Chapter 10
Classical Soil Geostatistics

R. Murray Lark and Budiman Minasny

"All the business of war, and indeed all the business of life, is to endeavour to find out what you don't know by what you do".

Arthur Wellesley
Duke of Wellington

10.1 Introduction

In 1971 Beckett and Webster reviewed the information on soil variability available to them. Their interest was, to use their term, in *lateral* variability of the soil. That is to say in the variation of the soil from place to place across the landscape (Beckett and Webster 1971). One way to capture this notion is by considering the variability of the soil as measured by sample variance, standard deviation or coefficient of variation (CV) within regions of different sizes. If the spatial location of soil observations is immaterial to their variation, then the variance and other quantities will be the same within regions of any size. Beckett and Webster (1971) found that, typically, half the variation of a soil property within an agricultural field, measured from conventional core samples, may be found within an area of 1 m^2. This shows us that the variability of the soil is spatially very intricate but also that a significant amount of variation is spatially structured over the intermediate scales. The fact that CV depends on area tells us something about the variability of soil. However, in the absence of intensive sensor measurements, the CV/area relationship is a cumbersome tool for describing spatial variability, requiring, in principle, sets of random samples from

R.M. Lark (✉)
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
e-mail: mlark@bgs.ac.uk

B. Minasny
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: budiman.minasny@sydney.edu.au

A.B. McBratney et al. (eds.), *Pedometrics*, Progress in Soil Science,
DOI 10.1007/978-3-319-63439-5_10

within randomly selected areas of different size. In a pioneering study Beckett and Bie (1976) sampled the soil on regular transects. They then grouped the observations into successive pairs, groups of three samples, groups of four samples, groups of n samples, etc. and for each $n = 2, 3, \ldots$ calculated the mean variances within the groups of n observations. Variance was then plotted against the length of the aggregated groups (log scale). It was shown that the form of the graphs related soil variability at different scales to vegetation and geomorphology in a predictable way.

This approach to analysis of spatial variability was innovative, developed from statistical literature available at the time (e.g. Yates 1948). However, other older statistical work had shown how the formalisms of analysis of variance could be used to describe spatial variation. This is the work of Youden and Mehlich (1937) who devised a sampling scheme to allow a partition of variance of a soil property between contrasts over different spatial intervals. This was achieved by a nested sampling design permitting a nested analysis in which, at each stage, the variance is partitioned into that observable between and within sampling stations separated by a particular interval. Webster and Butler (1976) revived the methodology and applied it to the analysis of soil variation in Australian Capital Territory, in a landscape where soil surveyors had struggled to represent the variation of soil by conventional soil maps. Webster and Lark (2013) discuss this sampling scheme and analysis in more detail.

While soil scientists were developing these innovations, statisticians in the mining industry, geostatisticians, were developing a formalism for the description of spatial variation which embraced and unified both the nested sampling of Youden and Mehlich and the transect analysis of Beckett and Bie. Their objective was to predict ore grades locally from a limited number of boreholes to support mine planning. The methods that they developed constitute what we might call classical geostatistics or mining geostatistics. The textbook of Journel and Huijbregts (1978) sets out the stall of these methods. Around the time that it was published, soil scientists became aware of geostatistics and the potential of geostatistical methods to solve the problems of soil survey: spatial prediction of soil conditions from limited sets of observations. Webster (2015) tells the story of how the connection between the armoury of mining geostatistical methods and the requirements of soil survey were first made.

This chapter discusses the classical methods of geostatistics as applied to soil science. It provides background to some of the more modern methods described in Chap. 11, but much of this standard methodology remains serviceable for the problem of spatial prediction. We first describe the random model of soil variation that underlies geostatistics and then show how this can be used for spatial prediction and inference.

10.2 Random Models of Soil Variation

This section presents a summary of the principles of geostatistical method, as set out by standard texts such as Journel and Huijbregts (1978), Isaaks and Srivastava (1989), Goovaerts (1997), Chiles and Delfiner (1999) and Webster and Oliver

(2007). Classical references on this methodology in the soil science literature are the papers of Burgess and Webster (1980a, b), McBratney and Webster (1983) and Webster and Burgess (1980).

10.2.1 Stationary Random Functions

A powerful approach to problems of inference is to treat data as if they are a realisation of a set of random variables. A simple example of a random variable is the number given by the throw of an unbiased die which is a random number, Y, which takes some value from the set $\{1,2,\ldots 6\}$. In the case of soil, we assume that a soil property s at location $\mathbf{x}$ is a realisation of a random variable. However, we can only ever observe one realisation, $s(\mathbf{x})$, the actual value of s at $\mathbf{x}$, which is not sufficient information to characterise a random variable. We can make progress if we assume there is a realisation not of a single variable but rather of a random variable $S(\mathbf{x})$ when the argument of the random function is location in space. If we can assume that certain properties of this function are constant for all $\mathbf{x}$ of interest, then a set of observations at different locations contains some replicated information from which we may infer properties of the random function.

10.2.1.1 Stationarity

The simplest, and strongest, assumption of stationarity is that the joint distribution of the random function over a set of locations, $\{S(\mathbf{x}_1),S(\mathbf{x}_2),\ldots S(\mathbf{x}_n)\}$, is identical to that for a set $\{S(\mathbf{x}_1+\mathbf{h}),S(\mathbf{x}_2+\mathbf{h}),\ldots S(\mathbf{x}_n+\mathbf{h})\}$ where $\mathbf{h}$ is any displacement or lag vector. This means that all moments of the distribution are constant so the mean $\mu = \mathrm{E}[S(\mathbf{x})]$, for all $\mathbf{x}$ and the covariance $\mathrm{E}[\{S(\mathbf{x})-\mu\}\{S(\mathbf{x}+\mathbf{h})-\mu\}]$, is constant for all $\mathbf{x}$ and $\mathbf{h}$, as are all higher moments. Note that E[.] denotes the statistical expectation of a random quantity in the square brackets. If we restrict these assumptions to the first two moments of the joint distribution (mean and covariances), then we have a weaker stationarity assumption – second-order or weak stationarity. Under the assumption of stationarity, it is therefore possible to define a covariance function

$$C(\mathbf{h}) = E\left[\{S(\mathbf{x})-\mu\}\{S(\mathbf{x}+\mathbf{h})-\mu\}\right]. \tag{10.1}$$

At lag zero the covariance function is equal to the variance of the random function. If the random variable shows no spatial dependence (it is a 'white noise' process in the terms of signal analysis or 'pure nugget' in geostatistical terms), then it is zero for all lag vectors which are non-zero. Spatial dependence is shown when the covariance declines with increasing lag separation reaching value zero for lags at which values of the random function are independent. Note that in some circumstances the covariance may not be strictly decreasing with lag distance, there may be a periodic component to the variation or quasiperiodic 'hole effect'.

It is possible to make a stationarity assumption weaker still (and so a plausible assumption about a broader class of soil variables). Intrinsic stationarity is, in effect, a second-order stationarity assumption about the increments $S(\mathbf{x}_i) - S(\mathbf{x}_i + \mathbf{h})$ $S(\boldsymbol{x}_i) - S(\boldsymbol{x}_i + \boldsymbol{h})$. So we assume that the mean increment is zero everywhere and that the variance of the increment is constant everywhere. Hence

$$E\left[\{S(\mathbf{x}) - S(\mathbf{x} + \mathbf{h})\}\right] = 0$$

and

$$E\left[\{S(\mathbf{x}) - S(\mathbf{x} + \mathbf{h})\}^2\right] = 2\gamma(\mathbf{h}). \tag{10.2}$$

The function $\gamma(\mathbf{h})$ given above is the semi-variogram. It is related to the covariance function in the second-order stationary case by

$$C(\mathbf{h}) = C(0) - \gamma(\mathbf{h}). \tag{10.3}$$

The covariance at lag 0, $C(0)$, is the variance of independently drawn values of S($\mathbf{x}$), also known as the a priori variance. We can use the variogram in a wider case of processes than the second- or higher-order stationary ones that can be described by the autocorrelation or covariance functions.

10.2.1.2 Variogram

Under the intrinsic hypothesis of stationarity, $\frac{1}{2}\{s(\mathbf{x}) - s(\mathbf{x} + \mathbf{h})\}^2$ and $\frac{1}{2}\{s(\mathbf{x}') - s(\mathbf{x}' + \mathbf{h})\}^2$ are both estimates of $\gamma(\mathbf{h})$. We may, therefore, combine all observations over lag $\mathbf{h}$ into an estimate of $\gamma(\mathbf{h})$. This is illustrated for a transect in Fig. 10.1. In the top row, the pairs of observations are combined to estimate the variogram for the lag equal to the basic sample interval. In the second row, the pairs

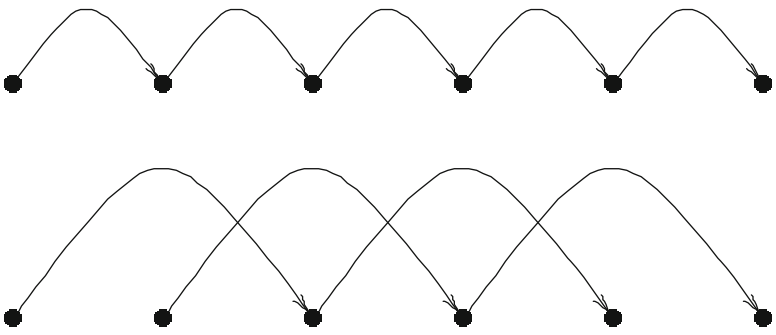

Fig. 10.1 Pair comparisons between sample points on a linear transect at intervals of (*top*) one times the basic spacing and (*bottom*) two times the basic spacing

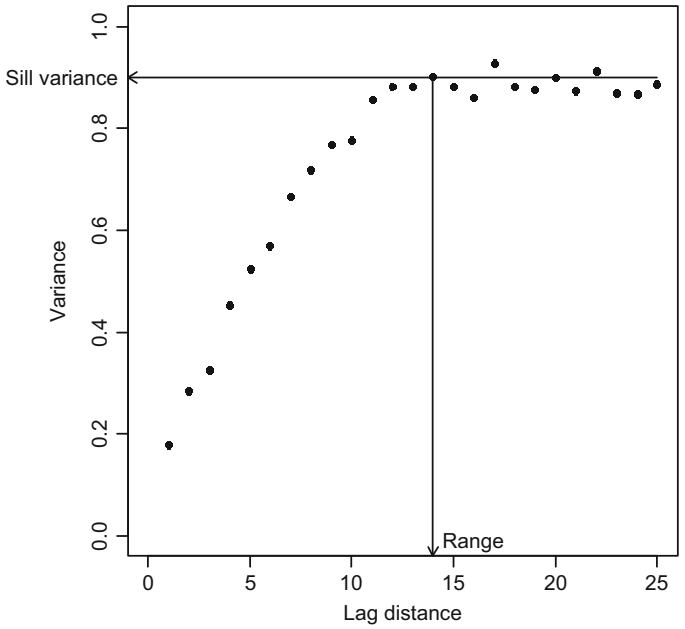

Fig. 10.2 An empirical variogram plot illustrating the concepts of sill variance and range

of observations are separated by twice the sample interval. On a transect the lag is a scalar and is written h. If h is some integer multiple of the basic sampling interval on the transect, then $\gamma(h)$ is estimated by taking all the N_h pairs separated by h and calculating from them half the mean-squared paired difference:

$$\widehat{\gamma}(h) = \frac{1}{2}\sum_{i=1}^{N_h}\{s(x_i) - s(x_i + h)\}^2. \tag{10.4}$$

We call this the empirical, or experimental variogram. Note that we can use the scalar lag in the analysis of data in two or more dimensions if we assume that the spatial autocorrelation of our variable depends on distance only and not direction. This is the assumption of isotropy. We consider isotropic variograms in most of the following discussion and address how to model directional dependence later in the chapter.

Consider Fig. 10.2. This empirical variogram increases with lag distance to a maximum value – the *sill* variance or *a priori* variance of the random function. In this particular figure, this happens at a particular distance – the *range*. If we consider Eq. 10.3, we see that at lags larger than the range, the autocovariance is zero. That is to say two points in space separated by a distance larger than the range are uncorrelated. At shorter distances, they will tend to be correlated. In this way the variogram describes the structure of spatial variability.

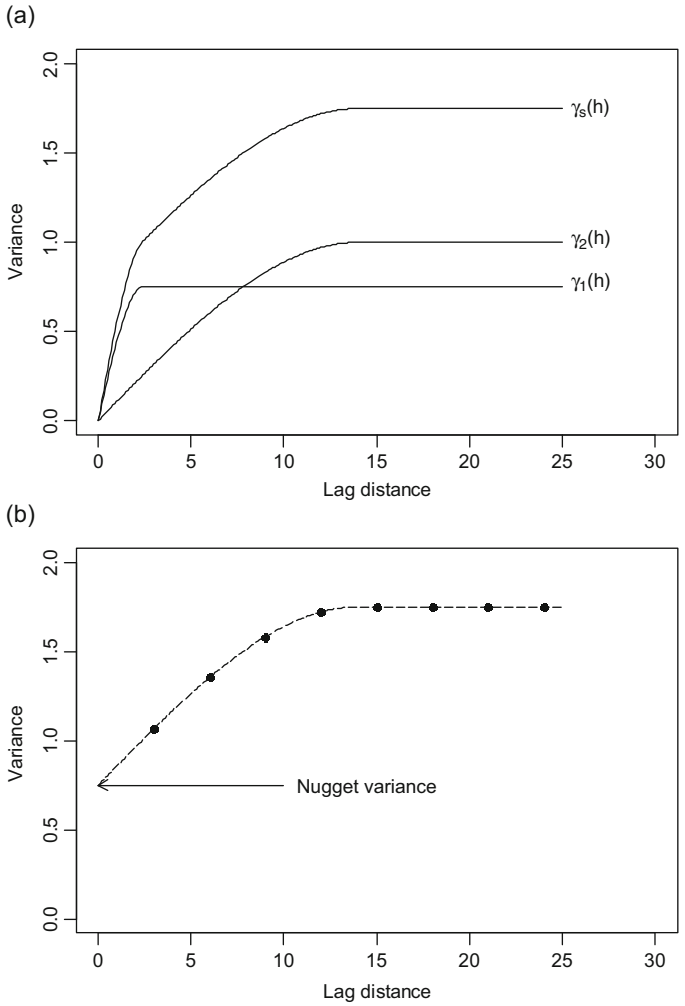

Fig. 10.3 (**a**) Two variograms, $\gamma_1(h)$ and $\gamma_2(h)$ for two mutually independent random variables with contrasting ranges of spatial dependence and the variogram of the sum of these two variables, $\gamma_s(h)$. (**b**) Points from the expected empirical variogram of the sum of the two variables, when sampled at basic interval 3 units. Note that the spatial structure of $\gamma_1(h)$ cannot be resolved, and its variance contributes an apparent intercept to the empirical variogram, the nugget variance

Imagine two isotropic random fields, statistically independent of each other, but with variograms $\gamma_1(h)$ and $\gamma_2(h)$. If we form a new random function by adding these two, then its variogram, $\gamma_s(h)$, is simply the sum, $\gamma_1(h) + \gamma_2(h)$. We describe the random function as a nested random function with a nested variogram. See Fig. 10.3. Note that, while nested variogram functions are commonly used in classical geostatistics and have an intuitive appeal representing combined effects of random processes at contrasting scale, they have been criticised in the setting of model-based geostatistics, a treatment of which is given in Chap. 11. See also Stein (1999).

Consider a soil process which can be regarded as the sum of two independent random functions with different ranges and with variograms $\gamma_1(h)$ and $\gamma_2(h)$. Its variogram, $\gamma_s(h)$, as seen above and as illustrated in Fig. 10.3a is the sum of the two variograms. If we sampled at a basic interval between the ranges of the two variograms, then we may obtain point estimates of the variogram as shown in Fig. 10.3b. The empirical variogram obtained under this sampling scheme resolves the structure of component 2 but not 1. The effect of random function 1 is that the experimental variogram appears to have a non-zero intercept, equal to the sill. This is called the *nugget* variance. Nugget variance will include elements due to measurement error, but as illustrated above, it also includes components of spatial variation which are spatially dependent at scales too fine to be resolved by our sampling scheme. It is possible to partition the nugget between these sources only if we have some independent estimate of the measurement error.

10.2.1.3 Variogram Models

Equation 10.4 above generates point estimates of the variogram for particular lags. In practice we need values of the variogram for any lag. This requires that we can express the variogram as a function of lag. At first glance this might seem like a simple problem in curve fitting. In reality there is a complication. The covariance function allows us to compute the variance of any linear combination of values of a random variable. Now the variance of a linear combination of values of a random function, subject to random variation, must have a variance which is positive and non-zero. We could write down a function of h that does not guarantee this (and so cannot actually be the covariance function of any actual random function). Such a covariance function is said to be non-positive definite. We avoid this by using variograms that are negative semi-definite (negative because the variogram is equal to the a priori variance minus the covariance function and semi-definite because the covariance function may not be zero at all lags). Variogram functions which meet this criterion are said to be authorised. We now describe some commonly used models.

