B.6 The Variogram and Kriging

Margaret A. Oliver

B.6.1 Introduction

Spatial statistics and geostatistics have developed to describe and analyze the variation in both natural and man-made phenomena on, above or below the land surface. Spatial statistics includes any of the formal techniques that study entities that have a spatial index (Cressie 1993). Geostatistics is embraced by this general umbrella term, but originally it was more specifically concerned with processes that vary continuously, i.e. have a continuous spatial index. The term geostatistics applies essentially to a specific set of models and techniques developed largely by Matheron (1963) in the 1960s to evaluate recoverable reserves for the mining industry. These ideas had arisen previously in other fields; they have a long history stretching back to Mercer and Hall (1911), Youden and Mehlich (1937), Kolmogorov (1941), Gandin (1965), Matérn (1960) and Krige (1966). Geostatistics has since been applied in many different fields, such as agriculture, fisheries, hydrology, geology, meteorology, petroleum, remote sensing, soil science and so on. In most of these fields the data are fragmentary and often sparse, therefore there is a need to predict from them as precisely as possible at places where they have not been measured. This chapter covers two of the principle techniques of geostatistics that solve this need for prediction; the variogram and kriging.

B.6.2 The theory of geostatistics

A brief summary only is given here of the theory that underpins geostatistics (for more detail see Journel and Huijbregts, 1978; Goovaerts, 1997; Webster and Oliver 2007). Most spatial properties vary in such a complex way that the variation cannot be defined deterministically. To deal with this spatial uncertainty a different approach from the traditional deterministic methods of spatial analysis was required that relies on a stochastic or probabilistic approach. The basis of modern geostatistics is to treat the variable of interest as a random variable. This implies

M.M. Fischer and A. Getis (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, DOI 10.1007/978-3-642-03647-7_17,

that at each point x in space there is a series of values for a property, $Z(x)$, and the one observed, $z(x)$, is drawn at random according to some law, from some probability distribution. At x, a property $Z(x)$ is a random variable with a mean, μ and variance, σ^2. The set of random variables, $Z(x_1)$, $Z(x_2)$, …, is a random process, and the actual value of Z observed is just one of potentially any number of realizations of the random process. In classical statistics this set of observed values, the realization, is the population.

To define the variation of the underlying random process, we can take into account the fact that the values of regionalized variables at places near to one another tend to be related. As well as estimating the mean and variance of the property, we can also estimate the spatial covariance to describe this relation between pairs of points. The covariance for the random variables is given by

$$C(x_1,x_2)=E[\{Z(x_1)-\mu(x_1)\}\{Z(x_2)-\mu(x_2)\}] \tag{B.6.1}$$

where $\mu(x_1)$ and $\mu(x_2)$ are the means of Z at x_1 and x_2, and E denotes the expected value. This solution is unavailable, however, because the means are unknown as there is only ever one realization of Z at each point. To proceed we have to invoke assumptions of stationarity.

Stationarity

Under the assumptions of stationarity certain attributes of the random process are the same everywhere. We assume that the mean, $\mu = E[Z(x)]$, is constant for all x, and so $\mu(x_1)$ and $\mu(x_2)$ can be replaced by μ, which can be estimated by repetitive sampling. When x_1 and x_2 coincide, Eq. (B.6.1) defines the variance (or the *a priori* variance of the process), $\sigma^2 = E[\{Z(x) - \mu\}^2]$, which is assumed to be finite and, as for the mean, the same everywhere. When x_1 and x_2 do not coincide, their covariance depends on their separation and not on their absolute positions, and this applies to any pair of points x_i, x_j separated by the lag $h = x_i - x_j$ (a vector in both distance and direction), so that

$$C(x_i x_j)=E[\{Z(x_i)-\mu\}\{Z(x_j)-\mu\}]=E[\{Z(x)\}\{Z(x+h)\}-\mu^2]=C(h) \tag{B.6.2}$$

which is also constant for a given h. This constancy of the first and second moments of the process constitutes second-order or weak stationarity. Equation (B.6.2) indicates that the covariance is a function of the lag and it describes quantitatively the dependence between values of Z with changing separation or lag distance. The autocovariance depends on the scale on which Z is measured; therefore, it is often converted to the dimensionless autocorrelation by

$$\rho(h) = C(h)/C(0) \tag{B.6.3}$$

where $C(0) = \sigma^2$ is the covariance at lag zero.

Intrinsic variation and the variogram

The mean often appears to change across a region and then the variance will appear to increase indefinitely as the extent of the area increases. The covariance cannot be defined because there is no value for μ to insert into Eq. (B.6.2). This is a departure from weak stationarity. Matheron's (1965) solution to this was the weaker intrinsic hypothesis of geostatistics. Although the general mean might not be constant, it would be for small lag distances and so the expected differences would be zero as follows:

$$E[Z(x) - Z(x+h)] = 0 \tag{B.6.4}$$

and the expected squared differences for those lags define their variances

$$E[\{Z(x) - Z(x+h)\}^2] = \text{var}[Z(x) - Z(x+h)] = 2\gamma(h). \tag{B.6.5}$$

The quantity $\gamma(h)$ is known as the semivariance at lag h, or the variance per point when points are considered in pairs. As for the covariance, the semivariance depends only on the lag and not on the absolute positions of the data. As a function of h, $\gamma(h)$ is the semivariogram or more usually the variogram.

If the process $Z(x)$ is second-order stationary, the semivariance and covariance are equivalent:

$$\gamma(h) = C(0) - C(h) = \sigma^2\{1 - \rho(h)\}. \tag{B.6.6}$$

However, if the process is intrinsic only there is no equivalence because the covariance function does not exist. The variogram is valid, however, and therefore it can be applied more widely than the covariance function. This makes the variogram a valuable tool and as a consequence it has become the cornerstone of geostatistics.

B.6.3 Estimating the variogram

This section describes two methods for estimating the variogram from data, Matheron's method of moments and the residual maximum likelihood (REML) method, together with the main features that variograms are likely to have.

The method of moments estimator

The empirical semivariances can be estimated from data, $z(x_1)$, $z(x_2)$, …, by

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i + h)\}^2 \qquad \text{(B.6.7)}$$

where $z(x_i)$ and $z(x_i+h)$ are the actual values of Z at places (x_i) and (x_i+h), and $m(h)$ is the number of paired comparisons at lag h. By changing h, an ordered set of semivariances is obtained; these constitute the experimental or sample variogram. Equation (B.6.7) is the usual formula for computing semivariances; it is often referred to as Matheron's method of moments (MoM) estimator. The way that this equation is implemented as an algorithm depends on the configuration of the data. For a regular transect the lag becomes a scalar, $h = |h|$, for which the semivariances can be computed only at integral multiples of the sampling interval. The number of paired comparisons decreases one at a time as the lag interval is increased. The maximum lag should be set to no more than a third of the length of the transect. For a regular grid, semivariances can be calculated along the rows and columns of the grid and the lag increment is the grid interval. For irregularly sampled data in one or more dimensions, or to compute the omnidirectional variogram of data on a regular grid, the separations between pairs of points are placed into bins with limits in both separating distance and direction, Fig. B.6.1. In this figure, 0L is the nominal lag interval of length h, w is the width of the bin, $\alpha/2$ is the angular tolerance and θ is one of a set of directions. To calculate the variogram over all directions, the omnidirectional variogram, $\alpha/2$ is set to 180º and θ is set to zero.

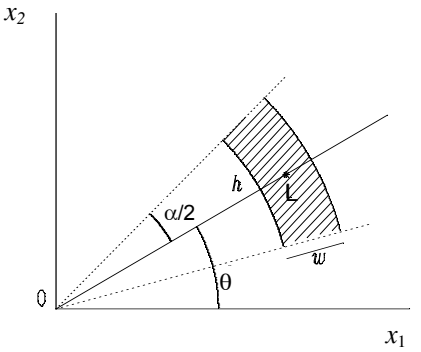

Fig. B.6.1. Discretization of the lag into bins for irregularly scattered data

The choice of narrow bins tends to give rise to erratic variograms, whereas wide bins tend to smooth and result in a loss of detail. You can see the effect of this in Fig. B.6.4. For a grid, it is usual to choose the grid interval as the nominal lag interval and for irregularly scattered data, the average distance between sampling points.

Webster and Oliver (1992) have shown that at least 100 sampling points are required to estimate the MoM variogram reliably. For many situations these are more data than can be afforded, for example where the costs of sampling and or sample analysis are considerable. In other situations this sample size might result in a closer sample spacing than is needed to resolve the variation adequately; this occurs where the property of interest has a large scale of spatial variation relative to the extent of the study area. This would result in over-sampling and a waste of resources. Pardo-Igúzquiza (1997) suggested the maximum likelihood (ML) approach as an alternative to Matheron's estimator. He also suggested that where the number of data is relatively small (a few dozen), the ML variogram estimator offers an alternative that gives an estimate of the variogram parameters and of their uncertainty (Pardo-Igúzquiza 1998, pp. 462-464).

