, and the second-phase sample on interpolation. In general, this two-phase sampling is more efficient, because the variogram estimated from the first phase sample can be used to optimize the sample for the second phase. This chapter describes principles for designing efficient samples for the first phase of such a phased sampling approach.

The first choice to be made is the size of the first phase sample. According to Webster and Oliver (1992), 150 locations might suffice in many situations,

and 225 locations would be almost certainly adequate in most applications where spatial variation is equal in all directions (isotropic). These sample sizes already make clear that for model-based prediction quite a few locations must be sampled. This is usually only realistic when many local means or the values at points must be predicted. Once one has decided on the number of locations to estimate the variogram from, one must choose the locations themselves. We distinguish two approaches for selecting the locations. In the first approach clusters of locations are selected. The locations within a cluster have a more or less regular, predetermined pattern. In this approach, one must choose a pattern, and a method or principle for selecting the clusters. In the second approach the sampling locations are optimized by minimizing a quality measure. In general this approach will result in an irregular pattern of locations. In this approach one must choose a quality measure and an algorithm for minimizing it.

9.2 Regular Patterns

9.2.1 Transect and Grid Sampling

Due to its operational advantages, a commonly used cluster type for estimating variograms is the transect. When sampling on transects, one has to choose the number of locations per transect, the distances between the locations, and the location and orientation of the transects. The distance between neighbouring locations on a given transect, i.e., the sampling interval, can be chosen constant or varying. Pettit and McBratney (1993) recommend transects in three directions, with the sampling locations in a given transect exponentially spaced. For instance, one may select 25 transects of 6 locations with interpoint distances of 0.2, 1, 5, 25 and 125 meter, i.e., each time the interval increases by a factor five. The transects must be evenly spread over the target area, for instance by dividing the area into squares or compact geographical strata of equal area (Sect. 7.2.4), and selecting one transect from each square (Fig. 9.1).

Transect sampling is appropriate when the average distance across the study area is large compared to the range of the variogram. For such areas sampling on a regular grid would give insufficient information on the semivariance at lags smaller than the range. If the distance across the area is small compared to the range of the variogram, then an alternative for transect sampling is sampling on a regular grid, supplemented by locations at a short distance of some grid nodes. To account for anisotropy, triangular grids are more suitable than square grids (Yfantis et al., 1987).

The 'short distance locations' are used to estimate the semivariance at lags smaller than the grid distance. Accurate estimates of the semivariance at small lags are important for estimating the nugget of the variogram, and for choosing between alternative models, for instance a spherical or a Gaussian model. It is

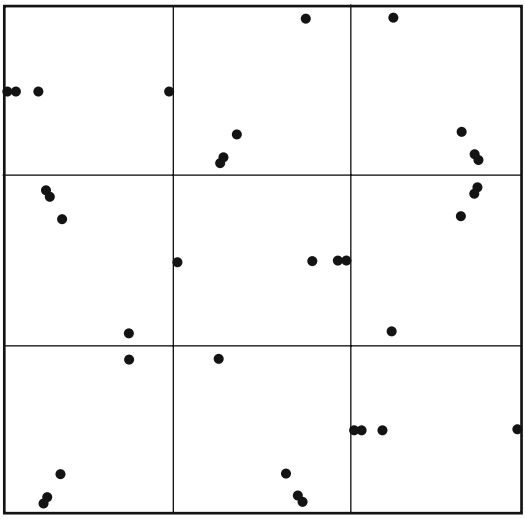

Fig. 9.1. Transect sampling from squares. After Pettit and McBratney (1993)

well known that if the variogram is used for geostatistical interpolation, then it is especially important to have reliable information on the variogram near the origin. We recommend to select at least 30 to 50 additional locations per lag, because this results into 30 to 50 disjoint pairs of locations, which is the minimum number of pairs mentioned by Journel and Huijbregts (1978).

To avoid spatial clustering of short distance locations in certain parts of the area, we recommend to select the grid nodes that will receive a short distance location purposively and not at random, for instance by subsampling the regular grid systematically (see Fig. 9.2). Also, we recommend to locate the additional locations on the sides of the grid cells, so that the directions for the smallest lag coincide with those of the larger lags.