10.2.1.4 Nugget

Imagine a situation where all the spatial dependence for random function is at scales finer than the basic sampling interval. In signal analysis such variation is called 'white noise'. The variogram will be flat, pure nugget. In practice such variograms are rare, but nugget models are almost always included as an additive term in a nested model to describe the variation which has not been resolved by sampling. The nugget variogram model may be written as

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \mathbf{h} = 0 \\ c_0 & \text{otherwise} \end{cases}. \tag{10.5}$$

Note that in practice, the nugget term appears as an intercept (i.e. a value at lag zero), but all variogram models must be zero at lag zero by definition. This is a practical importance since the variogram at lag zero appears in the kriging equation.

The nugget model is a model of discontinuance variation since the variogram jumps from zero to a finite value for any finite lag. This discontinuity can affect predictions of soil properties using a variogram with a significant nugget. When we invoke a nugget model, however, we are not necessarily claiming that the variation of a property is discontinuous at the limit, although it might be. The term nugget effect comes from the case of a nugget of pure gold embedded in rock with a discontinuous step in grade at its edge. All we can say in practice is that variation appears discontinuous at the scale of resolution of our data.

10.2.1.5 The Exponential Variogram

Imagine a linear transect on the ground intercepted at random by boundaries between regions within which the value of the soil property is a uniform value drawn from a random process of variance c_1. If the probability that two points on the transect separated by lag distance h lie either side of at least one such boundary is $p(h)$, then it is clear that the variogram of the process will be

$$\gamma(h) = c_1 p(h). \tag{10.6}$$

Let us assume that the boundaries occur at random as a Poisson process and that the mean interval between two boundaries is a. The mean number of boundaries falling on an interval of the transect of length h is therefore h/a. Under the Poisson distribution, the probability that *no* boundary falls on such an interval is

$$\left\{\left(\frac{h}{a}\right)^0 e^{-h/a}\right\} \Big/ 0! = e^{-h/a}. \tag{10.7}$$

The probability at one or more boundaries falls as such an interval is therefore $1 - e^{-h/a}$, so our variogram function is

$$\gamma(h) = c_1 \left\{1 - e^{-h/a}\right\}. \tag{10.8}$$

This is the exponential variogram for a process of variance c_1 and with distance parameter a. As the derivation shows, it describes the most basic concept of spatial randomness. The exponential function is a bounded one, with an upper bound at c_1, but it approaches this asymptotically so we cannot define a finite range at which $\gamma(h) = c_1$. Since $\gamma(h) \approx 0.95 c_1$ where $h = 3a$, it is common to define $3a$ as the *effective* range. Note that while we may argue from the Poisson boundary process to the exponential variogram, it is hazardous to reverse the argument when such a

variogram is found. Other spatial random processes will give rise to an exponential variogram – for example, a first-order auto-regressive process:

$$s(x_i) = \alpha + \beta\, s(x_{i-1}) + \varepsilon_i \tag{10.9}$$

where x_{i-1}, x_i are successive, equally spaced locations on a transect defines, at the limit as the interval between successive values approaches zero, a random process with an exponential variogram. The terms α and β are coefficients and ε_1 is an independent random variable of mean zero.

10.2.1.6 Bounded Linear Model

We might divide a transect into intervals of equal length, a, and then allocate to all sites within any segment a uniform value drawn from a random process with variance c_1. From Eq. 10.2 we may write the variogram of this process as

$$\gamma(h) = c_1 \left\{ \begin{array}{ll} h/a & \text{if } h \leq a \\ 1 & \text{otherwise} \end{array} \right\}. \tag{10.10}$$

This is a bounded linear variogram of range a. Burrough (1983) discusses this model and fitted it to data where a more or less regular pattern of geological boundaries was a dominant source of soil variation. As with the exponential model, however, this variogram can arise from a contrasting kind of spatial process. Imagine that we generated a random function on a transect with independent random values. This would have a nugget variogram. Let us now smooth this process by replacing each value by the sample average of all the values at locations within $\pm a/2$. Webster and Oliver (2007) demonstrate that such a process will have a bounded linear variogram.

Note that the model can only describe spatial variation in one dimension. It is not negative semi-definite in two or more dimensions.

10.2.1.7 Circular and Spherical Models

We can extend the bounded linear model of spatial variability to two dimensions by imagining a field of independent random variables filtered by replacing each value with the simple average of all values within a distance a. This new field has a circular variogram function:

$$\gamma(h) = c_1 \left\{ \begin{array}{ll} 1 - \frac{2}{\pi}\cos^{-1}\{h/a\} + \frac{2h}{\pi\, a}\sqrt{1 - h^2/a^2} & \text{if } h \leq a \\ 1 & \text{otherwise} \end{array} \right\}. \tag{10.11}$$

This is negative semi-definite in one or two dimensions but not in three dimensions. Like the bounded linear model, it reaches a distinct range at distance a beyond which it is flat, although its slope decreases as it approaches the range. The three-dimensional equivalent, of course, is a spherical model:

$$\gamma(h) = c_1 \left\{ \begin{array}{cc} \frac{3h}{2a} - \frac{1}{2}\{h/a\}^3 & \text{if } h \leq a \\ 1 & \text{otherwise} \end{array} \right\}. \tag{10.12}$$

This is negative semi-definite in one, two or three dimensions. Although it is fundamentally a three-dimensional random process, it is commonly used to describe variograms of two-dimensional data when these have a distinct range. By extending into five dimensions, the pentaspherical model is defined.

So far the variogram models described are fairly straightforward in their behaviour, but they will not always fit observed variograms comfortably. At this juncture we must consider some niceties of the behaviour of variograms of random processes. A more detailed account is given by Webster and Oliver (2007).

Consider a random function $S(x)$. Because this function is random, we cannot compute a derivative $\frac{\mathrm{d}}{\mathrm{d}x}S(x)$, it is not differentiable. For this reason the variogram of a function is not differentiable at the origin, it has an approximate linear approach to the origin. On the other hand two derivatives of a parabolic function can be obtained at the origin. A parabolic variogram would describe smooth variation, i.e. variation that is entirely predictable so not in any sense random. A variogram which approaches this behaviour may arise because of some short-range deterministic variation (*drift*) or perhaps as an artefact arising from a measurement process which has a strong smoothing effect over short distances (yield monitor data, remote sensors and electromagnetic inductance measurements of the soil's electrical conductivity might have such an effect). Some variogram functions may, therefore, appear concave upwards near the origin, suggesting possible local drift. As seen above, parabolic behaviour of the origin is not consistent with random variation. We may define a power function variogram,

$$\gamma(h) = c_1 h^{\alpha}. \tag{10.13}$$

This is subject to the constraint that $0 < \alpha < 2$. If $\alpha = 1$ then the variogram is an unbounded linear function (which unlike the bounded linear function is negative semi-definite in two dimensions). For any authorised power function, the variability appears to increase without bound so the function cannot describe a weakly stationary process, only an intrinsically stationary one.

Some variograms are concave upwards near the origin but are bounded. The Stable model is often used to describe such behaviour:

$$\gamma(h) = c_1 \left\{1 - e^{-\frac{h^{\kappa}}{a^{\kappa}}}\right\}, \tag{10.14}$$

where $\kappa < 2$ is a constant. The Gaussian model is the Stable model with $\kappa = 2$. Although widely used, the behaviour of the Gaussian model at the origin is inconsistent with random variation; it should not be used for the geostatistical methods of prediction which are described below because of this behaviour which can lead to artefacts. A more satisfactory alternative for spatial variation which shows a certain smoothness is the Matérn function (Matérn 1986). The Matérn variogram model is

$$\gamma(h) = c_1 \left\{ \frac{1}{2^{\nu}\Gamma(\nu)} (\varphi\, h)^{\nu} K_{\nu}(\varphi\, h) \right\} \tag{10.15}$$

where $\Gamma(\cdot)$ and $K_{\nu}(\cdot)$ are, respectively, the gamma function and a modified Bessel function (second kind) of order ν. Smoothness of the random process is controlled by parameter ν. If $\nu \geq 0$, then the process is continuous. If $\nu = 0.5$, the Matérn variogram is the exponential. With larger ν it is smoother than the exponential, and as $\nu \to \infty$, the variogram function approaches the Gaussian. The Matérn function has been used increasingly in soil science (Minasny and McBratney 2005) but largely in the setting of model-based geostatistics. It is important that the model be used only when there are adequate observations at short lag distances.

The experimental variogram may appear to vary in a periodic way when the soil variation is controlled in part by some regular pattern such as ridge-and-furrow variation. A sine function is a semi-negative definite model in one dimension. It has gradient zero at lag zero which is unacceptable, but a sine function in combination with another semi-negative definite variogram will constitute an acceptable model. In two dimensions or more, the sine function is not negative semi-definite, but a damped function in which the fluctuations diminish the distance is negative semi-definite. A damped sinusoidal variogram is said to show a 'hole affect'. However, apparent fluctuations may be artefacts. Webster and Oliver (2007) recommend that periodic models are not used unless there is strong evidence for periodic behaviour, perhaps because of our a priori knowledge of the process. Some phenomena in soils can give rise to periodic fluctuations in the variogram, in particular "patterned ground" phenomena as formed in the Gilgai landscapes of Australia or in soils affected at some stage of their development by periglacial conditions.

It follows from our previous discussion of nested processes – additive combination of independent random functions – that two or more of the models discussed here may be combined to describe an experimental variogram. In fact most model variograms are a combination of a nugget variogram and a spatially structured one. But since any combination of negative semi-definite functions is itself negative semi-definite, other combinations are possible. Two models with different distance parameters may be combined to describe a variable with variability with different scales caused by different processes. A double spherical model is a common example of the nested structure with two distinct ranges.

10.2.1.8 Anisotropy

In our examples so far, the lag is a distance in space, i.e. a scalar, and we explained that a random function for which the variogram depends only on the scalar lag distance is said to be isotropic. In practice the direction of a comparison may affect the variance of the difference between two points, so our lags are vectors with both a distance and direction. This is called anisotropy. There are two kinds of anisotropy. In *geometric* or *affine* anisotropy, the sill variances are independent of direction, but the variogram does depend on direction at short distances. It therefore approaches the sill with different slopes at different directions. An affine transformation of the coordinate system will transform the variogram to isotropy. Under the model a locus that is the set of all locations $\mathbf{x}_i \in \mathbf{X}$ about a fixed point, $\mathbf{x}_0$, such that $\gamma(\mathbf{x}_0 - \mathbf{x}_i), \mathbf{x}_i \in \mathbf{X}$ is constant (and less than the sill variance) describes an ellipse, which can be transformed to a circle by the affine transformation of coordinates.

The other kind of anisotropy is *zonal* anisotropy. In zonal anisotropy the sill depends on direction. Such variation is not difficult to imagine. The variation and landscape scale where contrasting rocks outcrop with parallel strike will be largest perpendicular to strike and smallest parallel to the strike, but anisotropy will diminish as the lag tends to zero.

10.2.2 Estimating and Modelling the Variogram

10.2.2.1 Variogram Cloud

The variogram cloud is a plot of the individual values, $\frac{1}{2}\{s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})\}^2$, against the scalar lag, $h = |\mathbf{h}|$. Plotting and examining the variogram cloud can be useful in exploratory spatial analysis of the data. In particular, we may use it to examine evidence for anisotropy or to identify effects of a few outlying observations (Ploner 1999).

10.2.2.2 Lag Classes

In the example above, we illustrated the application of Matheron's (1962) estimator of the variogram with an idealised example of regularly sampled data in one dimension. In practice the problem is more complex, particularly in two dimensions or more when data are irregularly sampled and/or we wish to account for anisotropy. Matheron's estimator is now applied to pair comparisons $\{s(\mathbf{x}) - s(\mathbf{x} + \mathbf{h})\}$ where the lag vector is the mean or central vector of a lag class. Ignoring anisotropy the lag class may be defined by a range of lag distances or lag bin, a central distance plus a tolerance of $\pm w/2$ where w is the width of the bin. We may then take the mean lag distance within each class as the representative lag interval. Defining lag classes requires care and an element of trial and error. If the classes are defined

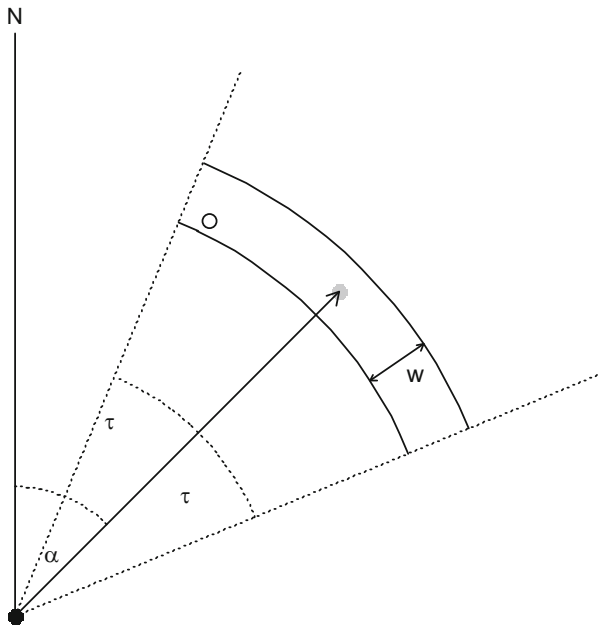

Fig. 10.4 Definition of a lag bin in two dimensions

too narrowly, then each will contain too few pair comparisons, and the resulting variogram estimates will be too noisy. If the classes are too wide, however, then they will smooth the spatial structure which is seen at lags within each interval and may obscure the underlying spatial dependence.

When anisotropy is a concern, lag classes must be defined relative to a compass bearing as well as to a distance. This is illustrated in Fig. 10.4. Consider the two points in space shown by a black and a grey disc. These are separated by a lag vector of scalar value (distance) h along a bearing of α from due north. With irregular sampling it is necessary to define lag bins with some tolerance into which to combine pair comparisons to form point estimates of the variogram. The tolerance is defined by a width of the lag distance bin ($\pm w/2$ in the Figure) and an angular tolerance of $\pm\tau$. With the tolerances specified, the lag between the black and the grey disc is the central lag for a lag bin such that any point within the region between the two solid arcs and the dotted lines indicating the angular tolerance would be separated from the black disc by a lag which belongs in the bin. The open circle illustrates such a point.

10.2.2.3 Estimating Variogram Parameters

The simplest variogram model to fit is an unbounded linear model, which may be fitted with an ordinary least squares criterion. However, most variogram models have non-linear parameters which must be estimated by more complex methods.

Nonetheless, the least squares criterion can be applied; that is to say we find a vector of variogram parameters, $\boldsymbol{\theta}$, such that the mean-squared difference between our point estimates of the variogram, $\widehat{\gamma}(h)$, and the corresponding model values, $\gamma(h|\boldsymbol{\theta})$, is minimised:

$$\text{MSE} = \frac{1}{N_l}\sum_{i=1}^{N_l}\{\widehat{\gamma}(h_i) - \gamma(h_i|\boldsymbol{\theta})\}^2, \tag{10.16}$$

where h_i is the representative lag for the ith lag class of which there are N_l, $\widehat{\gamma}(h)$ is the estimated variogram for this lag class and $\gamma(h|\boldsymbol{\theta})$ is the fitted model. We need an iterative algorithm to minimise this criterion for models with non-linear parameters.

Models can be fitted by eye and there is software available to help this process. However, fitting by eye is generally best avoided particularly where the variogram estimates are very variable and where the number of pair comparisons in each lag class varies so that different lag classes have different uncertainty. Visual inspection can be useful to guide the choice of a class of model and also to check that a fitted model seems reasonable.

Most variogram models are fitted by a weighted least squares method using a standard non-linear fitting algorithm to minimise a generalisation of *MSE*. If the experimental variogram for lag class i is supported by n_i pair comparisons, then one weighted *MSE* is

$$\text{MSE}_1 = \frac{1}{N_l}\sum_{i=1}^{N_l} n_i\{\widehat{\gamma}(h_i) - \gamma(h_i|\boldsymbol{\theta})\}^2, \tag{10.17}$$

so that we give more weight to a point estimate supported by many pair comparisons. Cressie (1985) proposed a development of this:

$$\text{MSE}_2 = \frac{1}{N_l}\sum_{i=1}^{N_l} \frac{n_i}{\gamma(h_i|\boldsymbol{\theta})^2}\{\widehat{\gamma}(h_i) - \gamma(h_i|\boldsymbol{\theta})\}^2. \tag{10.18}$$

This gives greater weight to the lags with small semi-variance, which is reasonable since the reliability of an estimate of variance is inversely related to its size, and the variogram at short lags is generally more influential when used in particular applications. Note that, since the model variogram appears in the weight, and it also depends on the parameters that we are estimating, the process must be done iteratively. The iterative procedure means that after one run of the algorithm, we use the fitted model to specify the modelled values. Generally only one iteration is required before the solution converges to a single set of model parameters.