The residual maximum likelihood (REML) variogram estimator

By contrast to the MoM approach, the ML methods are parametric and they also assume that the process, Z, is second-order stationary. Following the notation of Kerry and Oliver (2007), it is assumed that the data, $z(x_i)$, $i = 1, \ldots, n$, a realization of this process, follow a multivariate Gaussian distribution with the joint probability density function (pdf) of the measurements defined by

$$p(z|\beta,\theta)=(2\pi)^{-\frac{n}{2}}|V|^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2}(z-X\beta)^{\mathrm{T}}V^{-1}(z-X\beta)\right\} \tag{B.6.8}$$

where z is a vector that contains the n data, θ contains the parameters of the covariance matrix, V is the n-by-n variance-covariance matrix, and $X\beta$ represents the trend. The matrix V can be factorized as

$$V = \sigma^2 A \tag{B.6.9}$$

where σ^2 is the variance and A is the autocorrelation matrix. The pdf can then be rewritten as

$$p(z|\beta,\sigma^2,\theta)=(2\pi)^{-\frac{n}{2}}\sigma^{-n}|A|^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2\sigma^2}(z-X\beta)^{\mathrm{T}}A^{-1}(z-X\beta)\right\} \tag{B.6.10}$$

where θ is the set of covariance parameters excluding the variance. The parameters, β, σ^2, θ, are estimated in such a way that they minimize the negative log-likelihood function given by

$$\ln L(\beta, \hat{\sigma}^2, \theta | z) = \tfrac{n}{2}\ln(2\pi) + n\ln(\sigma) + \tfrac{1}{2}\ln|A| + \tfrac{1}{2\sigma^2}(z - X\beta)^{\mathrm{T}} A^{-1}(z - X\beta). \quad \text{(B.6.11)}$$

In the ML approach the drift parameter, β, is estimated at the same time as the set of covariance parameters.

Simultaneous estimation of the trend and covariance parameters in the ML approach results in biased covariance parameter estimates (Matheron 1971; Kitanidis and Lane 1985). Residual maximum likelihood (REML) developed by Patterson and Thompson (1971) avoids this problem because instead of working with the original data, it uses linear combinations of the data. These latter, known as generalized increments, filter out the trend. The generalized increments, g, can be represented as

$$g = \Lambda z \quad \text{(B.6.12)}$$

where the matrix Λ is derived from the projection matrix

$$P = I - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} \quad \text{(B.6.13)}$$

by dropping p rows in Λ because there are p generalized increments that are linearly dependent on others (Kitanidis 1983). The matrix P has the property that

$$PX = 0 \quad \text{(B.6.14)}$$

then

$$Pz = PX\beta + Pe = Pe \quad \text{(B.6.15)}$$

which filters out the trend regardless of what the coefficients β are. The e are the residuals. Then

$$E(g) = 0 \quad \text{(B.6.16)}$$

and

$$E(g\, g^{\mathrm{T}}) = \Lambda V \Lambda^{\mathrm{T}}. \quad \text{(B.6.17)}$$

The increments, g, are assumed to be Gaussian and the covariance parameters are estimated by minimization of the negative log-likelihood function (NLLF), given by

$$\ln L^{\mathrm{T}}(\hat{\sigma}^2, \theta|g) = \tfrac{n-p}{2}\ln(2\pi) + \tfrac{n-p}{2} - \tfrac{n-p}{2}\ln(n-p) + \tfrac{1}{2}\ln|\Lambda\Lambda\Lambda^{\mathrm{T}}| + \tfrac{n-p}{2}\ln[g^{\mathrm{T}}(\Lambda\Lambda\Lambda^{\mathrm{T}})^{-1}g]. \tag{B.6.18}$$

The covariance parameters, θ, can include the nugget variance (see below for the definition), long- and short-range distance components for isotropic and anisotropic situations, together with the anisotropy ratio for the latter. Pardo-Igúzquiza's (1997) MLREML program computes these parameters for three covariance models, the spherical, exponential and Gaussian.

For both the ML and REML approaches there is no experimental variogram, and as a consequence there is no smoothing of the spatial structure because there is no *ad hoc* definition of lag classes (bins). This is particularly advantageous for irregularly spaced data.

Features of the variogram

Continuity. Most environmental variables are continuous, therefore we should expect $\gamma(h)$ to pass through the origin at $h = 0$ [Fig. B.6.2(a)]. In practice, however, the variogram often appears to approach the ordinate at some positive value as h approaches zero, Fig. B.6.2(b), which suggests that the process is discontinuous. This discrepancy is known as the *nugget variance*. For properties that vary continuously the nugget variance usually includes some measurement error, but mostly comprises variation that occurs over distances less than the shortest sampling interval. Figure B.6.2(c) is a pure nugget variogram which usually indicates that the sampling interval is too large to resolve the variation present.

Monotonic increasing. Figure B.6.2(a) and (b) shows that the semivariance increases with increasing lag distance. This indicates that at short distances the values of the $Z(x)$ are similar, but as the lag distance increases they become increasingly dissimilar on average. The monotonic increasing slope indicates that the process is *spatially dependent*.

Sill and range. Figure B.6.2(b) shows a variogram that reaches an upper bound after the initial slope; this bound is known as the *sill variance*. It is the *a priori* variance, σ^2, of the process. A bounded variogram describes a process that is second-order stationary. The distance at which the variogram reaches its sill is the *range*, i.e. the range of spatial dependence. Places further apart than the range are spatially independent, Fig. B.6.2(b).

Hole effect and periodicity. The variogram may decrease from its maximum to a local minimum and then increase again. This maximum is equivalent to a minimum in the covariance function in which it appears as a 'hole'. It suggests fairly regular repetition in the process. A variogram that fluctuates in a periodic way with increasing lag distance indicates greater regularity of repetition.

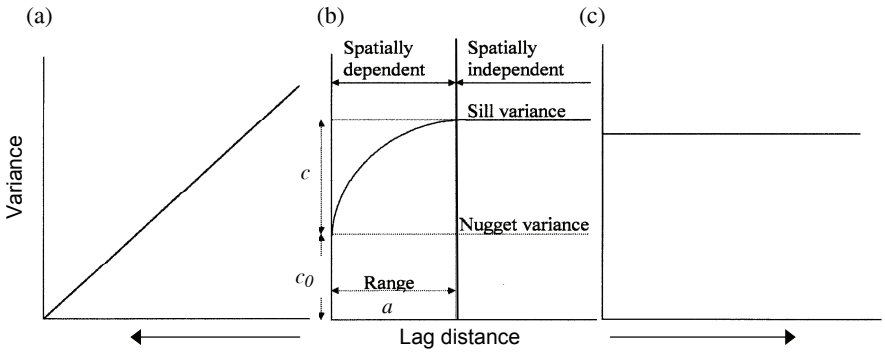

Fig. B.6.2. Three idealized variogram forms: (a) unbounded; (b) bounded; and (c) is the spatially correlated component [c_0 is nugget variance, a is the range of spatial dependence, $c + c_0$ is the sill variance, and c pure nugget]

Unbounded variogram. If the variogram increases indefinitely with increasing lag distance as in Fig. B.6.2(a), the process is intrinsic only.

Anisotropy. Spatial variation might not be the same in all directions. To explore data for any anisotropy, i.e. directional variation, the variogram must be computed in at least three directions. For a regular grid, it is usual to compute the variogram along the rows, columns and the principal diagonals. If there are four directions, start by setting the angular discretization to 22.5°, for example, and this angle can be decreased if there appears to be anisotropy. If the initial gradient or range of the variogram changes with direction and a simple transformation of the coordinates will remove it, then this is known as *geometric anisotropy.* An example of this is given in Fig. B.6.5 later in the case study; it shows the variogram of pH at Broom's Barn Farm computed in four directions from data on a regular grid. If the sill variance fluctuates with changes in direction, this might indicate the presence of preferentially orientated zones with different means. This is known as *zonal anisotropy.* It can sometimes be dealt with by stratifying the area of interest and then computing the variogram from the residuals of the class means. This is sometimes called the pooled within-class variogram.

Nested variation. Variation in the environment often occurs at several spatial scales simultaneously, and patterns in the variation can be nested within one another. This is usually evident when there are many data, for example from remote sensing etc. The experimental variogram will often appear more complex if more than one spatial scale is present; this can be seen in Fig. B.6.6. A combination of two or more simple models that are authorized can be used to model such a variogram. The simplest combined model is one with a nugget component. Spatial dependence may occur at two distinct scales and these can be represented in the variogram as two spatial components. Models describing more than one spatial structure are often known as nested functions; the nested or double spherical model has been the most commonly fitted, Fig. B.6.6(b).

B.6.4 Modeling the variogram

The experimental MoM variogram comprises a set of discrete estimates at particular lag intervals, which are subject to error that arises largely from sampling fluctuation. The underlying variogram, which represents the regional variation, is continuous. To obtain an approximation to this we can fit what are known as authorized functions that are conditional negative semi-definite (CNSD) to the experimental values. Functions that are CNSD will not give rise to negative variances when random variables are combined (see Webster and Oliver 2007 for more detail on this). There are a few principal features that the function must be able to represent:

(i) a monotonic increase with increasing lag distance from near the ordinate,
(ii) a constant maximum or asymptote (the sill),
(iii) a positive intercept on the ordinate (the nugget),
(iv) anisotropy.

There are a few simple functions only that encompass the above features and that are CNSD. They can be divided into those that are bounded, which represent processes that are second-order stationary, and those that are unbounded that are intrinsic only. There are several functions, but here we shall focus on those that are fitted most commonly in the environmental sciences. The formulae for the selected functions are given in their isotropic form, i.e. for $h = |h|$. A nugget variance, c_0, has been included because most experimental variograms if extended to the ordinate would have a positive intercept. The Gaussian model is included in many popular geostatistical packages, but it is excluded here. Its use can give rise to unstable kriging equations because the model approaches the origin with zero gradient (the limit for random variation), and this function will be replaced with the stable exponential model (Wackernagel 2003). Webster and Oliver (2007) describe a wide range of suitable variogram functions.

Circular model. The equation for the circular function is

$$\gamma(h) = \begin{cases} c_0 + c\left\{1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a}\right) + \frac{2h}{\pi a}\sqrt{1 - \frac{h^2}{a^2}}\right\} & \text{for } h \le a \\ c_0 + c & \text{for } h > a \\ 0 & \text{for } h = 0 \end{cases} \qquad \text{(B.6.19)}$$

where $\gamma(h)$ is the semivariance at lag h, c is the *a priori* variance of the autocorrelated process, c_0 is the nugget variance which represents the spatially uncorrelated variation at distances less than the sampling interval and measurement error, and a is the distance parameter, the range of spatial dependence or spatial autocorrelation. Values at places less than this apart are correlated, whereas those further apart are not. The combined $c_0 + c$ is the sill of the model. Theoretically the semivariance at lag zero is itself zero, but in practice there are usually too few estimates of $\gamma(h)$ near to the ordinate to fit a model through the origin. This function is CNSD in two dimensions. It curves tightly as it approaches the range (see Fig. B.6.4(i)).