9.2.2 Nested Sampling

In Nested Sampling the sampling locations are selected in stages (batches) in such a way that the distance between a randomly selected location of a given stage to a location of a previous stage is controlled. Figure 9.3 shows a nested design with four stages: in the first stage three locations with a mutual distance of h_1 m are randomly selected from the area. In the second stage at each of these locations in a random direction a location is selected at a distance of $h_2 = h_1/3$ m. We now have $3 \times 2 = 6$ locations. In the third stage at each of these six locations in a random direction a locations is selected at a distance of $h_2/3$ m, which makes the total number of locations $3 \times 2 \times 2 = 12$ locations.

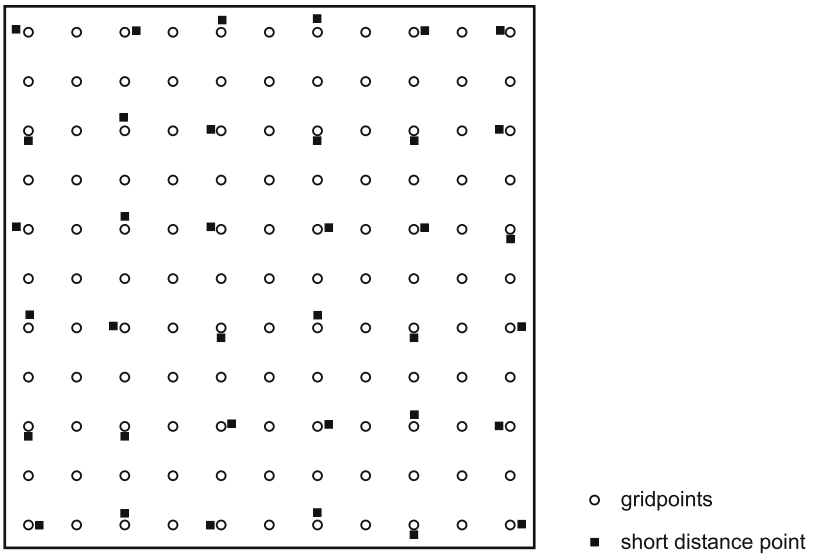

Fig. 9.2. Square grid sample with additional short distance locations for variogram estimation

This procedure is repeated once more, leading to a total of 24 locations. Note that in two dimensions it is impossible to select more than three mutually equidistant locations.

In practice, in the first stage often more than three locations are selected, for instance at the nodes of a square grid with a grid distance of h_1 m (Oliver, 1984). However, in that case one must be aware that the largest lag at which the semivariance is estimated is larger than h_1 m. Note that in the design of Fig. 9.3 the sample size doubles at each stage. So, if one has five stages with 9 locations in the first stage, the size of the nested sample becomes $9 \times 2 \times 2 \times 2 \times 2 = 144$. Adding a sixth stage implies an increase of the sample size of 144 locations.

Clearly, with nested designs one generally cannot afford many stages, because that would imply a too large sample size. However, full replication at each stage is unnecessary because with this design the variogram at the smaller lags is estimated much more precisely than at the larger distances. Therefore one may decide to replicate for the lower stages at only a proportion of the units, leading to so-called unbalanced designs. For the nested sample of Fig. 9.3 a 50% replication at the fourth stage leads to a total of 18 locations. For a nested design with five levels, 9 locations in the first stage, and a 50% replication at the fifth stage the total sample size becomes 108. One can now afford a sixth stage of 36 locations for the same budget, leading to an estimate of the variogram at one extra lag near the origin.

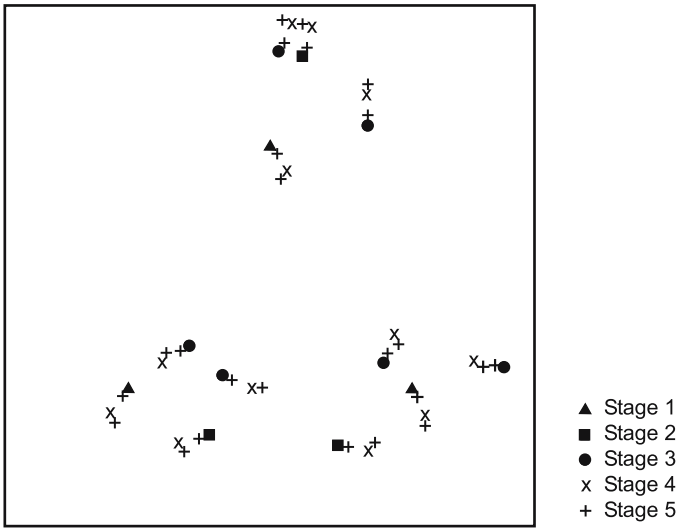

Fig. 9.3. Balanced nested sample with four stages. In the first stage three locations are selected. In subsequent stages at each of the sampling locations of the previous stages one additional location is selected