As we have seen, there are many variogram models which can be fitted to data. Visual inspection of a set of point estimates and consideration of the factors causing the variation will generally enable us to narrow the field to a few plausible models, but we still have to choose between them. One basis for choice is to fit

all possible models and select that one for which the *MSE* is smallest. This is a rational procedure when all models have the same number of parameters (e.g. if we are comparing an exponential with a spherical model). However, we may have a choice between a double spherical model (two independent spherical processes) and an exponential. If we exclude the nugget (common to both), then the double spherical has four parameters to the exponential's two. These additional parameters give the model more flexibility and mean that the *MSE* for the double spherical will almost inevitably be smaller. This does not mean that the model is necessarily better, however. The question is whether the inclusion of more parameters is justified by the improvement of the fit which they allow. The Akaike information criterion (AIC) is a basis for making this judgement (Akaike 1973; McBratney and Webster 1986). For a given set of point estimates for the variogram, AIC will be least for the model for which

$$\widehat{A} = N_l \ln(MSE) + 2P \tag{10.19}$$

is smallest where P is the number of parameters. That model for which $\widehat{A}$ is smallest is judged most parsimonious. This means that an exponential model will be favoured over a double spherical unless the *MSE* for the double spherical, MSE_{DSP}, is so much smaller than that for the exponential, MSE_{Exp}, that

$$\ln \frac{MSE_{\mathrm{DSP}}}{MSE_{\mathrm{Exp}}} < -4N_l. \tag{10.20}$$

10.2.3 Departures from the Standard Model

10.2.3.1 Trends

It is worth recapitulating from Chap. 1 the basic model of soil variation with which we are working. In most general terms $S(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x})$ where $f(\mathbf{x})$ is some deterministic function of space and $\varepsilon(\mathbf{x})$ is a random variable, so $S(\mathbf{x})$ too has random properties. In our discussion so far, we assume that $f(\mathbf{x})$ is a constant, μ, and that $\varepsilon(\mathbf{x})$ is a spatially dependent random variable that can be regarded as a realisation of a stationary (strongly, weakly or intrinsically) random variable or as the sum of two or more independent such variables. The soil variation which we observed in reality is deterministic, the result of many processes, and may be poorly described by such a model. First μ may have to be replaced by a deterministic function $f(\mathbf{x})$. This could represent a pronounced trend, for example, a trend in the particle size distribution along a catena or in water content down a slope. Strong trends cannot be ignored. If $\gamma(\mathbf{h})$ is the variogram of $\varepsilon(\mathbf{x})$, then it is clearly not equal to $\frac{1}{2}E\left[\{(f(\mathbf{x}) + \varepsilon(\mathbf{x})) - (f(\mathbf{x}+\mathbf{h}) + \varepsilon(\mathbf{x}+\mathbf{h}))\}^2\right]$ when $f(\mathbf{x})$ is not a constant.

The experimental variogram will be affected by deterministic variation and becomes concave upwards. We may also find short-range drift or local trend which disappears at longer ranges. In either case we have to disentangle the variation into components which may be treated as deterministic and those which are treated as random. There is no unique way of doing this in any case; it is a matter of finding a particular model which is suitable given the variation and the uses to which we want to put it. We discuss how this might be done later, since it forms an important part of the process of spatial prediction.

10.2.3.2 Non-stationarity in the Random Variation

The term $\varepsilon(\mathbf{x})$ has constant mean by definition. However, in practice it may not be plausible to assume (as in the intrinsic hypothesis or any stronger stationarity assumption) that the variance is uniform. This may originate in many ways. Consider a landscape in which, as we move upslope from the river channel, the parent material changes from sorted alluvium to terrace gravels to clay to sandstone with wet flush zones and then chalk with pockets of overlying clay. It is clear that the variability of a soil property such as the saturated hydraulic conductivity or the water potential will change over the sequence. The development of sensors that generate large data sets, for example, on soil electrical conductivity, will allow us to model such complexity more effectively.

One approach is to compute local variograms within a moving window (Haas 1990; Corstanje et al. 2008). This allows us to account both for changes in the magnitude and spatial scale of variation. As the window moves across the landscape in one part of the landscape, the sill of the variogram might be relatively low (the magnitude of the variation is small); in another part of the landscape, there may be a longer range to the variogram than we see on average (the spatial scale is different). This approach has been incorporated into the Vesper software package discussed in more detail below (Minasny et al. 2006). However, there is greater scope for managing non-stationarity of the variance within the model-based geostatistical framework discussed in Chap. 11.

10.2.3.3 Contaminated Fields

Consider the spatial variation of the concentration of some metal in the soil of a region. This variation may be complex. It will have its origins in the variable composition of the parent material from which the soil is derived: solid rock, drift, loess, gravels and alluvium of varying mineralogical composition. There may also be different sources of pollution through atmospheric deposition or deposition by flood water. These factors considered so far are likely to give rise to a more or less continuous variation of the concentration of the metal in the soil which we might reasonably regard as a realisation of a random function and describe by a variogram. However, other sources of variation are possible. In particular we consider quasi-

point processes, effects of very intensive contamination which have a very localised effect. We say a quasi-point process because the actual area of soil affected by such an event will be of finite dimensions, but at a practical sampling intensity, such a patch will be represented by a maximum of one sample point. A patch, therefore, appears in our sample as a random event of probability η.

Let us assume for the moment, although it is not necessary, that the continuous background process is a normal random variable S, with mean μ_b and standard deviation σ_b. We may also assume that the contaminant process is normal but typically of larger mean μ_b and with standard deviation σ_c. The overall distribution function is, therefore, $(1-\eta)N(\mu_b, \sigma_b) + \eta N(\mu_c, \sigma_c)$. Matheron's (1962) estimator of the variogram is unduly influenced by extreme values since it is based on squared differences. Also, one contaminated value may appear in several comparisons in each of several lag classes. The contaminant process will therefore have a large effect on the estimate of the variogram in the case of such data. Ideally we should like to decompose our data into background and contaminant components. One way of doing this in spatial analysis is to use robust estimators of the variogram which estimate the variation of the background process.

Estimation of the variogram assumes that our $s(\mathbf{x})$ are realisations of an intrinsically stationary random function $S(\mathbf{x})$. Since $E[S(\mathbf{x}_i) - S(\mathbf{x}_i + \mathbf{h})] = 0$, we can estimate $\gamma(\mathbf{h})$ by $\frac{1}{2}\text{Var}[s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})]$ where Var[] denotes the variances of the term in the brackets. Matheron (1962) uses the standard method of moments estimator, the mean square difference. We may estimate the variogram robustly by using an alternative variance estimator, of which several are available. Lark (2000) reviews some of the principal ones, but we consider one in detail for illustration. The mean-squared error is non-robust because it is an arithmetic average. If just one number in a set of data becomes very large, so does the average. Medians, however, are robust. The median value of a set of data is that value such that 50% of the date are smaller and 50% are larger. If there is an even number of data, $2n$, then the median is the average of the nth and the $n+1$th value. The example in the box below shows how substituting one datum with a very large value can dramatically affect the mean, while the median is only slightly affected or not affected at all. For this reason the median absolute deviation has been proposed as a robust measure of variability as an alternative to the mean square error. The median absolute deviation (MAD) is a robust estimator of the variance based on the median absolute difference between each data value considered in turn and the median of the whole data set. For a set of n values of a variable, $s_i, i = 1, 2, \ldots, n$,

$$\text{MAD} = 2.198 \ \text{median}_{i=1}^{n}\left\{\left|s_i - \text{median}_{i=1}^{n}\{s_i\}\right|\right\}^2 \tag{10.21}$$

where $\text{median}_{i=1}^{n}\{\}$ denotes the median value of the n terms in the brackets. The constant 2.198 requires explanation. It is a consistency correction which, on the assumption that s is a normal variable of variance σ^2, ensures that $E[\text{MAD}] = \sigma^2$. Dowd (1984) proposed using median absolute deviation to estimate variograms. His estimate may be written

$$2\widehat{\gamma}_D(\mathbf{h}) = 2.198\{\text{median}(|y_i(\mathbf{h})|)\}^2, \tag{10.22}$$

where $y_i(\mathbf{h}) = s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h}), i = 1, 2, \ldots, N(\mathbf{h})$. Note that, since we are assuming an intrinsically stationary process which is bivariate normal so that $s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})$ $N\{0, 2\gamma(h)\}$, we have the information that the median value $\text{median}_{i=1}{}^n \{s(\mathbf{x}_i) - s(\mathbf{x}_i + \mathbf{h})\}$ is zero, which is implicit in this estimator.

Using this robust estimator entails stronger assumptions than does Matheron's estimator, specifically the bivariate normality of $[s(\mathbf{x}_i), s(\mathbf{x}_i + \mathbf{h})]$. The median absolute deviation is also a less efficient estimator than the mean square error in the statistical sense. All the data apart from the median value itself (or the nth and $n + 1$th in the even case) only influence the median absolute deviation by their relative values and so their order. The information contained in the actual values, which the mean square error uses, is not used. This is a penalty of robustness and is one reason why robust estimators should not be used without specific reason. A useful tool to examine variogram models is cross-validation. In cross-validation we use the ordinary kriging method, described below, to compute a prediction of each observation from all the others. Ordinary kriging returns, along with the prediction $\widehat{S}(\mathbf{x}_i)$ at location $\mathbf{x}_i$, the expected squared error of the kriging prediction, $\sigma^2_{\text{OK}}(\mathbf{x}_i)$, which depends on the variogram. Since we know the observed value at each location, we can compute the standardised square prediction error:

$$\text{SSPE}(x_i) = \frac{\left\{s(x_i) - \widehat{S}(x_i)\right\}}{\sigma^2_{\text{OK}}(x_i)}. \tag{10.23}$$

The expected value of this statistic over a set of data is one, and, assuming normal prediction errors, the expected median is 0.455. Lark (2000) suggested that the median standardised squared prediction error from cross-validation is used to select between variogram models fitted to the standard method of moment estimates and alternatives from robust estimators.

10.3 Geostatistical Spatial Prediction

We may sample the soil at discrete locations and analyse the collected material. In practice many problems that the pedometrician has to tackle can be expressed as how to use this information on a very small volume of soil from a region to make predictions about locations or subregions which have not been sampled. In general this problem emerges in two forms. First we may require an estimated value for the soil variable at a site or over a block, the latter being any region, regular or irregular in two or more dimensions such as a field, or a square panel corresponding to the pixel of a remote sensor image. This predicted value may then be used to make some management decision, for example, how much fertiliser to add. Second, we may be more interested in whether the true value of the point or over a block

exceeds some threshold value. This could be a regulatory threshold, for example. In such a case simple comparison of the predicted value with a threshold value is not entirely adequate since the prediction has attendant uncertainty. We need to have some idea of the risk that the true value exceeds or is less than the threshold so that we know how strong is the evidence that the soil at a location should be treated as contaminated (von Steiger et al. 1996) or salt affected (Wood et al. 1990) or deficient in a nutrient (Lark and Ferguson 2004). In mathematical terms we need to know the probability the true value exceeds the threshold value conditional on our observations. Geostatistics has solutions to these problems, and they have been used by soil scientists since the early 1980s. During this time the variants on the basic kriging equations have diversified. We outline here the key methods which soil scientists have found useful.

10.3.1 Kriging Predictions and their Uncertainty

We wish to obtain an estimate of the value of the soil property at an unsampled location $\mathbf{x}_0$. We call the estimate $\widehat{S}(\mathbf{x}_0)$. Later we extend the problem to the estimation of the mean value of S over a block. We have observations $s(\mathbf{x}_i), i = 1, 2, \ldots n$ from which our estimate is to be obtained. We assume that our $s(\mathbf{x}_i)$ are drawn from a realisation of a random function $S(\mathbf{x})$. At any location this has mean $\mu(\mathbf{x}_i)$.

The kriging estimate is based on familiar regression model. We may write

$$\widehat{S}(\mathbf{x}_0) - \mu(\mathbf{x}_0) = \sum_{i=1}^{n_0} \lambda_i \{S(\mathbf{x}_i) - \mu(\mathbf{x}_i)\}, \tag{10.24}$$

where λ_i is a coefficient or kriging weight. This equation can be rearranged giving

$$\widehat{S}(\mathbf{x}_0) = \sum_{i=1}^{n_0} \lambda_i S(\mathbf{x}_i) + \mu(\mathbf{x}_0) - \sum_{i=1}^{n_0} \lambda_i \mu(\mathbf{x}_i). \tag{10.25}$$

How we proceed from here depends on assumptions which we choose to make about the behaviour of $\mu(\mathbf{x}_i)$.

10.3.2 Ordinary Kriging (OK)

Ordinary kriging is the most widely used kriging method. It proceeds on the assumption that $\mu(\mathbf{x}_i)$ is constant at least for all n_0 sample sites at $\mathbf{x}_i$ within the vicinity of $\mathbf{x}_0$ which we use to estimate $s(\mathbf{x}_0)$. It is also assumed that this mean is unknown to us.

On the basis of these assumptions, we may reduce Eq. 10.25 to

$$\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0) = \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} S(\mathbf{x}_i) + \mu(\boldsymbol{x}_0)\left(1 - \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}}\right). \tag{10.26}$$

Now if we specify that $\sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} = 1$, then the unknown mean is filtered from the ordinary kriging estimator and $\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0) = \sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} S(\mathbf{x}_i)$. This means that $\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0)$ is an unbiased estimate of $\mu(\mathbf{x}_0)$, i.e. on average $\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0)$ is equal to the local mean, under the assumption of a stationary mean within the neighbourhood. We may also define a quantity $\sigma^2\mathrm{OK}(\mathbf{x}_0) = E\left[\left\{\widehat{S}^{\mathrm{OK}}(\mathbf{x}_0) - s(\mathbf{x}_0)\right\}^2\right]$. This is the ordinary kriging variance, the mean-squared error of the kriging estimate. We find the weights λ_i by solving a set of linear equations which minimise this value. The OK estimate is an optimal estimate in the least square sense; that is to say it is a best linear unbiased estimator, given the assumption of a locally constant (but unknown) mean.

The kriging equations are essentially obtained by writing expression for the kriging variance in terms of the covariances of $S(\mathbf{x}_0)$ and all the $S(\mathbf{x}_i)$ and the covariance among all the $S(\mathbf{x}_i)$. These depend on the covariance function $C(\mathbf{h})$ defined in Eq. 10.1. We then write partial derivatives of the kriging variance with respect to the weights λ_i, which are zero where the kriging variance is minimised. See Isaaks and Srivastava (1989) for further details. At first glance the problem may appear simple. We have obtained n_0 unknowns, the weights, λ_i and obtained n_0 equations from setting the n_0 partial derivatives to zero However, this ignores an additional equation, namely, the constraint used to filter out the unknown mean which requires that all the weights sum to one. This $n_0 + 1$th equation requires that we add an additional unknown the Lagrange parameter, $\psi(\mathbf{x}_0)$, which means that the OK equations actually minimise an auxiliary function ${\sigma^2}_{\mathrm{OK}}(\mathbf{x}_0) - 2\psi(\mathbf{x}_0)\left\{\sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} - 1\right\}$. This generates the ordinary kriging equations:

$$\begin{array}{l} \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} C\left(\mathbf{x}_i - \mathbf{x}_j\right) + \psi(\mathbf{x}_0) = C\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} = 1. \end{array} \tag{10.27}$$

At this point the reader might reasonably wonder why we have discussed ordinary kriging in terms of covariance functions. Why the previous focus on the variogram? But remember that $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$. This allows us to rewrite the OK equations as

$$\begin{array}{l} \sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} \left\{C(0) - \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right)\right\} + \psi(\mathbf{x}_0) = C(0) - \gamma\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} {\lambda_i}^{\mathrm{OK}} = 1. \end{array} \tag{10.28}$$

The unbiasedness constraint causes the term $C(0)$ to drop out, so

$$\begin{array}{l} \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right) + \psi\left(\mathbf{x}_0\right) = \gamma\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} \lambda_i^{\mathrm{OK}} = 1. \end{array} . \tag{10.29}$$

We still seem to depend on the assumption of weak stationarity since it is assumed that the covariance function can be defined. In fact OK can be conducted for intrinsically stationary random variables by substituting for $C(0)$ some arbitrary large value C_A which is filtered out. The expression C_A -$C(\mathbf{h})$ is called the pseudo-covariogram. Solving Eq. 10.29 above for the weights allows us to generate the kriging estimator with kriging variance

$$\sigma^2{}_{\mathrm{OK}}\left(\mathbf{x}_0\right) = \sum_{i=1}^{n_0} \lambda_i{}^{\mathrm{OK}} \gamma\left(\mathbf{x}_i - \mathbf{x}_0\right) + \psi\left(\mathbf{x}_0\right). \tag{10.30}$$

OK can be extended to the estimation of a regional mean, i.e. the average value of s over some block R. This can be conveniently expressed as an integral

$$s(R) = \frac{1}{R} \int_{x \in R} s(x) \mathrm{d}x.$$

The interpretation of $s(R)$ may require some care. It is the population mean that we would estimate by design-based sampling of R, measuring s at randomly selected locations in R. For some properties we can think of $s(R)$ as the value we would obtain if all the soil in R were taken and homogenised and then analysed for properties in which the arithmetic average values of a set of discrete samples is equivalent to the aggregate property of the sample. This includes compositional properties such as the clay content or volumetric water content, but not scale-dependent physical properties such as the saturated hydraulic conductivity or properties such as the soil solution concentration of an element which may depend on variable and non-linear exchange processes.