Spherical function. This is one of the two most widely fitted models in the environmental sciences. Its equation is

$$\gamma(h) = \begin{cases} c_0 + c\left\{\frac{3h}{2a} + \frac{1}{2}\left(\frac{h}{a}\right)^3\right\} & \text{for } h \le a \\ c_0 + c & \text{for } h > a \\ 0 & \text{for } h = 0. \end{cases} \qquad \text{(B.6.20)}$$

The symbols have the same meaning as above. This model curves more gradually as the sill is reached than the circular one, see Fig. 6.4.4(c). This function is CNSD in three dimensions. It represents transition features that have a common extent that appear as patches, some with large values and other with small ones. The average diameter of the patches is represented by the range of the model.

Pentaspherical function. This model curves more gently as it approaches its sill than the preceding models, see Fig. B.6.3(b). It is CNSD in three dimensions. The pentaspherical function has the equation

$$\gamma(h) = \begin{cases} c_0 + c\left\{\dfrac{15h}{8a} - \dfrac{5}{4}\left(\dfrac{h}{a}\right)^3 + \dfrac{3}{8}\left(\dfrac{h}{a}\right)^5\right\} & \text{for } h \le \text{a} \\ c_0 + c & \text{for } h > \text{a} \\ 0 & \text{for } h = 0 \end{cases} \tag{B.6.21}$$

Exponential function. The exponential and spherical functions together account for a large proportion of the models fitted in the environmental sciences. Its equation is

$$\gamma(h) = c_0 + c\left\{1 - \exp\left(-\frac{h}{r}\right)\right\} \tag{B.6.22}$$

where c_0 and c have the same meanings as above, but the distance parameter is now r. The exponential model approaches its sill even more gently than the preceding models and also asymptotically so that it does not have a finite range. In practice, an effective range is assigned at the distance at which the function has reached 95 percent of c. The effective range, a', is $3r$. It is CNSD in three dimensions. The exponential function also represents transition structures, but they now have random extents.

Stable exponential. This is a useful substitute for the Gaussian function for experimental variograms that appear to approach the origin with a reverse curvature; they can be represented by the general equation

$$\gamma(h) = c_0 + c\left\{1 - \exp\left(-\frac{h^\alpha}{r^\alpha}\right)\right\} \tag{B.6.23}$$

in which $1 < \alpha < 2$. For the Gaussian function $\alpha = 2$, which is excluded because it represents differentiable variation in the process, which is not random. Webster and Oliver (2006) used the stable exponential function to describe topographic variation.

Unbounded models. Variograms that are intrinsic only increase without bound as the lag distance increases. These can usually be fitted by power functions, which have the general equation including a nugget variance of

$$\gamma(h) = c_0 + wh^\alpha \tag{B.6.24}$$

where w describes the intensity of the process, and the exponent, α, describes the curvature. If $\alpha < 1$, the curve is convex upwards; if it is one it is a straight line and

w is the gradient; and if $\alpha > 1$ the curve is concave upwards. The exponent must lie strictly between zero and two.

Modeling anisotropy. If the experimental variogram is anisotropic, then the variation is a function of distance, h, and direction, θ. Geometric anisotropy can be made isotropic by a linear transformation of the coordinates. The transformation is defined by reference to an ellipse

$$\Omega(\theta)=\sqrt{\mathrm{A}^2\cos^2(\theta-\phi)+\mathrm{B}^2\sin^2(\theta-\phi)} \tag{B.6.25}$$

where A and B are the long and short diameters of the ellipse, respectively, and ϕ is its orientation, i.e. the direction of the long axis. For bounded models, Ω replaces the distance parameter of the isotropic variogram as follows for the exponential variogram (see Fig. B.6.5(b)).

$$\gamma(h,\theta) = c_0 + c\left[1 - \exp\left\{-\frac{|h|}{\Omega(\theta)}\right\}\right] \tag{B.6.26}$$

and for the power function it replaces the gradient

$$\gamma(h,\theta) = c_0 + [\Omega(\theta)h]^{\alpha} \tag{B.6.27}$$

Nested models. The nested spherical function is given by

$$\gamma(h)=\begin{cases} c_0+c_1\left\{\dfrac{3h}{2a_1}-\dfrac{1}{2}\left(\dfrac{h}{a_1}\right)^3\right\}+c_2\left\{\dfrac{3h}{2a_2}-\dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3\right\} & \text{for } 0 < h \le a_1 \\ c_0+c_1+c_2\left\{\dfrac{3h}{2a_2}-\dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3\right\} & \text{for } a_1 < h \le a_2 \\ c_0+c_1+c_2 & \text{for } h = a_2 \end{cases} \tag{B.6.28}$$

where c_1 and a_1 are the sill and range of the short-range component of the variation, and c_2 and a_2 are the sill and range of the long-range component. A nugget component can also be added as above (see Fig. B.6.6(b)).

B.6.5 Case study: The variogram

We illustrate some of the principles of geostatistics with results from a recent study on precision farming for the British Home-Grown Cereals Authority (Oliver and Carroll 2004). The field (UK National Grid reference SU 458174) covers 23ha on the Yattendon Estate, Berkshire, England. It is on part of the Chalk downland of southern England and has the typical undulating topography of this region. From the extensive set of survey data obtained during 2002 we have selected topsoil (0–15 cm) available potassium. Data on yield of winter wheat were available for 2001 to illustrate nested variation. Table B.6.1 gives the summary statistics for these two variables.

Sampling for the soil survey was at the nodes of a 30m × 30m grid, with additional observations at 15m intervals along short transects from randomly selected grid nodes. The sampling intervals were based on scales of variation determined from several years of yield data with the aim of ensuring that the variation in the soil (of which there was no prior knowledge) would be represented adequately and efficiently. At each site ten cores of soil were bulked from a support of 5m × 2m to form the sample; this helps to reduce the locally erratic variation that contributes to the nugget variance. There were 230 data points, which enabled any anisotropy in the variation to be determined; this sample size is close to the 250 data recommended by Webster and Oliver (1992).

Table B.6.1. Summary statistics

Statistic	Topsoil K [mg l^{-1}]	Yield 2001 [t ha^{-1}]
Number	230	4060
Mean	142.5	6.838
Median	143.0	7.050
Minimum	48.1	1.000
Maximum	254.4	14.600
Variance	1367.5	3.909
Standard deviation	37.0	1.977
Skewness	0.1	–0.298

Experimental variograms were computed by Eq. (B.6.7) in four directions to reveal any anisotropy in the variation. The results for topsoil K are shown in Fig. B.6.3(a) for the directions 0°, 45°, 90° and 135°. There is little divergence among the different directions until lag 130m, after which the sills start to diverge. This suggests that there is zonal anisotropy in the variation of topsoil K in this field. Since the directional variograms are close together for the initial lags, the variation can be treated as isotropic for kriging, and the solid line shows the best fitting isotropic exponential function to the omnidirectional variogram.

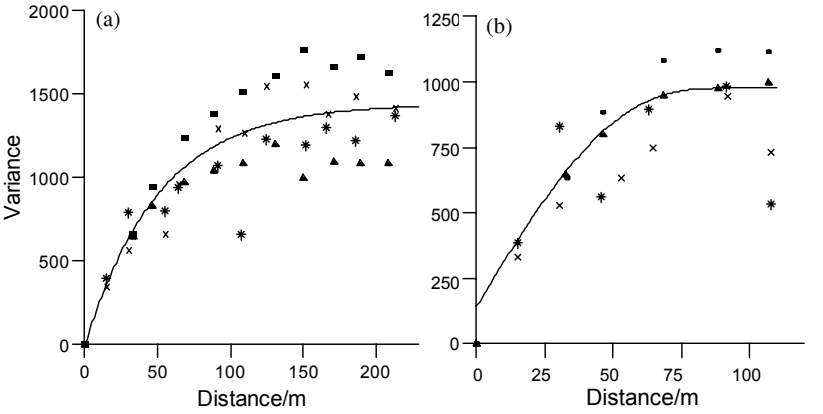

Fig. B.6.3. (a) Directional variogram computed on the raw data (230 points) from the Yattendon Estate, and (b) directional variogram computed on the residuals from the class means. The symbols represent: ∗ denotes 0º (E–W), ■ denotes 45º, × denotes 90º (N–S), ▲ denotes 135º

To illustrate the effect of sample size on the variogram, we subsampled the complete set of data (230 sampling points) to give subsets of 94 and 47 data. Experimental omnidirectional variograms were computed from the total data and two subsets for topsoil K. To explore the effect of different bin widths, variograms were computed for lag intervals of 15m (the sampling interval for the transects), 20m (mid-way between the transect and overall grid interval) and 40m (for illustration). Models were fitted to the experimental values using GenStat (Payne 2008).

Figure B.6.4 shows the experimental values as symbols and the fitted models as solid lines. The experimental variograms suggest that the 20m lag interval is a good comprise between the rather erratic result for the 15m interval and the loss of detail with the 40m lag interval. The experimental variograms also show the effect of decreasing the number of data; the variograms becomes more erratic and that computed from 47 data also shows a serious loss of variance.

Table B.6.2 gives the models and their parameters fitted to the experimental variograms. These show how sensitive the model parameters are to changes in lag interval and number of data. For the 230 data, the main difference in the model parameters for the variograms computed with different lag intervals is in the nugget variance, which is zero for the 15m lag. This suggests that the data from the transect sampling have resolved the local variation in topsoil K well. This is an important consideration when designing a sampling scheme. For a grid survey, it is worthwhile having some additional sampling points at shorter distances than the grid interval as in this survey because it helps to reduce the nugget variance. There were 40 sampling points at the shorter interval which is only 17 percent of the total data.