9.2.3 Independent Sampling of Pairs of Locations

If the method-of-moments is used to estimate the variogram (see Sect. 9.4.1), then one must be aware that the estimated semivariances for the M lags are logically correlated. These correlations should be taken into account when fitting the variogram model, but calculating the correlations is cumbersome. Moreover, from time-series analysis it is well known that due to these correlations one may be misled as to the type of model. For instance, the experimental variogram (the estimated semivariances for several lags) may show a hole effect which does not exist in reality.

These problems can be avoided by selecting pairs of locations independently as proposed by Brus and de Gruijter (1994). They proposed a design-based approach for estimating local (non-ergodic) variograms (Isaaks and Srivastava, 1988). In this approach, first a number of lags and the numbers of pairs per lag are chosen. To choose these lags and numbers the theory of experimental design can be used. Then for each lag the pairs of locations are selected by some type of design. For instance, in Simple Random Sampling (Sect. 7.2.3) of $M(h)$ pairs of locations with lag h, $M(h)$ locations are selected at random, with equal probability and independently from each other. Then for each location a counterpart is selected in a random or fixed direction at distance h from the starting location. If this counterpart is outside

the target area, also the starting location is omitted, and a new starting location is selected. Due to the independent selection of the pairs of locations, the estimated semivariances at the selected lags are uncorrelated, and their sampling variances can be estimated simply. The same holds for the sampling variance and covariance of the parameters of the fitted model. These advantages concerning the quantification of the uncertainty about the variogram are at the expense of the precision, because only $n/2$ (n is total number of locations) pairs of locations are used in the estimation, and the remaining $n\,(n-2)/2$ pairs are neglected. It depends on the situation how large this loss of information is and whether it is outweighed by the mentioned advantages.

9.3 Optimized Sampling Patterns

One may calculate the sampling pattern that explicitly has a minimum value for some objective function. In this case a quality measure and an algorithm to optimize it has to be selected. In the first papers on this subject a quality measure is proposed that quantifies how close the numbers of location pairs per lag class are to prespecified numbers (Warrick and Myers, 1987). However, the problem then shifts to the choice of the numbers of location pairs, and the question becomes what distribution of numbers of location pairs is best. Müller and Zimmerman (1999) and Lark (2002) have shown that a uniform distribution is sub-optimal; see also Müller (2001).

In subsequent papers it was proposed to base the quality measure on the variance-covariance matrix of the estimated parameters of a variogram model (Zimmerman and Homer, 1991; Bogaert and Russo, 1999; Müller and Zimmerman, 1999). For variograms that are non-linear functions of the parameters this is not straightforward. If one approximates the non-linear function by a first-order Taylor expansion, then the variance-covariance matrix of the parameters fitted by Generalized Least Squares to the estimated semivariances can be approximated by

$$\mathbf{V}_{\hat{\theta}} \approx \left(\mathbf{G}'_{\theta}\, \mathbf{V}_{\hat{\gamma}}^{-1}\, \mathbf{G}_{\theta} \right)^{-1} , \tag{9.3}$$

where $\mathbf{G}_{\theta}$ is the $M \times p$ matrix with the partial derivatives of the variogram, evaluated at the true (but unknown) values of the parameters:

$$\mathbf{G}_{\theta} = \begin{bmatrix} \frac{\partial \gamma(\mathbf{h}_1;\theta)}{\partial \theta_1} & \cdots & \frac{\partial \gamma(\mathbf{h}_M;\theta)}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \gamma(\mathbf{h}_M;\theta)}{\partial \theta_1} & \cdots & \frac{\partial \gamma(\mathbf{h}_M;\theta)}{\partial \theta_p} \end{bmatrix} , \tag{9.4}$$

and $\mathbf{V}_{\hat{\gamma}}$ is the $M \times M$ variance–covariance matrix of the estimated semivariances. Cressie (1993) shows how $\mathbf{V}_{\hat{\gamma}}$ can be calculated when the SF is assumed