Our block estimate $\widehat{S}(R)$ is found as a linear combination of the data $s(\mathbf{x}_i)$. The weights are found by solving a similar set of equations to those for point estimates.

$$\begin{array}{l} \sum_{i=1}^{n_0} \lambda_{R,i}^{\mathrm{OK}} \gamma\left(\mathbf{x}_i - \mathbf{x}_j\right) - \psi(R) = \overline{\gamma}\left(\mathbf{x}_j - R\right) \forall j \\ \text{and} \\ \sum_{i=1}^{n_0} \lambda_{R,i}^{\mathrm{OK}} = 1. \end{array} . \tag{10.31}$$

These equations are solved for the weights. The block kriging variance is

$$\sigma^2{}_{\mathrm{OK}}(R) = \sum_{i=1}^{n_R} \lambda_{R,i}^{\mathrm{OK}} \overline{\gamma}\left(\mathbf{x}_i, R\right) + \psi\left(R_0\right) - \overline{\gamma}\left(R, R\right). \tag{10.32}$$

The two components $\overline{\gamma}(\mathbf{x}_i, R)$ and $\overline{\gamma}(R, R)$ require explanation. The former represents the mean value of the variogram between $\mathbf{x}_i$ and some point in R. The latter is the mean value of the variogram between points in R. This latter quantity is called the dispersion variance or within-block variance. It is equal to the mean-squared difference form the mean of the values of s drawn at random from within R. These two values can be calculated for regular blocks using specific functions, the auxiliary functions (see Journel and Huijbregts 1978). However, with modern computers they are usually obtained by numerical integration which can also be done efficiently for regions of any shape or size.

10.3.3 Simple Kriging (SK)

We assume that the mean is everywhere constant and known. The kriging weights are calculated to minimise the prediction variance subject to an unbiasedness constraint that $E\left[\widehat{S}(\mathbf{x}_0) - s(\mathbf{x}_0)\right] = 0$. From this we obtain n_0 equations

$$\sum_{i=1}^{n_0} \lambda_i^{\mathrm{SK}} C\left(\mathbf{x}_i - \mathbf{x}_j\right) = C\left(\mathbf{x}_j - \mathbf{x}_0\right) \forall j. \tag{10.33}$$

In SK the value of $C(0)$ is not filtered out in the kriging equation. This means that the variogram function must be bounded, i.e. the random function is assumed to be second-order stationary. This constraint, and the condition that the mean is known, makes SK generally less attractive than OK, and it is much less widely used in soil science. SK is most commonly used in special circumstances where the mean is predetermined. We shall touch on some of these shortly.

10.3.4 Non-linear Kriging

Ordinary kriging and simple kriging, as discussed above, are linear methods which return an estimate of a soil variable at an unsampled site or over a block, the estimate being a linear combination of the data. Information about the soil for a point or block is often needed to make decisions about its management. This management may involve an intervention, for example, the addition of a fertiliser or other amendments to the soil or the removal or remediation of contaminated soil. This intervention entails a cost, but if it is not undertaken where it is actually needed, then a further cost may be incurred through loss of yield or environmental damage which may result in a fine. Regulations or rules are commonly stated in terms of threshold values s_t. These lead to rules:

If $s(R) > s_t$, then action (A), else ($s(R) \leq s_t$) action (B).

If we follow the rule, substituting our estimate $\widehat{S}(R)$ for $s(R)$ above, then errors in the estimate will inevitably result in some incorrect decisions. Should the uncertainty in $\widehat{S}(R)$ affect the decision-making rule above? This depends on the distribution of the estimation errors and the shape of the loss function. By the latter we mean the additional cost entailed as a result of an error (positive or negative) in $\widehat{S}(R)$ when the rule above is applied, the loss being expressed as a function of the error. If the loss function is symmetrical (the cost of an overestimate of s by x units is equal to the cost of an underestimate by the same amount) and the distribution of the errors is symmetrical (e.g. normal), then the best procedure is to follow the rule with the estimate $\widehat{S}(R)$. The only way to reduce the costs incurred as a result of error is to reduce the error variance of $\widehat{S}(R)$ (e.g. by sampling more intensively).

In practice, however, loss functions are usually asymmetric. The fine for leaving a region of land unremediated where remediation was, in fact, the correct decision may exceed the cost of remediation. The yield loss on under-fertilising a region of a field would generally exceed the cost of overfertilising. In these conditions the correct decision in the presence of uncertainty about the true value of $s(R)$ requires that we can quantify the uncertainty, conditional on the observations of the variable that we do have.

There are two general groups of kriging techniques that have come to prominence for tackling this problem, both are non-linear – a conventional linear kriging estimator is applied to the data after these have been transformed non-linearly. The first group are indicator methods (Journel 1983) of which the basic tool is ordinary indicator kriging (IK). Here the non-linear transform is to a binary indicator variable. These techniques have been widely applied by soil scientists (e.g. Meirvenne and Goovaerts 2001; Halvorson et al. 1995). The second group of techniques involve a non-linear transformation of the data to a continuous variable, usually a normally distributed variable. This method is exemplified by disjunctive kriging (Matheron 1976), but multi-Gaussian kriging is a similar method. These techniques have also found widespread use in soil science (e.g. Wood et al. 1990; von Steiger et al. 1996).

Indicator kriging (IK) As previously soil property s at location $\mathbf{x}$ takes value $s(\mathbf{x})$. An indicator transform of $s(\mathbf{x})$ can be defined by

$$\omega_t(\mathbf{x}) = \begin{array}{l} 1 \text{ if } s(\mathbf{x}) \leq s_t \\ 0 \quad \text{otherwise,} \end{array} \tag{10.34}$$

where s_t is a threshold value of the property such as one of the management thresholds mentioned above. In indicator geostatistics, $\omega_t(\mathbf{x})$ is regarded as realisation of the random function $\Omega_t(\mathbf{x})$. It can be seen that

$$\text{Prob}\left[S(\mathbf{x}) \leq s_t\right] = E\left[\Omega\left\{S(\mathbf{x}), s_t\right\}\right] \tag{10.35}$$

where Prob[] denotes, respectively, the probability of the event within the square brackets and $G\{S(\mathbf{x}), s_t\}$ is the cumulative distribution function of $S(\mathbf{x})$ at the threshold value s_t. In indicator kriging we estimate the conditional probability that

$s(\mathbf{x})$ is smaller than or equal to the threshold value s_t, conditional on a set of observations of s at neighbouring sites, by kriging $\omega_t(\mathbf{x})$ from a set of indicator-transformed data.

A set of data on s is transformed to the indicator variable $\omega_t(\mathbf{x})$ using Eq. 10.34 above. The variogram of the underlying random function $\Omega_t(\mathbf{x})$ is then estimated and modelled in the usual way. An estimate of the indicator random function may then be obtained for a location $\mathbf{x}_0$ by kriging from the neighbouring indicator transform data. Ordinary indicator kriging is equivalent to simple kriging of the indicator variable $\Omega_t(\mathbf{x})$ using the mean within the kriging neighbourhood as the expectation. Goovaerts (1997) and Webster and Oliver (2007) give more details. The IK estimate $\Omega_t(\mathbf{x})$ is an estimate of the conditional probability that $s(\mathbf{x}) \leq s_t$ and of the conditional cumulative distribution function (ccdf) $G(S(\mathbf{x}), s_t)$. This direct estimate $\widehat{\Omega}_t(\mathbf{x})$ is not generally used as the conditional probability. Instead it is recommended to obtain estimates $\widehat{\Omega}_t(\mathbf{x})$for several s_t that include the threshold of practical importance. There is no guarantee that the estimates $\widehat{\Omega}_t(\mathbf{x})$ will meet the order relation constraint for a cumulative distribution function, that is, $\widehat{\Omega}_{t1}(\mathbf{x}) \leq \widehat{\Omega}_{t2}(\mathbf{x}) \leq \widehat{\Omega}_{t2}(\mathbf{x})$ for any $s_{t1} < s_{t2} < s_{t3}$. Because of this the original set of estimates $\widehat{\Omega}_{t1}(\mathbf{x}), \widehat{\Omega}_{t2}(\mathbf{x}) \widehat{\Omega}_{t2}(\mathbf{x})$ must be smoothed to give a set of revised estimates $\tilde{\Omega}_{t1}(\mathbf{x}), \tilde{\Omega}_{t2}(\mathbf{x}) \tilde{\Omega}_{t2}(\mathbf{x})$ such that the order relations hold. Deutsch and Journel (1998) describe methods for doing this. The estimate of the conditional probability that $s(\mathbf{x}) \leq s_t$ is now given by $\tilde{\Omega}_t(\mathbf{x})$. It is likely that $\tilde{\Omega}_t(\mathbf{x})$ is a better estimate of the conditional probability for a cut-off than is $\widehat{\Omega}_t(\mathbf{x})$ since it incorporates information from adjacent parts of the ccdf. Variants of ordinary indicator kriging are available. In particular we can cokrige the indicator variables for the different thresholds (see the later discussion of cokriging) although Goovaerts (1994) found no substantial benefit from doing so.

Disjunctive kriging (DK) Disjunctive kriging is an alternative to indicator kriging. It is based on the assumption that our data are a realisation of a process with a second-order stationary distribution. Further we assume that the underlying process is a diffusion process (Rivoirard 1994; Webster and Oliver 2007). That is, it varies continuously so that if the variable takes values s_1 and s_2 at locations $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, then all intervening values between s_1 and s_2 must occur at locations on a straight line between $\mathbf{x}_1$ and $\mathbf{x}_2$. The commonest model which we use for DK is the normal diffusion process. Webster and Oliver (2007) explain how the plausibility of this assumption is tested. Since data may often not resemble a normal random variable, the first step in DK is to apply a non-linear transform to the data to achieve normality. This is done using hermite polynomials, a procedure described in more detail by Rivoirard (1994). A variogram is estimated for the hermite transform data in the usual way and modelled. The hermite polynomials are then kriged to target locations of interest. From these may then be estimated the original soil variable $s(\mathbf{x})$ and the conditional probability for specified thresholds, $\widehat{\Omega}_t^{\mathrm{DK}}(\mathbf{x})$.

The relative merits of IK and DK may be summarised as follows. IK may be implemented using widely available kriging software. Since indicators may be estimated by simple kriging or ordinary kriging, we need not assume that the mean of the indicator variable is known. The DK on the other hand is restricted to simple

kriging. The indicator transform of the continuous soil variable discards a good deal of information, and it is tedious to estimate model indicator variograms for a large number of threshold values. The fact that IK estimates of the ccdf do not necessarily conform to the order relations and generally require an arbitrary correction to meet this constraint is a further disadvantage.

DK requires only one variogram model for the hermite transform data. Indicator variograms and cross-variograms must also be estimated to test the assumption of a diffusion process, but these need not be modelled. The hermite transformation transforms a continuous variable monotonically to another continuous variable, so, unlike the indicator transform, there is no loss of information. Further DK returns an estimate of $s(\mathbf{x})$ along with the conditional probabilities. An estimate of $s(\mathbf{x})$ is not generated by IK without additional effort. On the other hand, DK requires stricter assumptions of stationarity than does IK. It can only be implemented if the hermite transform data have a bounded variogram. DK is a complex procedure, and friendly software is not widely available, although the related procedure of multi-normal kriging is available in the GSLIB package (Deutsch and Journel 1998).

In practice we make the choice between IK and DK on practical considerations such as some of those above. In general we might expect that DK or other non-linear kriging techniques that use a continuous transform of the original data will perform better than IK since the indicator transform will inevitably lose some information, while a continuous transform retains all the information in the original data. There have been few studies to compare IK and DK on real data. Those of which we are aware (Papritz and Dubois 1999; Lark and Ferguson 2004) suggest that there may be very little difference between the results obtained with the two methods.

10.3.5 Kriging with a Nonstationary Mean

Again we return to our basic model of soil variations $s(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x})$. So far we have considered $f(\mathbf{x})$ to be a constant value, the mean. In practice it may be necessary for $f(\mathbf{x})$ to express sources of variation in the soil which are not constant and cannot be regarded as any kind of random variable. It is not difficult to think of examples where the variation of a soil property is systematically linked to location and space. Consider the familiar figures in pedological textbooks of the depth of weathered material and how it changes from a crest down a slope to the foot slope and toe slope. Consider again the systematic variation of soil texture, mineralogy, redox potential and organic carbon content associated with the familiar catenas of Central Africa. In all these cases, $f(\mathbf{x})$ must express some deterministic relationship between location in space and a soil variable.

Generally we recognise a distinction between two kinds of systematic variation although it is more a distinction of degree and spatial scale than a fundamental difference. The first kind of systematic variation includes long-range trends where the component represented by $f(\mathbf{x})$ is a broad variation from large to small values of a variable between different parts of a landscape. The second kind of systematic

variation is where short-range differences are dominated by what appear to be deterministic processes. This results in the parabolic form of the experimental variogram that we discussed above. It should be noted that variation that we treat as a trend within field scale might be reasonably attributed to random variation if we are sampling a catchment. This underlines the fact that our model of variation is in part a consequence of the scale and intensity of our sampling.

Various approaches have been taken to the problem of kriging in the presence of a nonstationary mean; these include universal kriging, regression kriging and the method of intrinsic random functions of order k. In universal kriging (UK) (e.g. Webster and Burgess 1980), we express $f(\mathbf{x})$ as some function linear in polynomials of the elements of $\mathbf{x}$ (i.e. trend surfaces), and the kriging estimator implicitly estimates the coefficients of these terms along with a minimum variance estimate of the random term $\varepsilon(\mathbf{x})$. This requires that we have a variogram of $\varepsilon(\mathbf{x})$. Regression kriging (RK) is formally equivalent to UK, but we start by finding an ordinary least squares estimate of the coefficients of the trend function and then estimate the variogram from the residuals (e.g. Odeh et al. 1994). We may then use the variogram to compute a weighted least squares estimate of the coefficients and iterate this process to convergence. The RK prediction of $s(\mathbf{x}_0)$ is then computed by first calculating the value of the trend at $\mathbf{x}_0$ and then estimating the random component by simple kriging (with a known mean of zero) from the residuals from the trend surface at the observation sites. RK and UK are formally equivalent, given the variogram of $\varepsilon(\mathbf{x})$, but it is this variogram that is the problem for both methods. We saw above that the experimental variogram of a variable with a trend will be affected by this trend, and what we need is a variogram of the random component only. This might be obtained for UK by only estimating the variogram from pair comparisons that are not strongly affected by the trend (e.g. over lags perpendicular to the direction of the trend), but this is not always possible, and neither is it efficient. The variogram obtained from residuals from a trend surface is biased (Cressie 1993), so the RK solution is not satisfactory either, although it is reasonably robust when applied to large data sets.

Intrinsic random functions of order k (IRFk) are generalised increments of our data which filter out the trend. The concept is easily grasped if we consider a linear trend in one dimension. It is clear that the differences between adjacent and regularly sampled observations of a variable with a linear trend and an intrinsically stationary random function will be intrinsically stationary since the trend component is replaced by its first derivative. If the trend is of higher order, then further differencing will remove it (and the lower-order components become zero). The method of IRFk works in this way, and a key stage is the determination of the order of increments that can be regarded as intrinsically stationary. This is done by examining generalised covariance functions, in effect variograms of differenced data. Chiles and Delfiner (1999) provide more detail, and Buttafuoco and Castrignano (2005) give an example of the application of this approach in soil science.

In RK and UK we use an unbiased linear model of the trend for prediction in combination with an unbiased minimum variance estimate of the random component of variation. UK does this simultaneously and RK separates the two stages. Given a variogram for the random component, the predictors are the best linear unbiased

predictor (BLUP) or empirical BLUP (E-BLUP), empirical in that it is conditional on a variance model derived from data. We noted above that the key problem is obtaining this variance model, a variogram of $\varepsilon(\mathbf{x})$. The state-of-the-art approach to this problem is to use residual maximum likelihood (REML), and it is the REML–E-BLUP that we advocate for spatial prediction in the presence of a trend. This, however, is out with the scope of the current chapter and is discussed in Chap. 11.

10.3.6 Sampling for Estimation by Kriging

McBratney et al. (1981) and Burgess et al. (1981) showed how a geostatistical survey can be designed to ensure that the variable may be estimated with a prespecified precision. They considered regular grids and assumed that the maximum kriging variance occurred at the centre of each regular grid cell. Thus, they ignored the increased kriging variances close to the boundaries of the region. They sought the maximum interval between observations on the regular grid such that the kriging variance was less than the prespecified tolerance, σ^2_{T}. Thus, for variogram parameter vector $\boldsymbol{\theta}$, they calculated the optimal interval, $I(\boldsymbol{\theta})$, such that

$$I(\boldsymbol{\theta}) = \max\left[i, \text{such that } \max\left\{\sigma_K^2\left[\boldsymbol{\theta}; \mathbf{x}_g(i)\right]\right\} < \sigma_T^2\right], \tag{10.36}$$

where $\sigma^2{}_K|\boldsymbol{\theta}, \mathbf{x}_g(i)$ denotes the kriging variance at the centre of a cell within a square grid of interval i. The same approach can be followed when using cokriging (McBratney and Webster 1983) and when planning composite sampling (Webster and Burgess 1984).