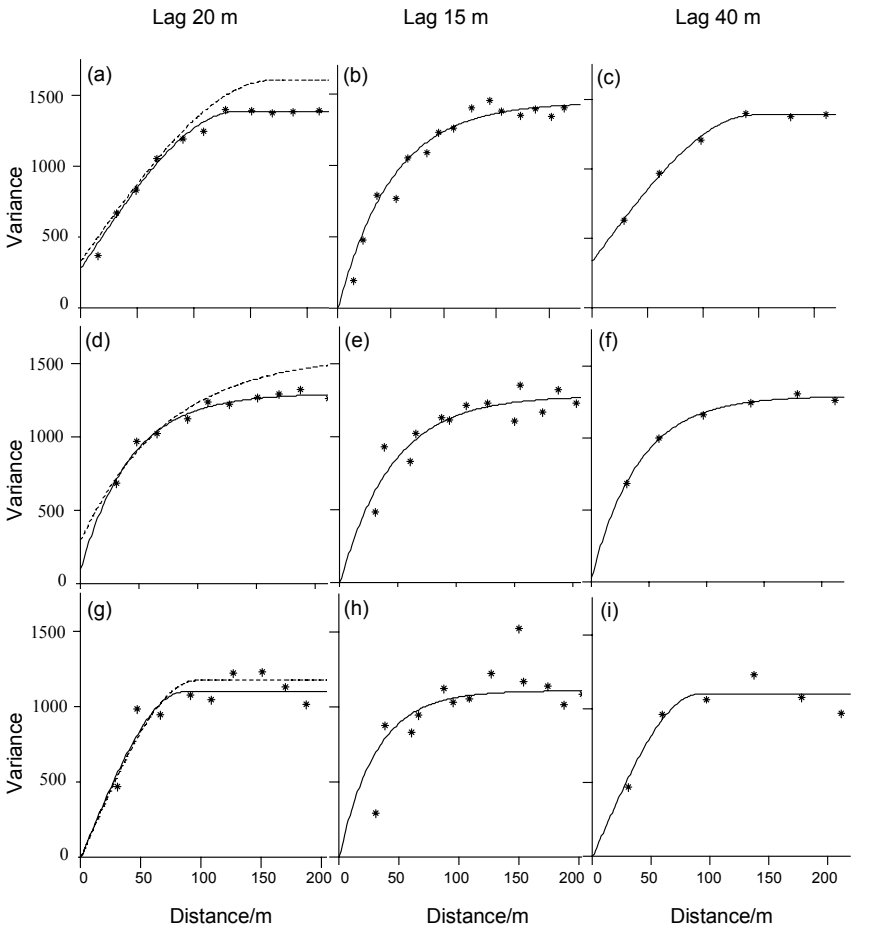

Fig. B.6.4. Experimental variograms (∗) computed by the method of moments (MoM) estimator for lag distances of 20m, 15m and 40m, and for the complete data set of 230 sites [(a), (b) and (c), respectively], subset of 94 data [(d), (e), (f)] and subset of 47 data [(g), (h), (i)] for topsoil K on the Yattendon Estate. The solid line is the model fitted to the MoM variogram and the dashed line is the variogram estimated by residual maximum likelihood (REML)

For the subsample of 94 data, the difference in model parameters from those for complete set of data is small; this indicates that Webster and Oliver's (1992) recommendation of a minimum of 100 data is adequate to obtain a reliable variogram. The model parameters for the smallest data set are considerably different from those of the complete set of data, suggesting that the variograms of the smallest data set are not an accurate reflection of the structure of the variation. For example, the sill variances are markedly less and the ranges of spatial dependence

Table B.6.2. Variogram model parameters

		Parameter			
Topsoil Property	Model type	Nugget variance	Correlated component	Range (m)	Sill variance
MoM estimator		c_0	c_1 c_2◆	a_1/m a_2/m+ or r/m*	
K (230 sites)					
Lag 20 m	Spherical	319.3	1070.0	142.9	1389.3
Lag 15 m	Exponential	0	1441.0	151.7	1441.0
Lag 40 m	Spherical	355.6	1035.7	148.4	1391.3
K residuals					
Lag 20 m	Pentaspherical	145.5	830.7	90.8	976.2
K (94 sites)					
Lag 20 m	Exponential	163.7	1138.0	44.6	1301.7
Lag 15 m	Exponential	0	1282.0	44.6	1109.0
Lag 40 m	Spherical	338.9	1051.0	146.4	1389.9
K (47 sites)					
Lag 20 m	Spherical	0	1098.0	85.3	1098.0
Lag 15 m	Exponential	0	1109.0	30.5	1109.0
Lag 40 m	Circular	0	1100.0	79.6	1100.0
pH Broom's Barn	Exponential	0	0.37	89.70	0.37
	Anisotropic	0	0.38	69.54	ϕ=1.09
	Exponential			114.50	
Yield 1995	Double	1.76	1.04	44.19	0.8882
	Spherical		1.16◆	277.50+	
REML estimator					
REML 230	Spherical	334.5	1273.5	170.6	1608.0
REML 94	Exponential	300.0	1262.6	74.0	1562.6
REML 47	Spherical	1.9	1171.1	95.7	1173.0

Notes: ◆ is the spatially correlated variance of the long-range spatial component, + is the range of the long-range spatial component, * is the distance parameter of the exponential function; to obtain a working range $a' = 3r$

are shorter. Table B.6.2 shows that the models are all bounded functions indicating that the variation has a patchy distribution.

Variograms were also computed by REML for the 20m grid interval, and are shown as the dashed line in Fig B.6.4(a), (d) and (g). The variograms estimated by REMLfor the two larger data sets are not as similar to those computed by MoM as one might expect. The sill variances are larger than the variance of the data. The range of the exponential model for the subset of 94 data is also much longer than that for the MoM variogram. The variograms estimated by REML and MoM are more similar to one another for the smallest data set, yet it is for these data that one would expect the greatest difference in model parameters. Although Kerry and Oliver (2007) showed a distinct advantage in computing variograms by REML for small sets of data, this is not particularly evident in the study described here.

The experimental variogram computed from the yield data of a crop of winter wheat (2001) shows a complex structure [see Fig. B.6.6(a)]. The best fitting model was a spherical function with two spatial components; one with a range of 44m

and the other of 278m. Figure B.6.6(b) shows the experimental variogram with the fitted model; the nugget, short- and long-range components of the model are also shown separately.

Anisotropy. Figure B.6.3(a) shows the directional variogram for topsoil K. It is evident that the sill variances disperse after a lag of about 130m. Zonal anisotropy cannot be dealt with by a simple transformation of the coordinates. If the region can be stratified into zones, then this is one way in which zonal anisotropy can be resolved. The variogram models suggest that the variation is patchy, which could arise from zones that are preferentially orientated and with different means. A classification of these data had been done previously (see Frogbrook and Oliver 2007 for details), therefore the class means were subtracted from the values of K for the appropriate class.

The directional variogram was then computed on the residuals from the class means, Fig. B.6.3(b). The directional variogram is shown by the symbols for the four directions and the isotropic models fitted to the omnidirectional variograms by the solid black line for both the raw data and the residuals. Stratification has effectively removed the zonal anisotropy – some scatter remains in the different directions but this is to be expected from sampling fluctuations. The model parameters have also changed considerably; the best fitting model is now a pentaspherical function with a sill variance of less than 1000 and a range of 91m. The model now has a much shorter range of spatial dependence, Table B.6.2, and so the variogram has been plotted to a maximum lag of 150m to take into account this difference. There is no marked evidence of anisotropy over distances less than the range.

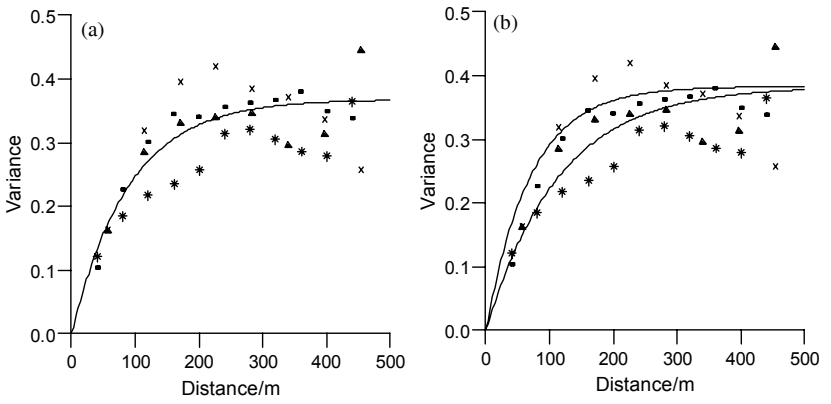

Fig. B.6.5. Directional variogram computed on the pH data from Broom's Barn Farm (433 sampling points): (a) with the best fitting isotropic model (solid line), and (b) with an isotropic exponential function (the solid lines show the envelope of this function). The symbols represent: ∗ denotes 0º (E–W), ▪ denotes 45º, × denotes 90º (N–S), ▲ denotes 135º, and the solid lines are the isotropic models fitted to the omnidirectional variograms

To illustrate geometric anisotropy we have used the data for pH from Broom's Barn Farm. This is an experimental sugar beet farm near to Bury St. Edmunds, Cambridgeshire, UK (see Webster and Oliver 2007, for more detail on these data). Figure B.6.5(a) shows the directional variogram which illustrates how the semivariances in the different directions start to diverge after a lag of 80m. The solid line is the best fitting isotropic model, an exponential function (Table B.6.2). Figure B.6.5(b) shows the directional variogram with the fitted anisotropic exponential function. The two lines show the envelope of this function and Table B.6.2 gives the parameters of the fitted function. The direction of maximum variation and of the shorter range is about 60º (where 0º is E–W) and the direction of minimum variation is perpendicular to this.