to be second-order stationary (Appendix B.1) and multivariate Gaussian. Bogaert and Russo (1999) and Müller and Zimmerman (1999) proposed to optimize the locations by minimizing the determinant of $\mathbf{V}_{\hat{\theta}}$, which is equivalent to maximizing the determinant of the so-called information matrix $\mathbf{G}'_{\theta}\,\mathbf{V}^{-1}_{\hat{\gamma}}\,\mathbf{G}_{\theta}$ in (9.3). Such designs are referred to as D-optimal designs. Note that both $\mathbf{G}_{\theta}$ and $\mathbf{V}_{\hat{\gamma}}$ depend on the parameters of the variogram, and therefore to minimize $\det(\mathbf{V}_{\hat{\theta}})$ with respect to the sampling locations, one must know the variogram. So, there is a circular problem.

A way out is to sample in two phases, and to use estimates of the parameters from the relatively small first phase sample to optimize the sample of the second phase. Bogaert and Russo (1999) optimized the pattern of 100 locations for an exponential variogram without nugget (sill = 1; effective range = 1/2 of side of square) and for an exponential variogram with nugget (nugget = 0.5; sill = 1; effective range = 1/2 of side of square). For the exponential variogram without nugget the gain in precision of the estimated parameters compared to Simple Random Sampling was limited. For the exponential variogram with nugget the optimized sample had many locations at very short distance, and as a result the estimated nugget was considerably more precise than for Simple Random Sampling. There was also gain for the sill, however for the range parameter the gain was again limited. Müller and Zimmerman (1999) and Boer et al. (2001) studied the effect of ignoring the correlations between the location pairs. This implies that the variogram is estimated by Weighted Least Squares (weights equal to $n(\mathbf{h})/\gamma^2(\mathbf{h};\theta)$) instead of Generalized Least Squares (Cressie, 1985). The quality measure then slightly changes: in the matrix $\mathbf{V}_{\hat{\gamma}}$ of (9.3) the off-diagonal elements are substituted by zeroes. Müller and Zimmerman (1999) found that this simplification led to a very similar pattern which was only slightly inferior to the optimal pattern. Boer et al. (2001) found that ignoring correlations led to strong clustering of locations, even for a variogram without nugget. Boer et al. (2001) also found that the value of the quality measure, $\det(\mathbf{V}_{\hat{\gamma}})$, and the pattern of locations was rather insensitive to changes in the parameters of the variogram, although there was some influence of the range and (for large ranges) of the nugget.

Finally, Lark (2002) proposed a quality measure based on the suitability of the estimated variogram for geostatistical interpolation. This seems rational because in many cases the ultimate goal is not the variogram itself, but a map obtained by kriging with the variogram. It is well-known that the kriging variance is more sensitive to changes in the variogram than the kriging prediction itself, and this is the main reason why Lark proposed to look at the kriging variance (Appendix B, equation B.19). Due to uncertainty about the variogram, there is also uncertainty about this kriging variance, and Lark (2002) proposed to use as a quality measure the variance of the kriging variance. This variance is approximated by a first-order Taylor expansion:

$$V(V_{\mathrm{K}}) \approx \mathbf{g}'_{V}\,\mathbf{V}_{\hat{\theta}}\,\mathbf{g}_{V}\ , \tag{9.5}$$

where $\mathbf{g}_V$ is the p-vector with partial derivatives of the kriging variance to the variogram parameters:

$$\mathbf{g}_V = \begin{bmatrix} \frac{\partial V_{\mathrm{K}}}{\partial \theta_1} \\ \vdots \\ \frac{\partial V_{\mathrm{K}}}{\partial \theta_p} \end{bmatrix} , \tag{9.6}$$

and $\mathbf{V}_{\hat{\theta}}$ is the $p \times p$ variance–covariance matrix of the estimated variogram parameters. Lark (2002) optimized 49 sampling locations assuming an exponential variogram with nugget, $\gamma(h) = c_0 + c_1\{1 - \exp(-h/a)\}$, at three levels for the distance parameter a, and for the ratio of spatial dependence $c_1/(c_0+c_1)$, resulting into nine combinations. Lark (2002) considered the kriging variance at the centre of a square grid of 5 units. Figure 9.4 shows the optimized sampling locations. For a small ratio of spatial dependence (large nugget-to-sill ratio) and/or a small range the optimized sample showed several clusters of locations. For the intermediate ratio of spatial dependence combined with the two largest ranges the optimized sample showed a more or less regular distribution with some of the locations supplemented by an additional location at short distance. For the largest ratio of spatial dependence and the two largest ranges the optimized sample has a more or less regular pattern with several chains of locations. Lark (2002) compared the optimized samples with 7 randomly selected transects of 7 locations with a regular spacing. He found comparable values for the quality measure, and therefore concluded that when one is ignorant about the variogram, then the most robust approach is to sample on transects. In an experiment where a first phase sample of 7 randomly selected transects of 7 locations each was supplemented by an optimized sample of 31 locations, there was a benefit from the optimization compared to sampling entirely on transects.