It is possible (Marchant and Lark 2007) to write an expression for the OK prediction error variance that accounts both for the distance to neighbouring observations and the error in the estimated variogram parameters. It is possible to minimise the average value of this statistic across a proposed sample region by modifying the location of sample points. Brus and Heuvelink (2007) show how the approach can be extended to universal kriging.

The problem here is that in most cases, the variogram is unknown when sampling is designed. We may use approximate variograms, as discussed in Chap. 10. An alternative is the adaptive approach proposed by Marchant and Lark (2006). This is appropriate for circumstances where the sampling can be done in phases (ideally when the data can be collected in real time with a sensor). For example, in their Bayesian adaptive approach, spatial simulated annealing is used to minimise an objective function which is the mean square error of the kriging variance at the centre of a notional grid cell. This objective function is obtained as a Bayesian integration over the space of possible variogram parameters with a probability density function for these parameters. The sampling proceeds in phases, and the pdf of the variogram parameters is updated as new data are collected. At the end of each sampling phase, we consider the cumulative distribution function of grid spacings which are sufficient to ensure some maximum kriging variance. When the uncertainty on this interval is sufficiently small, the sampling is completed

by finding those locations needed to complete a sample grid of adequate intensity across the study region.

10.3.7 Kriging for Large Data Sets

Sensors such as gamma-ray spectrometer and electromagnetic induction instruments attached to a vehicle with a GPS have been used to collect intensive soil data for high-resolution mapping. These on-the-go proximal soil sensors can collect a large volume of data over an area (approximately 5,000–100,000 data points per km^2). Data collected from these sensors need to be interpolated to a regular grid.

Kriging of such data poses several shortcomings: the time taken to calculate an empirical variogram can be excessive; in addition, solving the kriging equations for a data set of size n involves inversion of an $(n \times n)$ covariance matrix, which requires $O(n^3)$ operation. A way to circumvent this is to use the spherical model (a finite range variogram), where a sparse matrix can be approximated for the variance–covariance matrix (Barry and Kelley Pace 1997). Others called for covariance matrix tapering, where covariances at large distances are set to zero (Kaufman et al. 2008). Kriging using a single variogram model for the whole area usually resulted in a smooth map, where local variation captured by data can be lost.

Another solution is kriging that takes into account the local spatial structure (Haas 1990). This is implemented as kriging with local variograms, also known as kriging and automated variogram modelling within a moving window. It involves searching for the closest neighbourhood for each prediction site, estimating the empirical variogram from the neighbourhood, fitting a variogram model to the data automatically by a non-linear least squares approach, kriging with local neighbourhood and variogram parameters and calculating the uncertainty of kriging prediction. All these steps need to be done automatically, and thus the program adapts itself spatially in the presence of distinct differences in local structure over the whole field. Local variogram estimation and kriging can preserve the local spatial variation in the predictions. In most cases, local variograms could circumvent the problems of anisotropy and the need for trend analysis.

Minasny et al. (2006) developed a program called Vesper (Variogram Estimation and Spatial Prediction plus Error), a PC-Windows software program that can calculate and model global local variograms and do global and local kriging in either punctual or block form. Sun et al. (2012) extended this approach to local regression kriging to take into account both the local relationship between the covariates and soil observations and the spatial variance of the residuals.

Cressie and Kang (2010) regarded this local kriging method as an ad hoc solution and suggested the fixed ranked kriging (FRK) approach. FRK uses covariance functions that are flexible through a set of r basis functions, where the $(n \times n)$ variance–covariance matrix can be approximated by $(r \times r)$ positive-definite matrices (Cressie and Johannesson 2008). Cressie and Kang (2010) demonstrated the application of FRK on proximally-sensed gamma counts ($n = 34{,}266$) in a field of 2.66 km^2. They used a flexible, nonstationary spatial covariances represented as

77 basis functions, for which exact kriging can be carried out. Nevertheless, the FRK approach still needs tuning with respect to the type of basis functions and estimation of the parameters of those functions.

10.4 A Case Study

We now illustrate some of the key concepts introduced in previous sections with a case study. This uses data on the concentration of copper in the topsoil of part of the east of Scotland. The data were first described by McBratney et al. (1982). Figure 10.5 shows histograms of the raw data and of the data transformed to natural logarithms, and Table 10.1 presents summary statistics of these two variables.

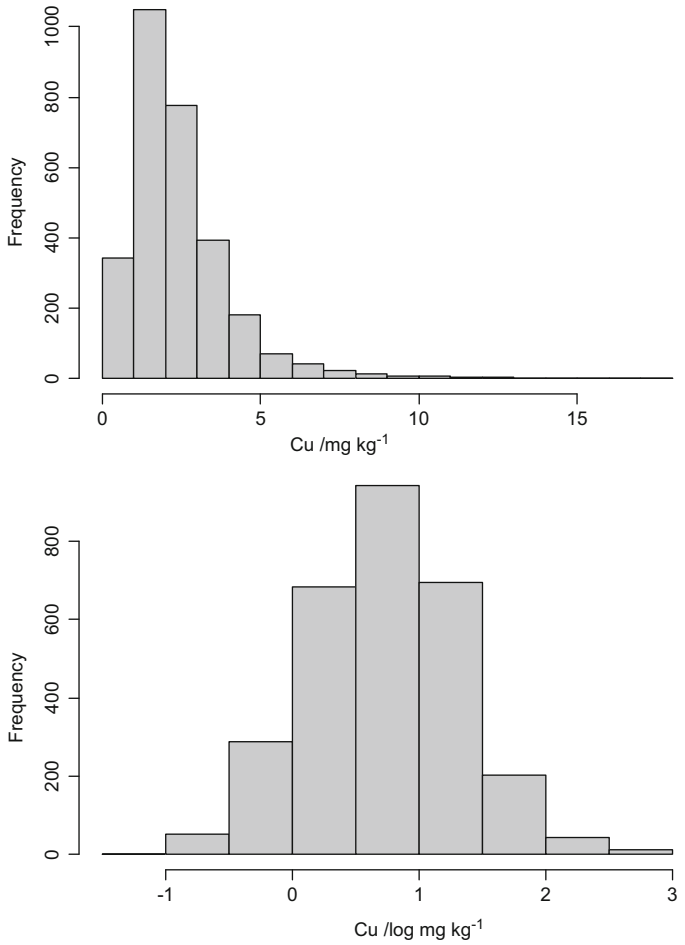

Fig. 10.5 Histograms of topsoil copper content in soils from the east of Scotland on (*top*) original and (*bottom*) natural log scales

Table 10.1 Summary statistics of data on topsoil copper content from the east of Scotland

Variable	Mean	Median	Standard deviation	Minimum	Maximum	Skewness
Copper/mg kg^{-1}	2.48	2.10	1.68	0.3	18	2.66
Copper/$\log_e$ mg kg^{-1}	0.73	0.74	0.59	−1.2	2.89	0.09

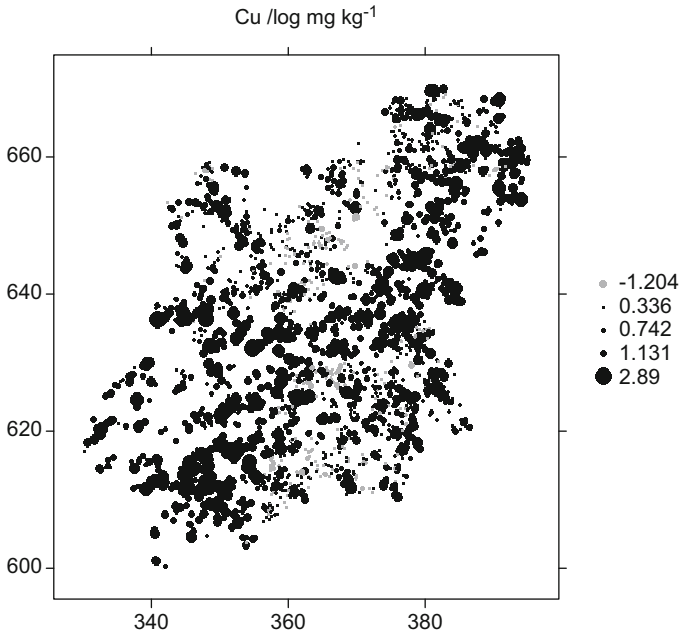

Fig. 10.6 Quintiles of soil copper content (legend is on log scale) shown on a post-plot of the copper data. Coordinates are kilometres relative to the origin of the British National Grid

The histograms and the summary statistics show that the data appear symmetrically distributed on a log scale, so this is used for further analysis. Figure 10.6 shows the distribution of the observations.

Figure 10.7 shows the empirical variogram of the log-transformed data, estimated for four different directions. There is no evidence of systematic anisotropy, so isotropic variograms were estimated and modelled. Matheron's estimator (Eq. 10.4) and the robust estimator of Dowd, described in Eq. 10.22, were both used, and the resulting empirical variograms were fitted with double spherical variogram models, which were preferred to simpler ones on the grounds of Akaike information criterion. Both models were cross-validated, and the standardised squared prediction error, Eq. 10.23, was computed. The estimates, models and values of *SSPE* (see Eq. 10.23) are shown in Fig. 10.8. On the basis of the cross-validation, the model fitted to estimates from Dowd's estimator was chosen. Figure 10.9 shows the ordinary kriging estimates of soil copper content (log scale) across the study area and the kriging variance. As expected the variance is largest where the data are most sparse.

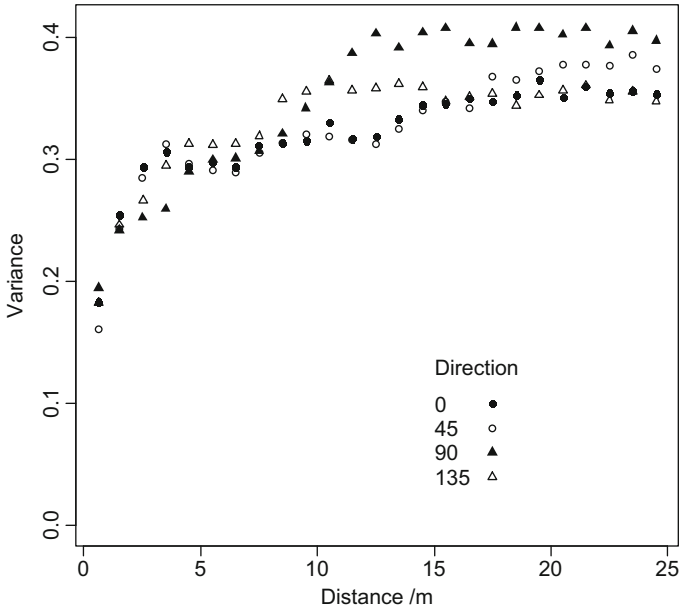

Fig. 10.7 Directional empirical variograms of log copper content

Figure 10.10 shows the ordinary kriging variance for (log) soil copper content, computed at the centre of a cell of a notional square sampling grid with spacing ranging from 500 m to 5 km. This graph, following McBratney et al. (1981), allows one to identify the grid spacing which would be required in a survey in comparable conditions if one wanted to achieve a kriging variance no larger than 0.2. The graph shows that a grid spacing of no more than 1.5 km is required to achieve this target.

10.5 Spatial Covariation and Coprediction

The methods we have discussed so far are all univariate. They consider just one variable and its variation in space. The geostatistical model of spatial variation can be readily extended to two or more variables. Why should we wish to do this? First, because there are cases where we are interested both in estimates of variables and in estimates of some linear function of the variables. An example from soil science quoted by Webster and Oliver (2007) is where one variable, $s_1(\mathbf{x})$, is the depth to the top of a particular soil horizon and the second, $s_2(\mathbf{x})$, is the depth to the bottom of the horizon. A linear combination of these variables is their difference, the thickness of the horizon. If $\widehat{S}_1(\mathbf{x})$ and $\widehat{S}_2(\mathbf{x})$ are estimates of the two variables at $\mathbf{x}$, then one estimate of the horizon thickness at $\mathbf{x}$ is $\widehat{S}_1(\mathbf{x}) - \widehat{S}_2(\mathbf{x})$. However, we could also krige the difference variables $\{s_1(\mathbf{x}) - s_2(\mathbf{x})\}$ directly from the differences at

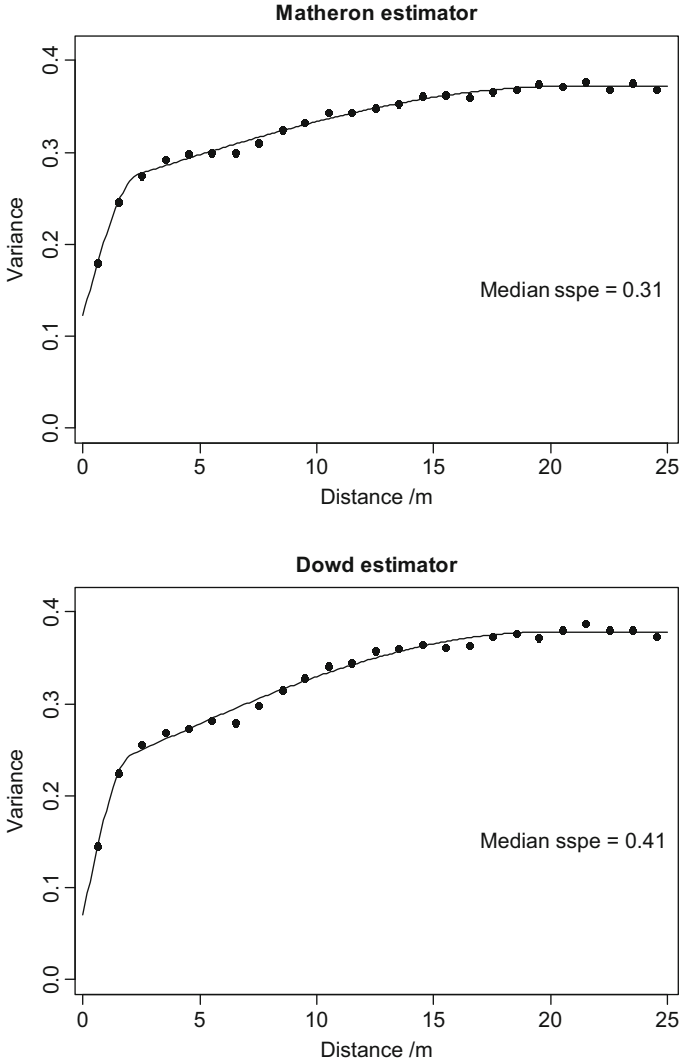

Fig. 10.8 Estimated isotropic variograms of soil copper content with fitted double spherical models and median standardised squared prediction error using Matheron's or Dowd's variogram estimators

our observation sites. There is no guarantee that the two approaches will yield the same estimate if we use the univariate kriging equations described above. That is to say, the kriging estimate of $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$ is not guaranteed to be equivalent to the kriging estimate of any linear combination of $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$. The kriging estimates are not coherent. For some applications this may be important.