Nested variation. Figure B.6.6(a) shows the experimental variogram for yield 2001 at the Yattendon Estate; it appears to have a complex structure. Several models were fitted and the one with the smallest residual sums of squares was a nested spherical function, which is shown as the solid line fitted to the experimental values in Fig. B.6.6(b). The model parameters for yield 2001 are given in Table B.6.2. To illustrate the individual components of this model, we have shown them separately in Fig. B.6.6(b) as lines with different ornament. The complex structure identified from the experimental variogram is evident as two markedly different ranges of spatial variation of 44m and 278m.

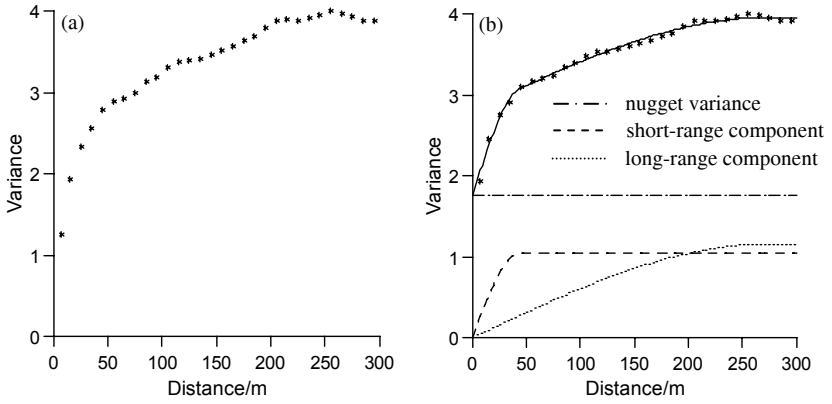

Fig. B.6.6. Variogram of yield 2001 for the Yattendon Estate: (a) experimental variogram (symbols), and (b) the experimental variogram with the fitted double spherical model (solid line); the ornamented lines represent the individual model components

B.6.6 Geostatistical prediction: Kriging

Kriging is a method of optimal prediction or estimation in geographical space, often known as a best linear unbiased predictor (BLUP). It is the geostatistical method of interpolation for random spatial processes. Matheron (1963) first used the term 'kriging' for the method in recognition of D. G. Krige's contribution to improving the precision of estimating concentrations of gold and other metals in ore bodies. Krige (1951) had observed that he could improve estimates of ore grades in mining blocks by taking into account the grades in neighbouring blocks. Matheron (1963) expanded Krige's empirical ideas and put them into the theoretical framework of geostatistics. However, Matheron's developments were not in isolation; the mathematics of simple kriging had been worked out by A. N. Kolmogorov in the 1930s (Kolmogorov 1939, 1941), by Wold (1938) for time series analysis and later by Wiener (1949). Cressie (1993) gives a brief history of the origins of kriging.

Kriging provides a solution to a fundamental problem faced by environmental scientists of predicting values from sparse sample data based on a stochastic model of spatial variation. Most properties of the environment (soil, vegetation, rocks, water, oceans and atmosphere) can be measured at any of an infinite number of places, but for economic reasons they are measured at relatively few. Several mathematical methods of interpolation are available, for example, Thiessen polygons, triangulation, natural neighbour interpolation, inverse functions of distance, least-squares polynomials (trend surfaces) and splines. Most of these methods take account of systematic or deterministic variation only and disregard the errors of prediction. Kriging, on the other hand, overcomes the weaknesses of these mathematical interpolators. It makes the best use of existing knowledge by taking account of the way a property varies in space through the variogram or covariance function. Kriging also provides not only predictions but also the kriging variances or errors. It can be regarded simply as a method of local weighted moving averaging of the observed values of a random variable, Z, within a neighbourhood, V. Kriging can be done for point (punctual kriging) or block supports of various size (block kriging), depending upon the aims of the prediction, even though the sample information is often for points.

Since its original formulation, kriging has been elaborated to tackle increasingly complex problems in disciplines that use spatial prediction and mapping. It is used in mining, petroleum engineering, meteorology, soil science, precision agriculture, pollution control, public health, monitoring fish stocks and other animal densities, remote sensing, ecology, geology, hydrology and other disciplines. As a consequence, kriging has become a generic term for a range of BLUP least-squares methods of spatial prediction in geostatistics. The original formulation of kriging, now known as ordinary kriging (Journel and Huijbregts 1978), is the most robust method and the one most often used.

Types of kriging

Ordinary kriging assumes that the mean is unknown and that the process is locally stationary. Simple kriging, which assumes that the mean is known, is used little because the mean is generally unknown. However, it is used in indicator and disjunctive kriging in which the data are transformed to have known means. Lognormal kriging is ordinary kriging of strongly positively skewed data transformed by logarithms to approximate a lognormal distribution. Kriging with trend enables data with a strong deterministic component (non-stationary process) to be analyzed;

Matheron (1969) originally introduced universal kriging for this purpose, but the state-of-the-art is empirical-BLUP (Stein 1999), which uses the REML variogram (Lark et al. 2006). Matheron (1982) developed factorial kriging or kriging analysis for variation that is nested. It estimates the long- and short-range components of the variation separately, but in a single analysis. Ordinary cokriging (Matheron 1965) is the extension of ordinary kriging to two or more variables that are spatially correlated. If some property that can be measured cheaply at many sites is spatially correlated or coregionalized with others that are expensive to measure and recorded at many fewer sites, the latter can be estimated more precisely by cokriging with the spatial information from the former.

Disjunctive kriging (Matheron 1973) is a non-linear parametric method of kriging. It is valuable for decision-making because the probabilities of exceeding (or not) a predefined threshold are determined in addition to the kriged estimates. Indicator kriging (Journel 1982) is a non-linear, non-parametric form of kriging in which continuous variables are converted to binary ones (indicators). It can handle distributions of almost any kind and can also accommodate 'soft' qualitative information to improve prediction. Probability kriging was proposed by Sullivan (1984) because indicator kriging does not take into account the proximity of a value to the threshold, but only its geographic position. Bayesian kriging was introduced by Omre (1987) for situations in which there is some prior knowledge about the drift or trend.

Ordinary kriging

Ordinary kriging is by far the most widely used type of kriging. It is based on the assumption that the mean is unknown. Consider that a random variable, Z, has been measured at sampling points, x_i, $i = 1, \ldots n$, and we want to use this information to estimate its value at a point x_0 (punctual kriging) with the same support as the data by

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i) \qquad \text{(B.6.29)}$$

where n usually represents the data points within the local neighbourhood, V, and is much less than the total number in the sample, N, and λ_i are the weights. To ensure that the estimate is unbiased the weights are made to sum to one

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{B.6.30}$$

and the expected error is $E[\hat{Z}(x_0) - Z(x_0)] = 0$. The prediction variance is

$$\mathrm{var}[\hat{Z}(x_0)] = E[\{\hat{Z}(x_0) - Z(x_0)\}^2] = 2\sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i, x_j) \tag{B.6.31}$$

where $\gamma(x_i, x_j)$ is the semivariance of Z between points x_i and x_j, $\gamma(x_i, x_0)$ is the semivariance between the ith sampling point and the target x_0. The semivariances are derived from the variogram model because the experimental semivariances are discrete and at limited distances.

Kriged predictions are often required over areas (block kriging) that are larger than the sample support of the data. The estimate is a weighted average of the data, $z(x_1)$, $z(x_2)$, …, $z(x_n)$, at the unknown block,

$$\hat{Z}(B) = \sum_{i=1}^{n} \lambda_i z(x_i). \tag{B.6.32}$$

The estimation variance of $Z(B)$ is:

$$\mathrm{var}[\hat{Z}(B)] = E[\{\hat{Z}(B) - Z(B)\}^2] = 2\sum_{i=1}^{n} \lambda_i \bar{\gamma}(x_i, B) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i, x_j) - \bar{\gamma}(B, B) \tag{B.6.33}$$

where $\bar{\gamma}(x_i, B)$ is the average semivariance between data point x_i and the target block B, and $\bar{y}(B, B)$ is the average semivariance within B, the within block variance.

Equation (B.6.31) for a point leads to a set of $n + 1$ equations in the $n + 1$ unknowns

$$\sum_{i=1}^{n}\lambda_i\,\gamma(x_i,x_j)+\psi(x_0)=\gamma(x_j,x_0) \qquad \text{for all } j \tag{B.6.34}$$

$$\sum_{i=1}^{n}\lambda_i=1 \tag{B6.35}$$

the Lagrange multiplier, ψ, is introduced to achieve minimization. The kriging equations in matrix form for punctual kriging are

$$A\lambda = b \tag{B.6.36}$$

where A is the matrix of semivariances between data points, $\gamma(x_i, x_j)$, b is the vector of semivariances between data points and the target, $\gamma(x_i, x_0)$ and λ is the vector of weights and the Lagrange multiplier. The kriging weights are obtained as follows by inverting matrix A,

$$\lambda = A^{-1}\,b. \tag{B.6.37}$$

The weights, λ_i, are inserted into Eq. (B.6.29) to give the prediction of Z at x_0. The kriging (prediction or estimation) variance is then

$$\sigma^2(x_0)=\sum_{i=1}^{n}\lambda_i\,\gamma(x_i,x_0)+\psi(x_0) \tag{B.6.38}$$

and in matrix form

$$\sigma^2(x_0)=b^{\mathrm{T}}\lambda. \tag{B.6.39}$$

Punctual kriging is an exact interpolator – the kriged value at a sampling site is the observed value there and the estimation variance is zero. The equivalent kriging system for blocks is

$$\sum_{i=1}^{n}\lambda_i\,\gamma(x_i,x_j)+\psi(B)=\bar{\gamma}(x_j,B) \qquad \text{for all } j \tag{B.6.40}$$

$$\sum_{i=1}^{n}\lambda_i=1 \tag{B6.41}$$

and the block kriging variance is obtained as

$$\sigma^2(B) = \sum_{i=1}^{n} \lambda_i \bar{\gamma}(x_i, B) + \psi(B) - \bar{\gamma}(B, B) \tag{B.6.42}$$

and in matrix form

$$\sigma^2(B) = b^{\mathrm{T}} \lambda - \bar{\gamma}(B, B). \tag{B.6.43}$$

Block kriging results in smoother estimates and smaller estimation variances overall because the nugget variance is contained entirely in the within-block variance, $\bar{\gamma}(B, B)$, and it does not contribute to the block kriging variance.