9.4 Estimating the Variogram

We shall now describe how the variogram can be estimated from the sample data. The most widely used method for variogram estimation is the method-of-moments. Alternatively, the variogram can be estimated by the maximum likelihood method. In principle for all sampling designs described above both estimation methods can be used. For nested designs the natural way of estimating the variogram is ANOVA. By summing the variance components associated with the stages we get the variogram for the chosen lags (Webster and Oliver, 1990). Miesch (1975) and Corsten and Stein (1994) have shown that for balanced designs ANOVA and the method-of-moments are equivalent and lead to identical experimental variograms. For unbalanced designs these two methods lead to different estimates.

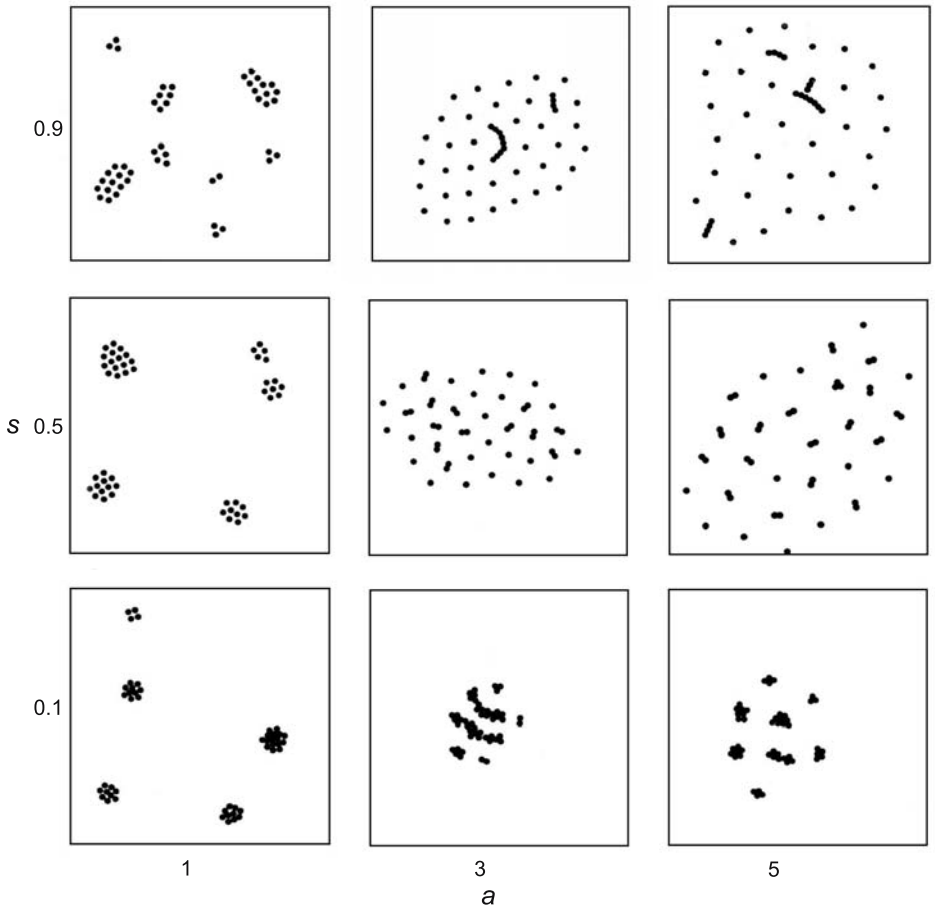

Fig. 9.4. 49 sampling locations optimized for the kriging variance of predictions at the centre of the square grid cells with a spacing of 5 distance units. An exponential variogram is assumed, with varying distance parameters, and ratios of spatial dependence. (Reproduced from Lark (2002, p. 69) with permission from Elsevier.)