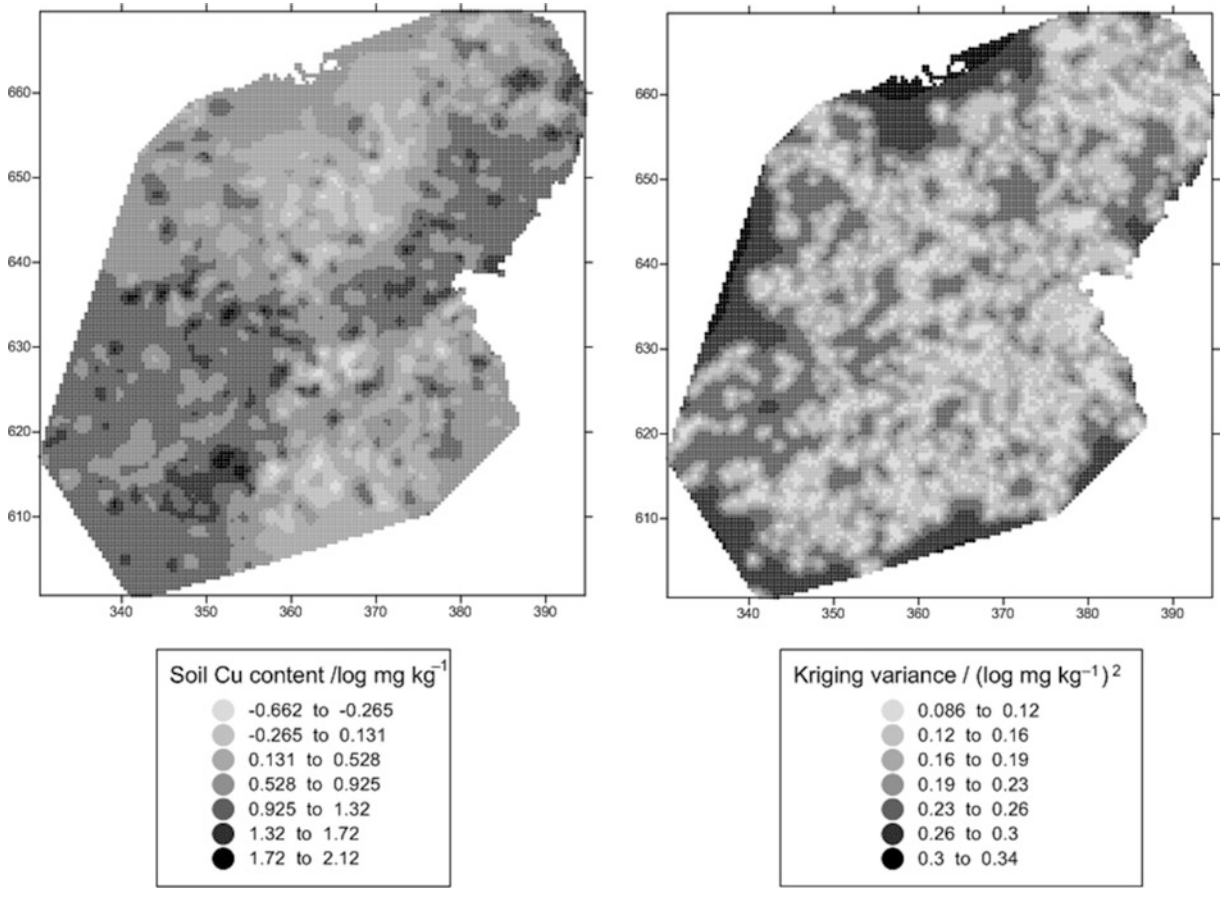

Fig. 10.9 Ordinary kriging prediction of topsoil copper content (log scale) and ordinary kriging variance across the east of Scotland

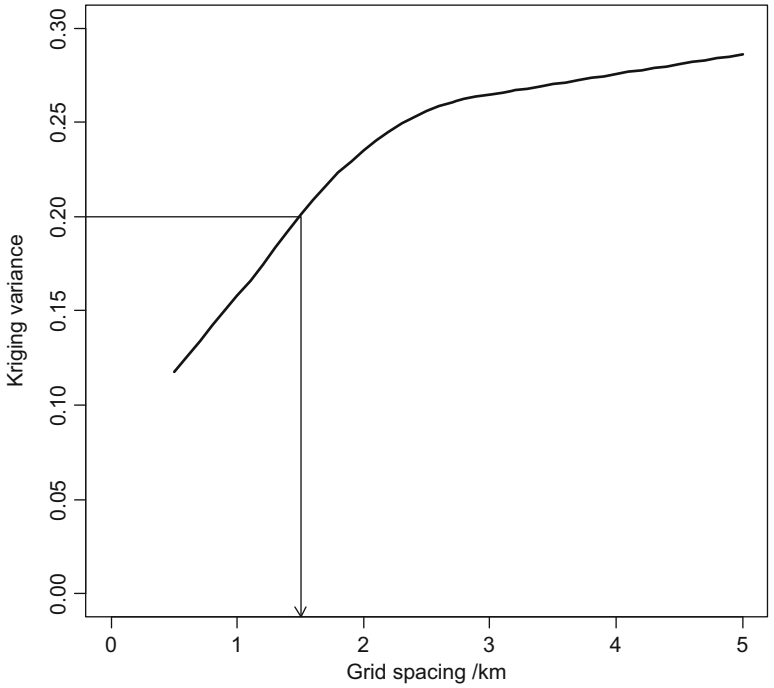

Fig. 10.10 Ordinary kriging variance at the centre of a cell of square grids of different spacing, based on the variogram for topsoil copper content in the east of Scotland

The second reason for considering multivariate statistical methods is probably the commonest. In many circumstances it may be possible to supplement a set of relatively costly direct measurements of a soil variable with a denser but cheaper set of measurements of a second correlated variable. For example, remote sensor measurements of the earth's surface may be correlated with a soil variable of interest. If this can be exploited through a geostatistical method, then we may be able to obtain better predictions of the soil variable by incorporating the cheapest second variable, without added costly measures.

10.5.1 Spatial Co-regionalisation

Let $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$ denote measurements of two soil variables at location $\mathbf{x}$. We assume that these are realisations of, respectively, random functions $s_1(\mathbf{x})$ and $s_2(\mathbf{x})$. It is assumed that these are both intrinsically stationary, and there exists a cross-variogram $\gamma_{2,1}(\mathbf{h})$ dependent only on $\mathbf{h}$ where

$$\gamma_{2,1}(\mathbf{h}) = E\left[\{S_1(\mathbf{x}_i) - S_1(\mathbf{x}_i + \mathbf{h})\}\{S_2(\mathbf{x}_i) - S_2(\mathbf{x}_i + \mathbf{h})\}\right]. \tag{10.37}$$

The cross-variogram can take negative values unlike the auto-variograms $\gamma_{1,1}(\mathbf{h})$ and $\gamma_{2,2}(\mathbf{h})$ which are positive by definition. In fact the ratio

$$\frac{\gamma_{2,1}(\mathbf{h})}{\sqrt{\gamma_{1,1}(\mathbf{h})\,\gamma_{2,2}(\mathbf{h})}} \tag{10.38}$$

is known as the co-dispersion coefficient and measures the correlation of variables $S_1(\mathbf{x})$ and $S_2(\mathbf{x})$ at lag $\mathbf{h}$ which may be positive (large values of $S_1(\mathbf{x})$ are associated with large values of $S_2(\mathbf{x})$) or negative (small values of $S_1(\mathbf{x})$ are associated with large values of $S_2(\mathbf{x})$).

The cross-variogram may be estimated in much the same way as the ordinary variogram for a single variogram, sometimes called the auto-variogram. If we define a lag class centred at scalar lag h with N_h pair differences, then an estimate of the cross-variogram for the lag class is

$$\widehat{\gamma}_{2,1}(h) = \frac{1}{2}\sum_{i=1}^{N_h}\{s_2(\mathbf{x}_i) - s_2(\boldsymbol{x}_i + h)\}\{s_1(\mathbf{x}_i) - s_1(\mathbf{x}_i + h)\}. \tag{10.39}$$

This estimator is non-robust, like the comparable auto-variogram estimator we have already discussed (of which it is a generalisation). Lark (2003) has proposed and demonstrated robust estimators of the cross-variogram which should be used with either s_1 or s_2 or both s_1 and s_2 containing outlying values.

It is notable that the cross-variogram estimator requires that we have paired comparisons of both variables made at the same location. There are circumstances in which few or none of our measurements on different variables are made at the same location. For example, Papritz and Webster (1995) point out that when monitoring change of soil properties over time (with s_1 equal to the variable at time 1 and s_2 the variable at time 2), there can be such problems since at the limit, we cannot measure the same soil twice by destructive sampling and analysis. In these circumstances we may estimate the pseudo-cross-variogram introduced by Clark et al. (1989) defined by Myers (1991) as

$$\widehat{\gamma}^{P}_{2,1}(h) = \mathrm{Var}\left[s_2(\mathbf{x}_i) - s_1(\mathbf{x}_i + h)\right], \tag{10.40}$$

where Var[] denotes the variance of the term in brackets. While the pseudo-cross-variogram allows us to extract information on spatial covariation where variables are not measured at the same locations, it does require more restrictive assumptions, and in practice these amount to the assumption of weak rather than only intrinsic stationarity.

In order to use the cross-variogram, we must model it along with the corresponding auto-variograms. In the multivariate case, it is necessary but not sufficient to use authorised variogram functions; since a model co-regionalisation comprising authorised functions fitted separately to the cross, the auto-variograms is not necessarily positive definite. The most widely recommended strategy to ensure a positive-definite model of co-regionalisation is to use a linear model of co-

regionalisation (LMCR). In the LMCR it is assumed that all the constituent random functions are linear combinations of a common set of independent random functions of mean zero and unit variance, $Y_k^l(\mathbf{x})$ (l is an index not a power). Thus

$$S_u(\mathbf{x}) = \mu_u + \sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{u,k}^l Y_k^l(\mathbf{x}), \tag{10.41}$$

where μ_u is the mean of the random function. The coefficients $a_{u,k}^l$ are specific to the random functions $S_u(\mathbf{x})$ and $Y_k^l(\mathbf{x})$. As stated above the random functions $Y_k^l(\mathbf{x})$ are mutually independent, but if two such functions have a common index l, then they have the same spatial correlation structure. There are n_l such structures with variogram functions $g_l(\mathbf{h})$. It can be shown that the cross-variograms of any two of the correlated random variables S_u and S_v can be expressed as a linear combination of the $L+1$ basic variogram functions:

$$\gamma_{\mathrm{uv}}(\mathbf{h}) = \sum_{l=0}^{L} b_{u,v}^l g_l(\mathbf{h}), \tag{10.42}$$

where

$$b_{u,v}^l = \sum_{k=1}^{n_l} a_{u,k}^l a_{v,k}^l.$$

Thus the cross-variogram matrix for N_w variables,

$$\mathbf{G}(\mathbf{h}) \equiv \begin{bmatrix} \gamma_{1,1}(\mathbf{h}) & \gamma_{1,2}(\mathbf{h}) & \cdots & \gamma_{1,N_w}(\mathbf{h}) \\ \gamma_{2,1}(\mathbf{h}) & \cdot & \cdots & \gamma_{2,N_w}(\mathbf{h}) \\ \cdot & \cdot & \cdots & \cdot \\ & & & \cdot \\ \gamma_{N_w,1}(\mathbf{h}) & \gamma_{N_w,2}(\mathbf{h}) & \cdots & \gamma_{N_w,N_w}(\mathbf{h}) \end{bmatrix},$$

can be written as

$$\mathbf{G}(\mathbf{h}) = \sum_{l=0}^{L} g_l(\mathbf{h}) \mathbf{S}_l, \tag{10.43}$$

where

$$S_l \equiv \begin{bmatrix} b^l{}_{1,1} & b^l{}_{1,2} & \cdots & b^l{}_{1,N_w} \\ b^l{}_{2,1} & \cdot & \cdots & b^l{}_{2,N_w} \\ \cdot & \cdot & \cdots & \cdot \\ & & & \cdot \\ b^l{}_{N_w,1} & b^l{}_{N_w,2} & \cdots & b^l{}_{N_w,N_w} \end{bmatrix}.$$

If the $L + 1$ matrices $\mathbf{S}_l$, co-regionalisation matrices, are all positive semi-definite and the $L + 1$ basic variogram functions are all authorised, then the LMCR has a positive-definite covariance structure.

To fit an LMCR, we must find a set of variogram functions in the corresponding co-regionalisation matrices that optimise a measure of the fit of $\mathbf{G}(\mathbf{h})$ to the point estimates of the variograms, subject to the constraints on the co-regionalisation matrices which we describe above. This is not trivial, particularly if the number of variables is increased. Goovaerts (1997) describes how an LMCR may be fitted by visual assessment and trial and error. A semiautomated procedure was devised by Goulard and Voltz (1992). If the distance parameters of the basic variogram functions are given, then this algorithm will find estimates of the corresponding co-regionalisation matrices optimising the measure of the goodness of fit and meeting the constraints on the co-regionalisation matrices. Kunsch et al.(1997) describe a way of fitting the LMCR using a non-linear regression method to fit the distance parameter and Goulard and Voltz's (1992) algorithm to fit the co-regionalisation matrices. Lark and Papritz (2003) show how the model can be fitted subject to these constraints using simulated annealing. Fitting the LMCR with estimates of the pseudo-cross-variogram is more complex but can be done under certain assumptions discussed by Papritz et al. (1993) and by Lark (2002).

10.5.2 Cokriging

When we have an LMCR for two variables, then we can proceed to spatial prediction by cokriging. The cokriging estimator is a linear predictor, like the kriging estimator we have already discussed. If soil variable s_1 is to be estimated at location $\mathbf{x}_0$ from observations of $s_1, s_2, \ldots, s_m$ at neighbouring locations, then the general linear predictor is

$$\widehat{s}_1(\mathbf{x}_0) - \mu_1(\mathbf{x}_0) = \sum_{i=1}^{n_{\mathbf{x}_0,1}} \lambda_{i,1} \{s_1(\mathbf{x}_i) - \mu_1(\mathbf{x}_i)\} + \sum_{j=2}^{m} \sum_{i=1}^{n_{\mathbf{x}_0,j}} \lambda_{i,j} \left\{s_j(\mathbf{x}_i) - \mu_j(\mathbf{x}_i)\right\}, \tag{10.44}$$

where $\mu_j(\mathbf{x}_i)$ is the expected value of s_j at $\mathbf{x}_i$. The cokriging estimator is the best linear unbiased estimator best in the sense that the cokriging variance

$$\sigma_C^2(\mathbf{x}_0) = E\left[\{\widehat{s}_1(\mathbf{x}_0) - S_1(\mathbf{x}_0)\}^2\right] \tag{10.45}$$

is minimised and unbiased in that $E[\widehat{s}_1(\mathbf{x}_0) - S_1(\mathbf{x}_0)] = 0$. As with univariate kriging, there are two alternative treatments of the local mean. The first is to regard it either as a known constant value for all locations (simple cokriging) or as an unknown value fixed within the neighbourhood about $\mathbf{x}_0$ from which kriged

estimates are derived (the ordinary cokriging). The second approach is to model it as a combination of known functions of location (simple cokriging with a trend). Here we consider the first option under ordinary kriging version. Let variable 1 be s_u the variable to be predicted from observations $s_u(\mathbf{x}_i)$ and observations of a secondary variable $s_v(\mathbf{x}_i)$ although this can be extended to a number of secondary variables. With a similar rearrangement of Eq. 10.44) to that done above for univariate ordinary kriging, it is clear that constraining the weights so that

$$\sum_{i=1}^{n_{x_0,u}} \lambda_{i,u} = 1$$

and

$$\sum_{i=1}^{n_{x_0,v}} \lambda_{i,v} = 0$$

will filter the (unknown) means from the estimator, while the resulting estimator is unbiased. These are two independent constraints, so the kriging equations need two Lagrange parameters. By a similar derivation to univariate ordinary kriging, we obtain our cokriging equations which may be solved for the weights to be substituted into the cokriging estimator

$$\widehat{s}_u(\mathbf{x}_0) = \sum_{i=1}^{n_{x_0,u}} \lambda_{i,u} s_u(\mathbf{x}_i) + \sum_{i=1}^{n_{x_0,v}} \lambda_{i,v} s_v(\mathbf{x}_i). \tag{10.46}$$

There are three general instances in which we use cokriging. The classical case is where the secondary variable is more densely sampled than the target variable, and so the cross-covariance information allows us to obtain better predictions of the target variable than we could using the data on that variable alone. An interesting example of this approach is given by Leenaers et al. (1990). In general counting for the cross-covariance structure will give improved predictions when the cross-covariance structure includes information on scale-dependent relationships between the variables, i.e. the co-dispersion coefficient changes with lag. If the co-dispersion coefficient is constant, then the cross-covariance model adds no information.

When the primary and secondary variables are equally sampled, then there is generally little advantage from cokriging in terms of improved precision of estimates, but we do achieve coherence as discussed above. Cokriging linear combinations of variables directly (such as the change variable where s_u and s_v are measurements of the same variable at time u and time v) is described in more detail by Papritz and Fluhler (1994), and an example of this approach applied to soil monitoring is given by Lark et al. (2006). One special case in which these properties of cokriging are useful is when we work with compositional data, that is to say,

variables which sum to a fixed value (typically 1, or 100%). In soil science we are often interested in the sand, silt and clay content of soils. These are exhaustive particle size classes, and the values sum to one. It is inappropriate to analyse the proportions or percentages separately since separate OK predictions of sand, silt and clay percentage are not guaranteed to sum to 100. An alternative is to apply cokriging to the additive log-ratio of the particle size fractions. This methodology is set out in detail by Pawlowsky-Glahn and Olea (2004), and Lark and Bishop (2007) illustrate the method applied to data on soil particle size fractions.

A third instance is where we require information to support a decision which depends on two or more variables and when we are interested in the uncertainty in the information. In this case the joint prediction error, characterised by a covariance matrix for kriging estimates, can be obtained as output from cokriging. Lark et al. (2014) give an example in a study in the northern counties of the Irish Republic, and we present this as an example of the application of cokriging.

Advice to farmers in Ireland on the risk of cobalt deficiency for grazing livestock is based on the total concentration of cobalt and manganese in the soil. This is because manganese oxides can bind cobalt, reducing its availability. Lark et al. (2014) used data from a large regional survey of the soil. They conducted two separate geostatistical analyses in two geological domains with contrasting behaviour. Figure 10.11 shows the auto-variograms for Co and Mn and the cross-variograms for both domains, using both the standard estimator in Eq. 10.39 and a robust estimator from Lark (2003). The cokriged maps for Co and manganese are shown in Fig. 10.12. From the cokriging estimates and covariance matrix of cokriging errors, it was then possible, assuming jointly normal errors, to estimate the probability that the soil concentrations at any location fell within a range of values where a Co deficiency would be expected. This probability is mapped in Fig. 10.13.