For many environmental applications kriging is most likely to be used for interpolation and mapping. The values of the property are usually estimated at the nodes of a fine grid, and the variation can then be displayed by isarithms or by layer shading. The estimation variances or standard errors can also be mapped similarly: they are a guide to the reliability of the estimates, where sampling is irregular, such a map may indicate if there are parts of a region where sampling should be increased to improve the estimates.

Kriging weights

The kriging weights depend on the variogram and the configuration of the sampling. The way in which the data points within the search radius are weighted is one feature that makes kriging different from classical methods of prediction where the weights are applied arbitrarily. Webster and Oliver (2007) illustrate how the weights vary according to changes in the nugget: sill ratio, the range, type of model, sampling configuration and the effect of anisotropy. The weights are particularly sensitive to the nugget variance and anisotropy. Weights close to the point or block to be estimated carry more weight than those further away, which shows that kriging is a local predictor. As the nugget: sill ratio increases the weights near to the target decrease and those further away increase. For a pure nugget variogram, the kriging weights are all the same and the estimate is simply the mean of the values in the neighbourhood. The effect of the range is more complex than for the nugget: sill ratio because it is also affected by the type of variogram model. In general, however, as the range increases the weights increase close to the target. For data that are irregularly distributed, points that are clustered carry less weight individually than those that are isolated.

The fact that the points nearest to the target generally carry the most weight has practical implications. It means that the search neighbourhood need contain no more than 16–20 data points, which in turn means that matrix A in the kriging system need never be large.

Factorial kriging

If the variogram of $Z(x)$ is nested, it can be represented as a combination of S individual variograms

$$\lambda(h) = \gamma^1(h) + \gamma^2(h) + \cdots + \gamma^S(h) \tag{B.6.44}$$

where the superscripts refer to the component variograms. If we assume that the processes represented by these components are uncorrelated, then Eq. (B.6.44) can be written as

$$\lambda(h) = \sum_{k=1}^{S} b^k g^k(h) \tag{B.6.45}$$

where $g^k(h)$ is the kth basic variogram function and b^k is a coefficient that measures the relative contribution of the variance $g^k(h)$ to the sum.

The components on the right-hand side of Eq. (B.6.45) correspond to S random functions that in sum form $Z(x)$, which can be represented as

$$Z(x) = \sum_{k=1}^{S} Z^k(x) + \mu \tag{B.6.46}$$

in which μ is the mean of the process. Each $Z^k(x)$ has an expectation zero, and the squared differences are

$$\tfrac{1}{2}E\left[\left\{Z^k(x) - Z^k(x+h)\right\}\left\{Z^{k'}(x) - Z^{k'}(x+h)\right\}\right] = \begin{cases} b^k g^k(h) & \text{if } k = k' \\ 0 & \text{otherwise.} \end{cases} \tag{B.6.47}$$

The last component, $Z^S(x)$ could be intrinsic only, so that $g^S(h)$ in Eq. (B.6.45) is unbounded with gradient b^S. This equation expresses the mutual independence of the S random functions, and enables the values of the contributing processes to be estimated separately by factorial kriging. Each spatial component $Z^k(x)$ is estimated as a linear combination of the observations, $z(x_i)$, $i = 1, \ldots, n$

$$\hat{Z}^k(x_0) = \sum_{i=1}^{n} \lambda_i^k \, z(x_i). \tag{B.6.48}$$

The λ_i^k are weights assigned to the observations, but now they must sum to zero, not to one, to ensure that the estimate is unbiased and to accord with Eq. (B.6.46).

Subject to this condition, they are chosen to minimize the kriging variance. This leads to the kriging system

$$\sum_{j=1}^{n} \lambda_j^k \gamma(x_i, x_j) - \psi^k(x_0) = b^k g^k(x_i, x_0) \qquad \text{for all } i = 1, \ldots, n \tag{B.6.49}$$

$$\sum_{j=1}^{n} \lambda_j^k = 0 \tag{B.6.50}$$

where $\psi^k(x_0)$ is the Lagrange multiplier for the kth component. This system of equations is solved for each spatial component, k, to find the weights, λ_i^k, which are then inserted into Eq. (B.6.48) for that component. Estimates are made for each spatial scale, k, by solving Eq. (B.6.49).

Kriging is usually done in small moving neighbourhoods centred on x_0, as for ordinary kriging. Thus, from a theoretical point of view, it is necessary only that $Z(x)$ is locally stationary. Equation (B.6.46) can then be rewritten as

$$Z(x) = \sum_{k=1}^{S} Z^k(x) + \mu(x) \tag{B.6.51}$$

where $\mu(x)$ is a local mean that can be considered as a long-range spatial component. We need to krige the local mean, which is again a linear combination of the data:

$$\hat{\mu}(x_0) = \sum_{j}^{n} \lambda_j^{\text{mean}} z(x_j). \tag{B.6.52}$$

The weights are obtained by solving the kriging system:

$$\sum_{j=1}^{n} \lambda_j^{\text{mean}} \gamma(x_i, x_j) - \psi^{\text{mean}}(x_0) = b^k g^k(x_i, x_0) \quad \text{for all } i = 1, \ldots, n \tag{B.6.53}$$

$$\sum_{j=1}^{n} \lambda_j^k = 1. \tag{B.6.54}$$

Estimating the long-range component can be affected by the size of the moving neighbourhood (Galli et al. 1984). To estimate a spatial component with a given range, the distance across the neighbourhood should be at least equal to that range. If the sampling is intensive and the range is large, there are so many data within the chosen neighbourhood that only a small proportion of them is retained for kriging, and those are all near to the target. Although modern computers can

handle many data at a time, the inversion of such large matrices can be unstable. Further, only the nearest few data to the target contribute to the estimate because they screen the more distant data. Consequently, the neighbourhood used is smaller than the one specified, which means that the range of the component estimated is smaller than that determined from the variogram. Galli et al. (1984) suggested a way of overcoming this shortcoming by selecting only a proportion of the data within the specified neighbourhoods. Such a selection is arbitrary, and Jaquet (1989) proposed an alternative that involves adding the estimate of the local mean to the estimated long-range component. Following Oliver et al. (2000), this is the solution we have adopted for the case study below.

B.6.7 Case study: Kriging

The case study describes applications of ordinary kriging with an isotropic variogram model and with an anisotropic one where there are directional differences in the variation. Factorial kriging is applied to explore variation that is described best by a nested variogram function.

Ordinary kriging

The complete set of data and the two data subsets of topsoil potassium from Yattendon are used to illustrate ordinary kriging. Predictions were made at unsampled places at the nodes of a 5m × 5m grid by ordinary punctual and block kriging. A minimum of seven and a maximum of 20 points were the limits set for the number of data in the neighbourhood. For block kriging, estimates were made over blocks of 10m × 10m. The parameters of the variogram models fitted to the MoM experimental variograms of each data set for the 20m lag (Table B.6.2) were used with the respective data for kriging. The kriged predictions were mapped in Gsharp. Figure B.6.7 shows the maps of block kriged estimates; those from punctual kriging are not shown as they appear so similar. The map based on the 230 data, Fig. B.6.7(a), shows the detail in the variation of topsoil K from the intensive sampling. The areas of small concentrations are where the soil is more sandy and the largest concentrations are in a dry valley that extends from NW to SE across the field where the soil contains more clay and silt. The map based on the sample size of 94, Fig. B.6.7(b), which is close to Webster and Oliver's (1992) minimum recommended size for computing an accurate variogram, shows the main features of the variation in topsoil K, albeit with some loss of detail. From a management perspective this map would form a sound basis to manage applications of K in this field. This smaller sample size represents a saving of almost 60 percent in sampling effort. Figure B.6.7(c) is the block kriged map based on 47 data and the loss of detail is evident. It is clear that to reduce the sample size to this level would be unadvisable for managing K applications in this field.