9.4.1 Method-of-Moments

With the method-of-moments, the variogram is estimated in two steps. In the first step, the data are used to form pairs of locations. Then we estimate the variogram for a given lag $\mathbf{h}$ by selecting all pairs of sampling locations $\mathbf{h}$ apart, and calculating:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2M(\mathbf{h})} \sum_{i=1}^{M(\mathbf{h})} \{z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h})\}^2 \ , \tag{9.7}$$

where $M(\mathbf{h})$ is the number of pairs separated by lag $\mathbf{h}$. Unless locations are chosen at regular intervals along transects or on grids, it is necessary to group the lags by distance and direction, i.e., to use tolerance intervals both for the length and for the direction of the vector. If it is assumed that the variogram is independent of direction, i.e., an omnidirectional or isotropic variogram, then one may group all lags with (approximately) the same length, and estimate the experimental variogram for these lags. For kriging one needs a continuous function, and therefore in the second step a continuous model is fitted to the experimental variogram. Only models are permitted which ensure that the variances of the prediction errors are non-negative. Permissible models that are commonly used are the spherical, the exponential, and the Gaussian model (Appendix B). The estimates $\hat{\gamma}(\mathbf{h})$ are correlated, and therefore the parameters of the model must be estimated by taking into account the variances and covariances of the values in the experimental variogram. In statistical literature this is referred to as Generalized Least Squares estimation. This implies that the parameters must be estimated iteratively, because the variances and covariances depend on the variogram itself. Cressie (1985) recommended to neglect the covariances, and to fit the model by Weighted Least Squares, using the number of pairs divided by the squared semivariance in the previous iteration as weights.

9.4.2 Maximum Likelihood Estimation

Contrary to the method-of-moments, with the maximum likelihood method the data are not paired into couples, and the variogram is estimated in one step. To apply this method one typically assumes that the n sample data are a realization of a second-order stationary n-variate Gaussian Stochastic Function. Second-order stationarity is a slightly stronger assumption than the intrinsic hypothesis because the variance of the process is assumed finite (Appendix B). Especially the assumption that the process is multivariate Gaussian is a rather strong assumption. When the Spatial Cumulative Distribution Function is clearly non-Gaussian, we recommend transforming the data first, for instance by taking logarithms or square roots, and to estimate the variogram of the transformed data.

One needs the assumption of a multivariate Gaussian process, because then the joint probability density of the sample data can be calculated by:

$$P(\mathbf{z}, \boldsymbol{\mu}, \mathbf{p}) = (2\pi)^{-\frac{n}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})' \, \mathbf{V}^{-1} \, (\mathbf{z} - \boldsymbol{\mu}) \right\} , \tag{9.8}$$

where $\mathbf{z}$ is the vector with the n sample data, $\boldsymbol{\mu}$ is the vector with means (all values are equal), $\mathbf{p}$ is the vector with parameters of the covariance function, and $\mathbf{V}$ is the $n \times n$ matrix with variances and covariances of the sample data. This equation may not be familiar to the reader, however rewriting this equation for $n = 1$ gives

$$P(z, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\}, \tag{9.9}$$

the univariate normal density function. Note that there is a matrix with covariances in the equation, and not a matrix with semivariances, however for second-order stationary Stochastic Functions the variogram can be obtained from the covariance function (Appendix B, equation B.15). If the values at the sampling locations are considered as fixed, then (9.8) can be used to calculate the probability of finding these values for any combination of values for μ and for the parameters of the covariance function. The parameters of the covariance function can now be estimated by maximizing this probability. The estimates thus obtained are referred to as maximum likelihood estimates. Usually the logarithm of the probability is taken and multiplied by -1, and this negative log-likelihood is minimized, which is equivalent to maximizing the likelihood itself. Lark (2000) compared the method-of-moments and the maximum likelihood method. In general the maximum likelihood estimates were better than the method-of-moments estimates, especially for small sample sizes (say $n < 75$) and when spatial structure is moderate to strong (small nugget, large range). Lark (2000) also found that, although the method-of-moments does not make the multivariate Gaussian assumption, this method is equally sensitive to skewness of data as the maximum likelihood method. For larger sample sizes ($n > 150$) ML estimation becomes impractical because a huge number of computations are then required. For more information on this method we refer to Pardo-Igúzquiza and Dowd (1998).