We are most usually interested in multivariate methods of spatial prediction when the soil variable of primary interest is costly or difficult to measure and so is sampled relatively sparsely. However, one or more secondary variables are available which are easy and/or cheap to measure and so can be sampled densely and are also thought to contain information about the primary variable. Such variables may be remotely sensed measurements of the earth's surface, digital elevation models of terrain or data from geophysical surveys. When it is plausible to treat our variables primary and secondary as intrinsic random functions which conform to an LMCR, then the cokriging described above provides the best linear unbiased estimator of the primary variable given the secondary variable. There are other approaches that should be considered, however. In particular, we can use the secondary variable as a fixed effect in the REML–E-BLUP discussed above. This approach is known as kriging with an external drift, and examples of this approach are given by Bourennane et al. (1996).

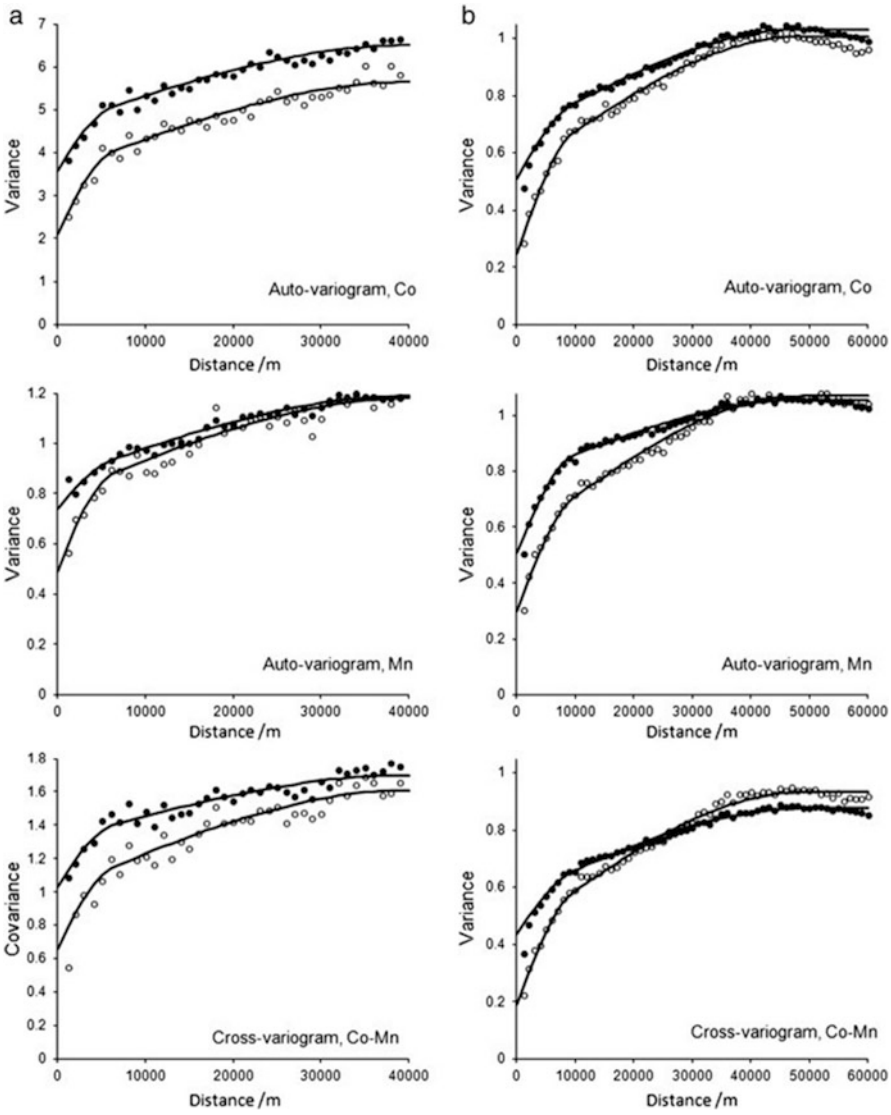

Fig. 10.11 Auto-variograms and cross-variograms for topsoil Co and Mn in two geological domains (**a**, **b**) of the northern counties of Ireland using (*solid symbol*) the standard estimators and (*open symbol*) a robust estimator, with fitted linear models of co-regionalisation (From Lark et al. (2014), under CC-BY licence. https://creativecommons.org/licenses/by/3.0/)

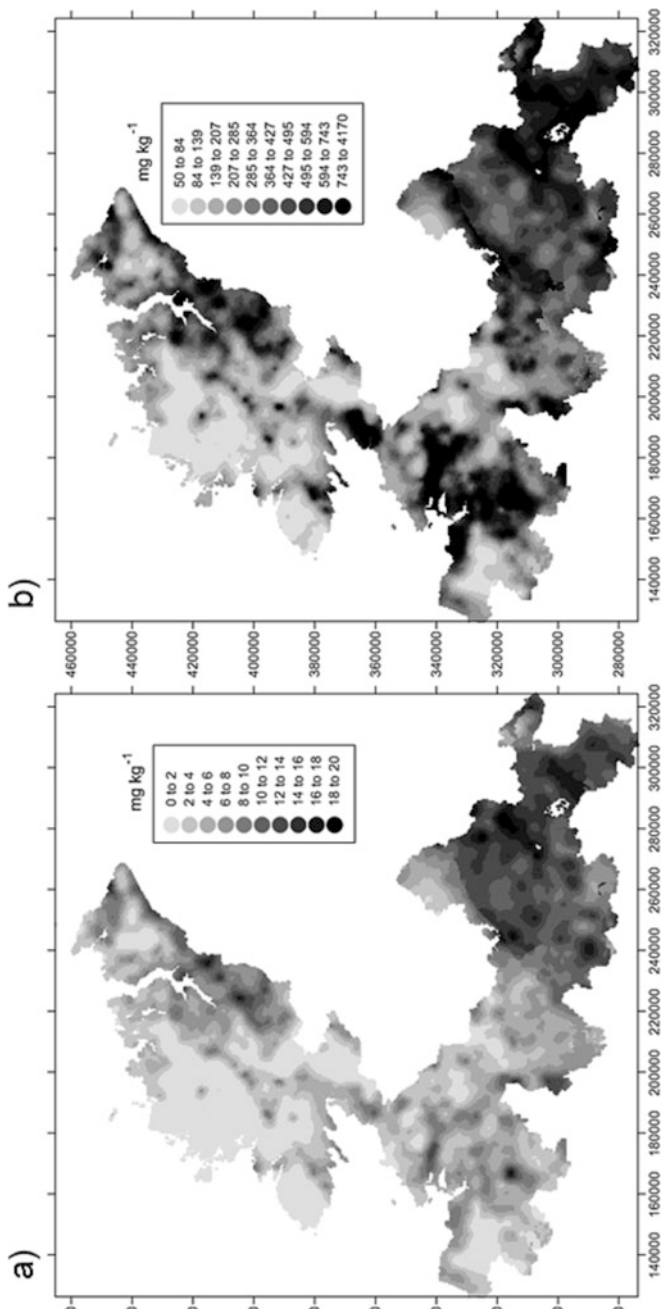

Fig. 10.12 Ordinary cokriging predictions of (**a**) topsoil Co and (**b**) topsoil Mn content across the northern counties of Ireland. Coordinates are the Irish National Grid (From Lark et al. (2014), under CC-BY licence. https://creativecommons.org/licenses/by/3.0/)

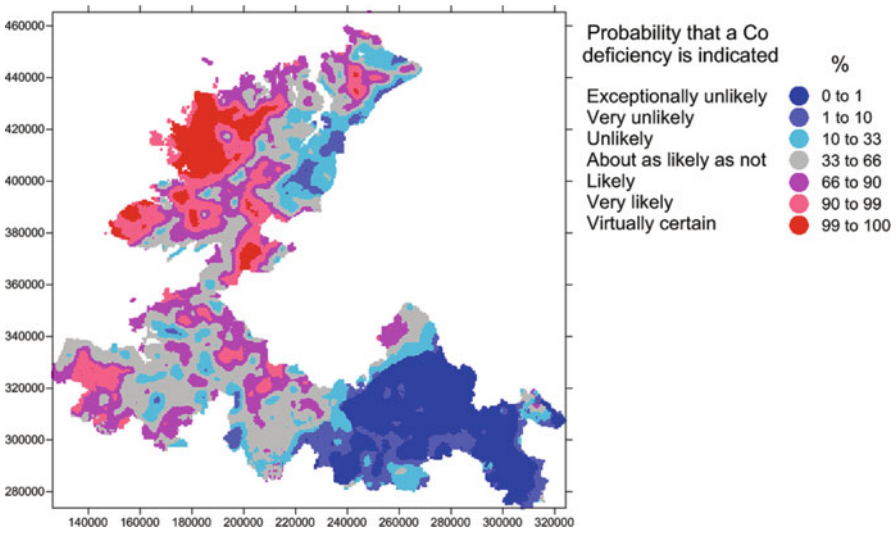

Fig. 10.13 Probability that the local Co and Mn content of soil indicates a risk of Co deficiency to livestock computed from the cokriging predictions and their error covariance matrices (From Lark et al. (2014), under CC-BY licence. https://creativecommons.org/licenses/by/3.0/)

10.6 Spatial Simulation

10.6.1 Simulation vs Prediction

We have discussed at length the problem of estimating a variable at a location from a set of observations on the assumption that our data are a realisation of a random function. Spatial simulation is a different problem. In simulation we draw several different realisations of the random function that we assume to be realised in our data. There are two general types of simulation. In *unconditional simulation* each set of simulated values represents a realisation of a random function with specified statistics (variogram, mean, histogram). In *conditional simulation* the realisation of the random function from which our values are drawn has specified statistics as in unconditional simulation, but is also required to reproduce the observed values at locations where we have sampled. We can think of unconditional simulation as drawing at random a set of values from a set of all possible realisations of a random function with specified statistics, and conditional simulation is drawn at random from a smaller subset of those sets of values which have the observed values at the sample sites.

Why might we want to do simulation? A good deal of simulation is done for pedometrical research or research in allied disciplines for a data with a realistic pattern of spatial variation in order to test the behaviour of pedometrical methods for estimation (e.g. Webster and Oliver 1992; Papritz and Webster 1995). But

simulation is a pedometrical tool in its own right, not simply a tool for research. It is seen in Chap. 13 how simulation can be used to examine the propagation of error in sources of information through models. If we generate multiple realisations of a random function conditional on observations and then look at distribution of simulated values at a particular location, these provide us with an estimate of the conditional cumulative distribution function (ccdf) of the variable at the location (Journel 1994a, b).

Conditional simulation is also used when we need to tackle problems where the variable of interest is obtained by some complex non-linear process model and our data are key model inputs. Consider for example the problem of predicting concentration of pollutant in water from a borehole when a plume of the pollutant has been released at the soil surface some distance away. The movement of the pollutant can be modelled if we know among other things the unsaturated conductivity functions of the soil at intervening points. This can only be measured at a few sites. You might interpolate the conductivities at intervening sites by kriging. However, since the model is non-linear, the fact that the kriged estimates of the conductivity are necessarily much less variable than the true values (although each estimate is unbiased) means that the simulated behaviour of the pollutant is a biased estimate of the true behaviour. An alternative is to generate a conditional simulation of the field of conductivities and to apply the model to these values. Multiples of such realisations can be generated, and from these we may generate a ccdf for an important model output such as the peak concentration of the pollutant in the well water. Other examples could be given where we wish to explore the aggregate impact on some region of some non-linear function of the spatially variable soil property, e.g. Viscarra Rossel et al. (2001) use simulation to estimate the cost function of different strategies for spatially variable application of lime to agricultural fields.

10.6.2 Methods for Simulation

LU Decomposition The most straightforward method for spatial simulation is known as LU decomposition. This is long established as a method to simulate correlated variables; an account of its use in geostatistics is given by Davis (1987). The name is somewhat misleading since it more properly describes the central algebraic step of the process where we perform a factorisation of a matrix into lower (L) and upper (U) triangular matrices. LU decomposition generates realisations of a multivariate normal process. This means that, unless we can assume that our data are from a multivariate normal process, we must transform them to normality. This is done typically by a standard normal score transform where we order our data values and replace each with a corresponding centile of a standard normal variable. The transformed data are stored in a row vector $\mathbf{s}$. We then test the plausibility of bivariate normality for pairs of data (see Goovaerts 1997 for details) and then estimate and model a variogram for the normal score transformation

of the data. From this, via Eq. 10.3, we can then compute a covariance matrix for the nodes at which we wish to simulate and the nodes at which we have data on which the simulation is to be conditioned. Element [*i,j*] of covariance matrix $\mathbf{C}$ therefore contains the covariance between node i and node j which, under the stationarity assumption, depends only on the distance in space between the nodes. We order the nodes so that our covariance matrix can be thought of as four component matrices $\mathbf{C}_{D,D}$ (covariance among conditioning data points), $\mathbf{C}_{N,N}$ (covariance among simulation nodes) and $\mathbf{C}_{N,D} = \mathbf{C}^T{}_{D,N}$ (covariance between conditioning data and simulation nodes), so

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{D,D} & \mathbf{C}_{D,N} \\ \mathbf{C}_{N,D} & \mathbf{C}_{N,N} \end{bmatrix}. \tag{10.47}$$

The Cholesky decomposition is a factorisation so that

$$\mathbf{C} = \mathbf{LU} = \begin{bmatrix} \mathbf{L}_{D,D} & \mathbf{0}_{D,N} \\ \mathbf{L}_{N,D} & \mathbf{L}_{N,N} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{D,D} & \mathbf{U}_{D,N} \\ \mathbf{0}_{N,D} & \mathbf{U}_{N,N} \end{bmatrix}, \tag{10.48}$$

where $\mathbf{0}_{D,N}$ ($=\mathbf{0}^T{}_{N,D}$) is a $D \times N$ matrix of zeros. Now, if N standard independent normal random variables are generated in column vector $\mathbf{y}$, then a realisation of our random variable is given in row vector $\breve{\mathbf{s}}$ by

$$\breve{\mathbf{s}} = \mathbf{L}_{N,D}\mathbf{L}_{D,D}\mathbf{s} + \mathbf{L}_{N,N}\mathbf{y}. \tag{10.49}$$

The standard normal variable is then back transformed with the empirical normal score transform originally used on the raw data.

In many ways this is the most theoretically satisfactory method for simulation. Its major drawback in theoretical terms is the assumption of multivariate normality. This assumption can never be disproved or proved for real data; the normal score transform can ensure reasonable representation of the desired histogram only, which is a necessary but not a sufficient condition for the multivariate assumption to be plausible. The multivariate normal assumption can cause problems in applications of simulation to the kinds of problem described above where we wish to use the simulated field in a model. Because the multivariate normal distribution of a set of data has maximum entropy, the extreme values of the distribution tend to be disconnected spatially. Thus in our example of pollutant movement, we find that regions with very large conductivity, which may occur in nature and have a disproportionately large effect on transport through the soil, will not be adequately reproduced by simulation. Another limitation of LU composition is computational. The Cholesky factorisation step is computationally expensive, and commonly it is impractical to simulate more than a few thousand data points. This is not a very large two-dimensional array. Once the decomposition has been done, however, as many realisations as required can be very rapidly computed.

Pebesma and Heuvelink proposed the use of Latin hypercube sampling (LHS) rather than complete random samples (Eq. 10.9) to more efficiently generate the random field. In their simulation study, they found that LHS of size 20 can perform equally as well as a simple random sample using 2,000 realisations.

Sequential methods Sequential simulation has been developed in response to the limitations that the computational load of the Cholesky composition imposes on the size of the fields that may be simulated using LU methods (Journel 1994b). A good account is given by Goovaerts (1997). Consider the simple problem of simulating a realisation of a random function at two locations $\mathbf{x}_1$ and $\mathbf{x}_2$, conditional on n data in the set γ_n. We may characterise the joint distribution of $S(\mathbf{x}_1)$ and $S(\mathbf{x}_2)$ by the bivariate conditional cumulative frequency distribution

$$F\left(\mathbf{x}_1, \mathbf{x}_2; z_1, z_2 \,|\zeta_n\right) \equiv \Pr\left\{s\left(\mathbf{x}_1\right) \leq s_1, s\left(\mathbf{x}_2\right) \leq s_2 \,|\zeta_n\right\}. \tag{10.50}$$

Following Bayes' rule, we can write

$$F\left(\mathbf{x}_1, \mathbf{x}_2; s_1, s_2 \,|\zeta_n\right) = F\left(\mathbf{x}_2; s_2 \,|\zeta_n, s\left(\mathbf{x}_1\right)\right) F\left(\mathbf{x}_1; s_1 \,|\zeta_n\right) \tag{10.51}$$

where the condition of the first ccdf on the right-hand side refers to the n conditioning data and one realisation generated at $\mathbf{x}_1$. This is why the simulated method is sequential because the data of simulated modes are generated as realisations of processes with conditional distribution functions conditioned on a sequence of modes and, ultimately, the conditioning data.