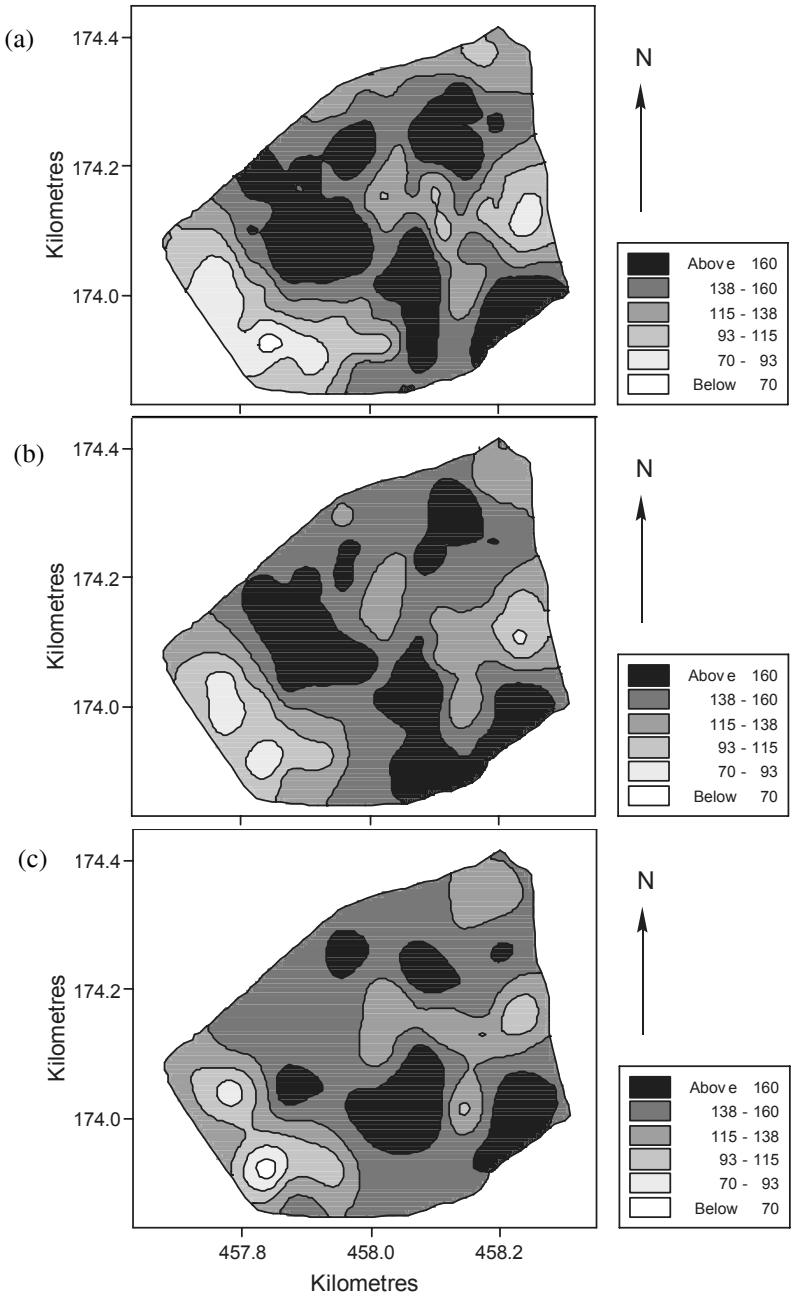

Fig. B.6.7. Maps of block kriged predictions of topsoil potassium at the Yattendon Estate for: (a) complete set of 230 data, (b) subset of 94 data, and (c) subset of 46 data

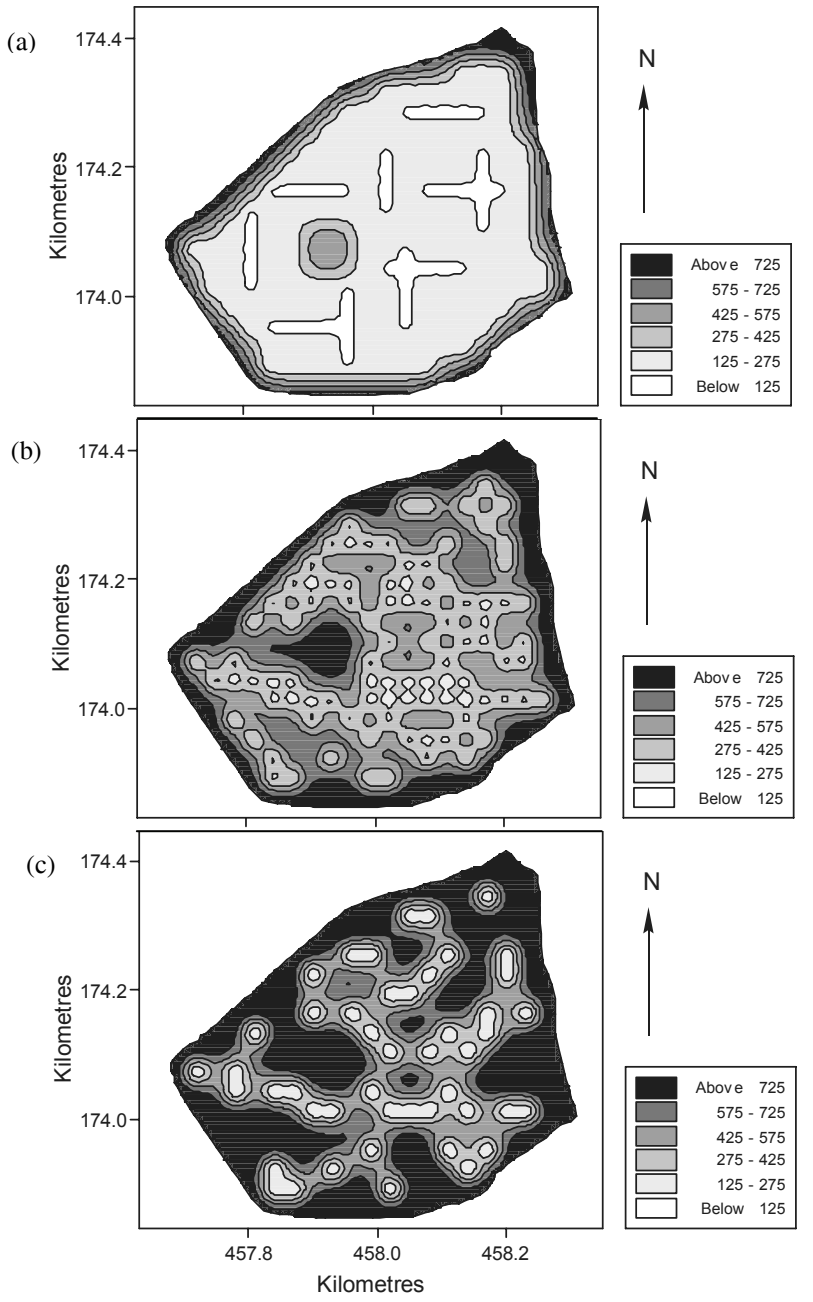

Fig. B.6.8. Maps of block kriged kriging variances for topsoil potassium at the Yattendon Estate for: (a) total of 230 data, (b) subset of 94 data, and (c) subset of 46 data

Figure B.6.8 (a), (b) and (c) shows the maps of block kriging variances for the three sizes of sample (230, 94 and 47, respectively); they show clearly how the variances of the predictions increase markedly with fewer data. The large kriging variances in the central part of the field in Fig. B.6.8(a) and (b) indicate an area with no sampling points where there is a copse. Figure B.6.8(a) shows that the smallest errors are along the short transects where the sampling was most intensive. Figure B.6.8(b) and (c) also shows that the kriging variances are smallest close to sampling points. The large variances around the field margins show the edge effects where there were fewer data from which to predict. These maps show that economizing on sampling to a sample size of 47 results in a loss of accuracy in the predictions that could have implications for subsequent management.

Figure B.6.9 shows the map of kriging variances from punctual kriging of the complete data set. Although the maps of estimates for punctual and block kriging were almost indistinguishable, the maps of kriging variance are quite different. The punctual kriging variances are much larger because the nugget variance sets a lower limit to the kriging variance. For block kriging the nugget variance disappears from the block kriging variance [see Eqs. (B.6.31) and (B.6.37)]. The larger is the proportion of nugget variance, the greater is the difference between the block and punctual kriging variances.

Kriging with an anisotropic model

The pH data from Broom's Barn Farm were used with the anisotropic exponential model for ordinary punctual kriging on a 10m × 10m grid. Figure B.6.10 shows the map of predictions. It is evident that there is more variation in pH from SSE to NNW than at right angles to this as the model in Table B.6.2 above describes.

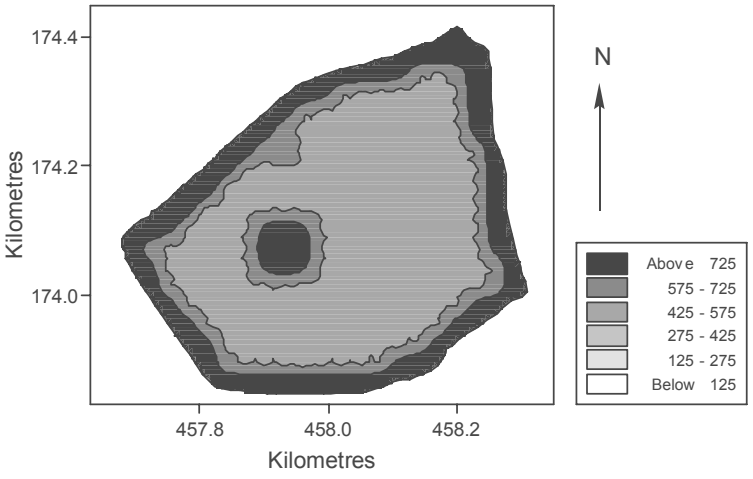

Fig. B.6.9. Map of punctually kriged kriging variances for topsoil potassium at the Yattendon Estate for the complete set of 230 data

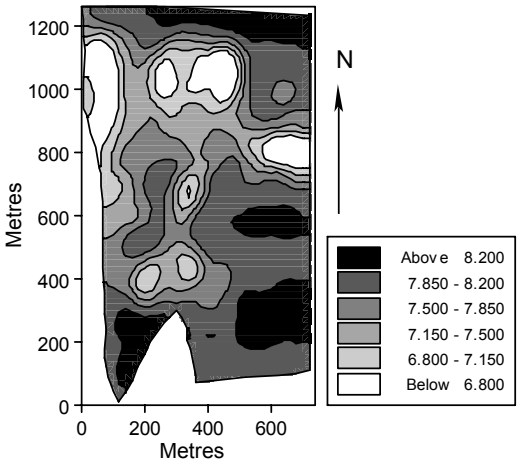

Fig. B.6.10. Map of punctually kriged predictions for topsoil pH at the Broom's Barn Farm

Nested variation: factorial kriging

The yield of winter wheat for 2001 from the Yattendon Estate is used to illustrate factorial kriging; its variogram (see Fig. B.6.6 and Table B.6.2) shows that there is more than one scale of variation present. Predictions were made at the nodes of a 5m × 5m grid as for topsoil K at Yattendon. The parameters of the double spherical model were used for ordinary kriging first; Fig. B.6.11(a) is the map of predictions. The pattern of variation appears complex because of the long- and short-range components of the variation. These components were then extracted separately and predicted by factorial kriging. Figure B.6.11(b) is the map of the long-range predictions. It is similar to that from ordinary kriging but it is less noisy because the short-range variation is no longer present. The regions of the field with large and small yields are clear in both maps. Many of the areas with large yield correspond to areas of large topsoil K concentrations [see Fig. B.6.11(a)]. The map of the short-range predictions, Fig. B.6.11(c) is quite different from the other two maps. It shows a much smaller scale of variation with a strong regular pattern.