The data are transformed to normality and the plausibility of the bivariate normal assumption is tested. The conditional cumulative distribution function (ccdf) at the first node to be simulated is specified as N (m,v) where m is a kriged estimate at the node and v is the simple kriging variance. Note that simple kriging, ordinary kriging or universal kriging can be used to generate m, but the simple kriging variance must always be specified. Similarly an indicator transform can be used if we wish to simulate indicator variables, but the simple indicator kriging must be used to generate v. The simulation precedes sequentially from one simulation node to another accumulating the simulated values as conditioning ones.

The sequential simulation procedure is efficient. It can be made more so. In kriging there is a "screening effect" whereby the influence of a datum on a krige estimate is masked by an intervening observation. Thus there is little loss of information but a gain in speed if local ccdfs are calculated from all the conditioning data but only the nearest neighbouring simulation modes. In fact the screening effect can cause problems for reproducing long-range correlation structures. For this reason it is sometimes preferred to simulate on a coarse sub-grid of modes and then to "fill in" in intervening locations. Such multiple grids should be visited in a sequence that is randomised between realisations.

In Fig. 10.14 we show five independent conditional simulations of soil copper across our study area in the east of Scotland. These were obtained at the same prediction sites as the kriging predictions in Fig. 10.9 by sequential Gaussian

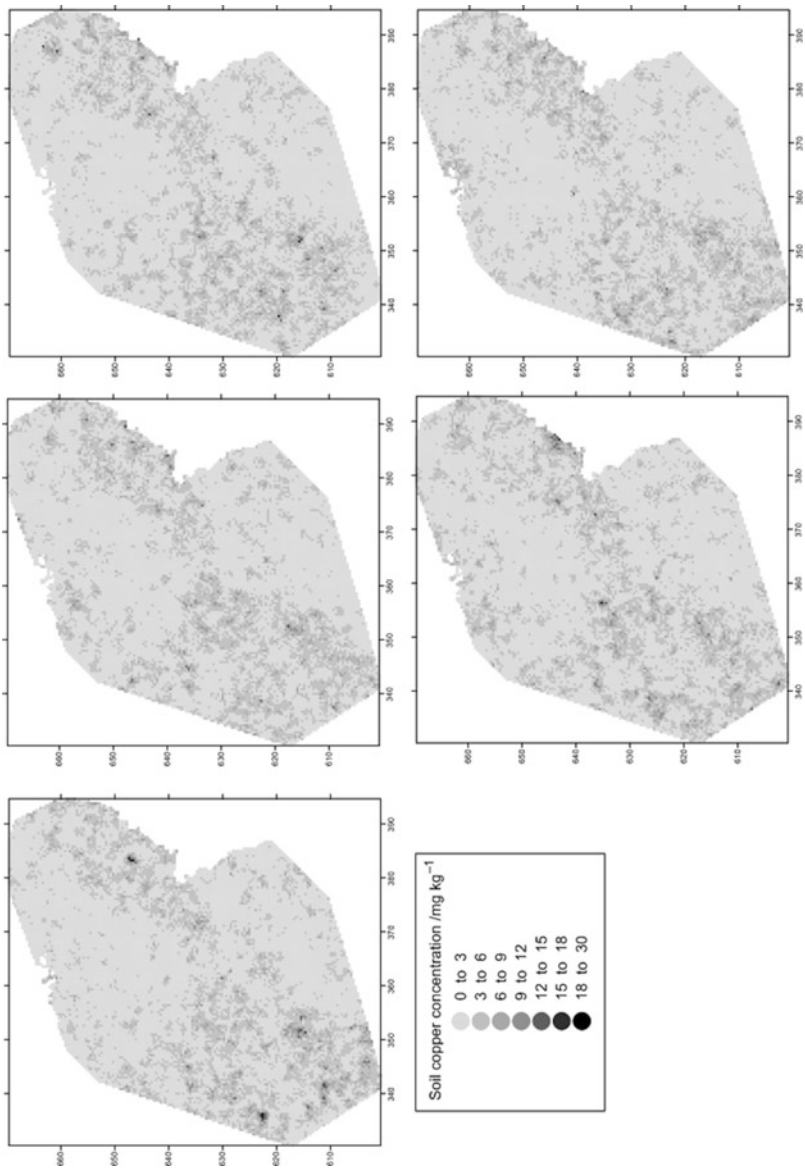

Fig. 10.14 Five mutually independent conditional simulations by sequential Gaussian simulation of lead content in the topsoil of the east of Scotland

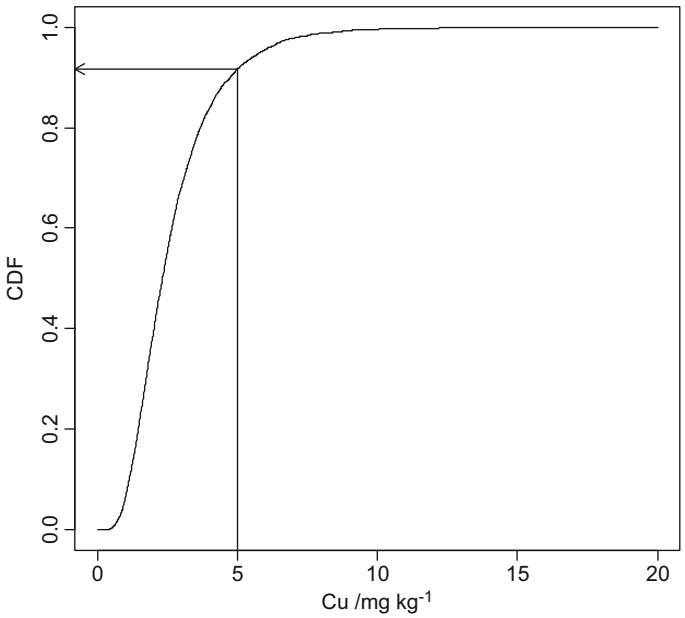

Fig. 10.15 Empirical conditional cumulative distribution function for topsoil lead content at a site (coordinates 340,642) in the east of Scotland, obtained from 5,000 mutually independent conditional simulations by sequential Gaussian simulation

simulation of the log-transformed variable. Each simulated value was then back transformed to the original scale of measurement. In Fig. 10.15 we consider a single location (coordinates 340,642). Here we simulated 5,000 independent realisations, conditional on the data. The figure shows the empirical cumulative distribution function (CDF) of these values, which is the conditional CDF of the variable at that site (conditional on the geostatistical model and the data). We may read off the value of the empirical CDF for some concentration of copper (here 5 mg kg^{-1}) from this graph. The value is 0.92, which allows us to estimate the conditional probability that soil copper content at that location exceeds 5 mg kg^{-1} as 0.08 or 8%.

The usual sequential Gaussian simulation is necessarily tied to assumptions of multivariate normality. These may be questionable and, as we discuss above in LU context, inappropriate assumptions of multivariate normality can cause problems for the simulation of conductivity fields or other variables where the connectedness of large values can have a substantial effect on aggregate behaviour. This is the motivation for recent studies of multiple-point geostatistics in soil science (e.g. Meerschman et al. 2013).

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petov BN, Csaki F (eds) 2nd international symposium on information theory. Akademia Kiado, Budapest, pp 267–281

Barry RP, Kelley Pace R (1997) Kriging with large data sets using sparse matrix techniques. Commun Stat-Simul Comput 26(2):619–629

Beckett PHT, Bie SW (1976) Reconnaissance for soil survey. II. Pre-survey estimates of the intricacy of the soil pattern. J Soil Sci 26:101–110

Beckett PHT, Webster R (1971) Soil variability: a review. Soils Fertil 34:1–15

Bourennane H, King D, Chéry P, Bruand A (1996) Improving the kriging of a soil variable using slope gradient as external drift. Eur J Soil Sci 47

Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138:86–95

Burgess TM, Webster R (1980a) Optimal interpolation and isarithmic mapping of soil properties I the semi-variogram and punctual kriging. J Soil Sci 31:315–331

Burgess TM, Webster R (1980b) Optimal interpolation and isarithmic mapping of soil properties II block kriging. J Soil Sci 31:333–341

Burgess TM, Webster R, McBratney AB (1981) Optimal interpolation and isarithmic mapping of soil properties IV sampling strategy. J Soil Sci 32:643–659

Burrough PA (1983) Multiscale sources of spatial variation in soil. II. A non-Brownian fractal model and its application in soil survey. J Soil Sci 34:599–620

Buttafuoco G, Castrignano A (2005) Study of the spatio-temporal variation of soil moisture under forest using intrinsic random functions of order *k*. Geoderma 128:208–220

Chilès J, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Clark I, Basinger KL, Harper WV (1989) MUCK – a novel approach to cokriging. In: Buxton BE (ed) Proceedings of the conference on geostatistical, sensitivity and uncertainty methods for ground-water flow and radionuclide transport modeling. Battelle Press, Columbus, pp 473–493

Corstanje R, Grunwald S, Lark RM (2008) Inferences from fluctuations in the local variogram about the assumption of stationarity in the variance. Geoderma 143:123–132

Cressie N (1985) Fitting variogram models by weighted least squares. Math Geol 17:563

Cressie N (1993) Statistics for spatial data. Wiley, New York

Cressie N, Johannesson G (2008) Fixed rank kriging for very large spatial data sets. J R Stat Soc: Ser B (Stat Methodol) 70(1):209–226

Cressie, N., Kang, E.L., 2010. High-resolution digital soil mapping: Kriging for very large datasets. In: Proximal soil sensing. Springer, Dordrecht, pp 49–63

Davis MW (1987) Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Math Geol 19:91–98

Deutsch CV, Journel AG (1998) GSLIB geostatistical software and user's guide, 2nd edn. Oxford University Press, New York

Dowd PA (1984) The variogram and kriging: robust and resistant estimators. In: Verly G, David M, Journel AG, Marechal AŽ (eds) Geostatistics for natural resources characterization. Part 1. Reidel, Dordrecht, pp 91–106

Goovaerts P (1994) Comparative performance of indicator algorithms for modeling conditional probability distribution functions. Math Geol 26:389–411

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

Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math Geol 24:269–286

Haas TM (1990) Lognormal and moving window methods of estimating acid deposition. J Am Stat Assoc 85:950–963

Halvorson JJ, Smith JL, Bolton H, Rossi RE (1995) Evaluating shrub-associated patterns of soil properties in a shrub steppe ecosystem using multiple-variable geostatistics. Soil Sci Soc Am J 59:11476–11487

Isaaks E, Srivastava RM (1989) An introduction to applied Geostatistics. Oxford University Press, New York
Journel AG (1983) Non-parametric estimation of spatial distributions. Math Geol 15:445–468
Journel AG (1994a) Resampling from stochastic simulations. Environ Ecol Stat 1:63–91
Journel AG (1994b) Modelling uncertainty: some conceptual thoughts. In: Dimitrakopoulos R (ed) Geostotistics for the next century. Kluwer Academic, Dordrecht, pp 30–43
Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London
Kaufman CG, Schervish MJ, Nychka DW (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. J Am Stat Assoc 103(484):1545–1555
Künsch HR, Papritz A, Bassi F (1997) Generalized cross-covariances and their estimation. Math Geol 29:779–799
Lark RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. Eur J Soil Sci 51:137–157
Lark RM (2002) Robust estimation of the pseudo cross-variogram for cokriging soil properties. Eur J Soil Sci 53:253–270
Lark RM (2003) Two robust estimators of the cross-variogram for multivariate geostatistical analysis of soil properties. Eur J Soil Sci 54:187–201
Lark RM, Bishop TFA (2007) Cokriging particle size fractions of the soil. Eur J Soil Sci 58: 763–774
Lark RM, Ferguson RB (2004) Mapping risk of soil nutrient deficiency or excess by disjunctive and indicator kriging. Geoderma 118:39–53
Lark RM, Papritz A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115:245–260
Lark RM, Bellamy PH, Rawlins BG (2006) Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. Geoderma 133:363–379
Lark RM, Ander EL, Cave MR, Knights KV, Glennon MM, Scanlon RP (2014) Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. Geoderma 226–227:64–78
Leenaers H, Okx JP, Burrough PA (1990) Comparison of spatial prediction methods for mapping floodplain soil pollution. Catena 17:535–550
Marchant BP, Lark RM (2006) Adaptive sampling for reconnaissance surveys for geostatistical mapping of the soil. Eur J Soil Sci 57:831–845
Marchant BP, Lark RM (2007) Optimized sample schemes for geostatistical surveys. Math Geol 39:113–134
Matérn B (1986) Spatial variation. Springer, Berlin
Matheron G (1962) Traité de géostatistique appliquée, Tome I: Mémoires du Bureau de RechercheesGéologiques et Minières, vol 14. Editions Technip, Paris
Matheron G (1976) A simple substitute for conditional expectation: the disjunctive kriging. In: Guarascio M, David M, Huijbregts C (eds) Advanced geostatistics in the mining industry. D. Reidel, Dordrecht, pp 221–236
McBratney AB, Webster R (1983) Optimal interpolation and isarithmic mapping of soil properties V. Co-regionalization and multiple sampling strategy. J Soil Sci 34:137–162
McBratney AB, Webster R (1986) Choosing functions for semivariances of soil properties and fitting them to sample estimates. J Soil Sci 37:617–639
McBratney AB, Webster R, Burgess TM (1981) The design of optimal sampling schemes for local estimation and mapping of regionalised variables. I. Theory and method. Comput Geosci 7:331–334
McBratney AB, Webster R, McLaren RG, Spiers RB (1982) Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland. Agronomie 2:969–982
Meerschman E, Van Meirvenne M, Van De Vijver E, De Smedt P, Islam MM, Saey T (2013) Mapping complex soil patterns with multiple-point geostatistics. Eur J Soil Sci 64:183–191
Minasny B, McBratney AB (2005) The Matérn function as a general model for soil variograms. Geoderma 128:192–207

Minasny B, McBratney AB, Whelan BM (2006) VESPER version 1.62. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney. NSW
Myers DE (1991) Pseudo-cross variograms, positive definiteness, cokriging. Math Geol 23: 805–816
Odeh IOA, McBratney AB, Chittleborough DJ (1994) Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63:197–214
Papritz A, Dubois J-P (1999) Mapping heavy metals in soil by (non-)linear kriging: an empirical validation. In: Gomez-Hernandez J, Soares A, Froidevaux R (eds) geoENV II: Geostatistics for environmental applications, Quantitative geology and geostatistics, vol 10. Kluwer Academic Publishing, Dordrecht, pp 429–440
Papritz A, Flühler H (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. Geoderma 62:29–43
Papritz A, Webster R (1995) Estimating temporal change in soil monitoring. 1. Statistical theory. Eur J Soil Sci 46:1–12
Papritz A, Kunsch HR, Webster R (1993) On the pseudo cross-variogram. Math Geol 25: 1015–1026
Pawlowsky-Glahn V, Olea RA (2004) Geostatistical analysis of compositional data. Oxford University Press, New York
Ploner A (1999) The use of the variogram cloud in geostatistical modelling. Environmetrics 10:413–437
Rivoirard J (1994) Introduction to disjunctive kriging and non-linear geostatistics. Oxford University Press, Oxford
Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York
Sun W, Minasny B, McBratney A (2012) Analysis and prediction of soil properties using local regression-kriging. Geoderma 171:16–23
Van Meirvenne M, Goovaerts P (2001) Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102:75–100
Viscarra Rossel RA, Goovaerts P, McBratney AB (2001) Assessment of the production and economic risks of site-specific liming using geostatistical uncertainty modelling. Environmetrics 12:699–711
von Steiger B, Webster R, Schulin R, Lehmann R (1996) Mapping heavy metals in polluted soil by disjunctive kriging. Environ Pollut 94:205–215
Webster R (2015) Daniel Krige's influence on soil science. J South Afr Inst Min Metall 11:165–117
Webster R, Burgess TM (1980) Optimal interpolation and isarithmic mapping of soil properties III changing drift and universal kriging. J Soil Sci 31:505–524
Webster R, Burgess TM (1984) Sampling and bulking strategies for estimating soil properties in small regions. J Soil Sci 35:127–140
Webster R, Butler BE (1976) Soil survey and classification studies at Ginninderra. Aust J Soil Res 14:1–24
Webster R, Lark RM (2013) Field sampling for environmental science and management. Routledge, London. pp viii+192
Webster R, Oliver MA (1992) Sample adequately to estimate variograms of soil properties. J Soil Sci 43:177–192
Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester. pp. x+328
Wood G, Oliver MA, Webster R (1990) Estimating soil salinity by disjunctive kriging. Soil Use Manag 6:97–104
Yates F (1948) Systematic sampling. Phil Trans R Soc A 241:345–377
Youden WJ, Mehlich A (1937) Selection of efficient methods for soil sampling. Contrib Boyce Thompson Inst Plant Res 9:59–70