This component of the variation appears to relate to the lines of management within the field in a NE–SW direction. The larger values are probably between the tramlines where the soil has suffered less compaction from machinery. There is some weak evidence of variation perpendicular to these lines that might reflect tramlines of previous operations. These management effects that have given rise to the short-range variation are not evident in the map of ordinary kriged predictions, Fig. B.6.11(a).

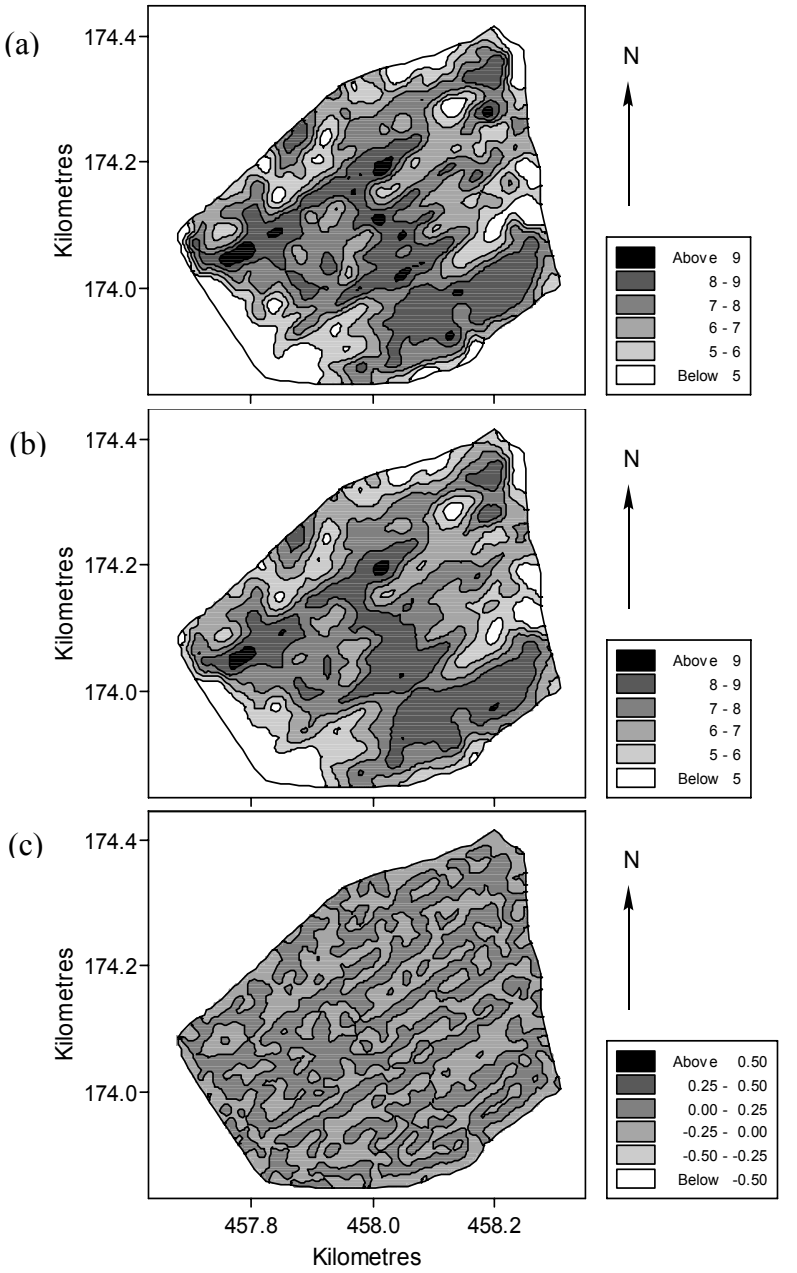

Fig. B.6.11. Maps of wheat yield for 2001 at the Yattendon Estate for: (a) ordinary kriged predictions, (b) predictions of the long-range component of the variation, and (c) predictions of the short-range component of the variation

For intensive data such as those from yield monitors, digital elevation models and satellites, factorial kriging is a valuable technique to explore the variation at different spatial scales. In this way it might be possible to gain some insight into the underlying processes that are responsible for variation at the different spatial scales.

Acknowledgements. The majority of the results in the case studies were from the author's project which was funded by the Home Grown Cereals Authority. We thank them for their support. We also thank Dr Z. L Frogbrook and Dr S. J Baxter for their work on this project. The data for Broom's Barn Farm were from an original survey of the Farm. We thank Dr J. D. Pidgeon for their use.

References

Cressie NAC (1993) Statistics for spatial data (revised edition). Wiley, New York, Chichester, Toronto and Brisbane

Frogbrook ZL, Oliver MA (2007) Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. Soil Use Mgmt 23(1):40-51

Galli A, Gerdil-Neuillet F, Dadou C (1984) Factorial kriging analysis: a substitute to spectral analysis of magnetic data. In Verly G, David M, Journel AG, Marechal A (eds) Geostatistics for natural resource characterization. Reidel, Dordrecht, pp.543-557

Gandin LS (1963) Objective analysis of meterological fields. Leningrad, Gidremeterologicheskoe Izdatel'stvo (GIMIZ) (translated by Israel Program for Scientific Translations, Jerusalem 1965)

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York

Jaquet O (1989) Factorial kriging analysis applied to geological data from petroleum exploration. Math Geol 21(7):683-691

Journel AG (1983) Non-parametric estimation of spatial distributions. J Int Ass Math Geol 15(3):445-468

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London

Kerry R, Oliver MA (2007) Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. Geoderma, Pedometrics 2005 140(4):383-396

Kitanidis PK (1983) Statistical estimation of polynomial generalized covariance functions and hydrological applications. Water Resources Research 19(4):909-921

Kitanidis PK, Lane RW (1985) Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method. J Hydrol79(1-2):53-71

Kolmogorov AN (1939) Sur l'interpolation et l'extrapolation des suites stationnaires. C R Acad Sci 208:2043-2045

Kolmogorov AN (1941) The local structure of turbulence in an incompressible fluid at very large Reynolds numbers. Doklady Academii Nauk SSSR 30:301-305

Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. J Chem Met Min Soc of South Africa 52(6):119-139

Krige DG (1966) Two-dimensional weighted moving average trend surfaces for ore-evaluation. J South African Inst Min Met 66(1):13-38

Lark RM, Cullis BR, Welham SJ (2006) On optimal prediction of soil properties in the presence of spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. Europ J Soil Sci 57(6):787-799

Matérn B (1966) Spatial variation: Stochastic models and their applications to problems in forest surveys and other sampling investigations. Meddelanden från Statens Skogsforskningsinstitut 49(5):1-144

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246-1266

Matheron G (1965) Les variables régionalisées et leur estimation, une application de la theorie de fonctions aleatoires aux sciences de la nature. Masson et Cie, Paris

Matheron G (1969) Le krigeage universel. Cahiers du Centre de Morphologie Mathématique No.1, Ecole des Mines de Paris, Fontainebleau

Matheron G (1971) The theory of regionalized variables and its applications. Cahiers No 5 du Centre de Morphologie Mathematique, Fontainebleau, France

Matheron G (1973) The intrinsic random functions and their applications. Adv Applied Probab 5(3):439-468

Matheron G (1982) Pour une analyse krigreante de données regionalisées. Note N-732 du Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, France

Mercer WB, Hall AD (1911) The experimental error of field trials. J Agricult Sci 4:107-132

Oliver MA, Carroll ZL (2004) Description of spatial variation in soil to optimize cereal management. Project Report 330 Home Grown Cereals Authority (HGCA), London

Oliver MA, Webster R, Slocum K (2000) Filtering SPOT imagery by kriging analysis. Int J Remote Sens 21(4):735-752

Omre H (1987) Bayesian kriging-merging observations and qualified guesses in kriging. Math Geol 19(1):25-39

Pardo-Igúzquiza E (1997) MLREML: a computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. Comp Geosci 23(2):153-162

Pardo-Igúzquiza E (1998) Inference of spatial indicator covariance parameters by maximum likelihood using MLREML. Comp Geosci.24(5):453-464

Payne R (ed) (2008) The guide to genstat release 10 - Part 2: statistics. VSN International, Hemel Hempstead

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. Biometrika 58(3):545-554

Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, Berlin, Heidelberg and New York

Sullivan J (1984) Conditional recovery estimation through probability kriging: theory and practice. In Verly G, David M, Journel AG, Marechal A (eds). Geostatistics for natural resource characterization. Reidel, Dordrecht, pp.365-384

Wackernagel H (2003) Multivariate geostatistics (3rd edition). Springer, Berlin, Heidelberg and New York

Webster R, Oliver MA (1992) Sample adequately to estimate variograms of soil properties. J Soil Sci 43(1):177- 192

Webster R, Oliver MA (2007) Geostatistics for environmental scientists (2nd edition). Wiley, New York, Chichester, Toronto and Brisbane

Wiener N (1949) Extrapolation, interpolation and smoothing of stationary time series. MIT Press, Cambridge [MA]

Wold H (1938) A study in the analysis of stationary time series. Almqvist and Wiksell, Uppsala

Youden WJ, Mehlich A (1937) Selection of efficient methods for soil sampling. Contributions of the Boyce Thompson Institute for Plant Research 9:59-70