Chapter 11
Model-Based Soil Geostatistics

Ben P. Marchant

"A model's just an imitation of the real thing".

Mae West

11.1 Introduction

In Chap. 10, we described how classical geostatistical methods can be used to interpolate measurements of soil properties at locations where they have not been observed and to calculate the uncertainty associated with these predictions. The idea that soil properties can be treated as realizations of regionalized random functions in this manner has perhaps been the most significant ever in pedometrics (Webster 1994). The approach has been applied in thousands of studies for every imaginable soil property at scales varying from the microscopic to the global and has greatly enhanced our understanding of the spatial variability of soil properties.

However, despite its popularity amongst pedometricians, the classical geostatistical methodology has met with some criticism (e.g. Stein 1999). This has primarily been because of the subjective decisions that are required when calculating the empirical semivariogram and then fitting the variogram model. When calculating the empirical semivariogram, the practitioner must decide on the directions of the lag vectors that will be considered, the lag distances at which the point estimates are to be calculated and the tolerance that is permitted for each lag distance bin (see Sect. 10.2). Furthermore, when fitting a model to the empirical variogram, the practitioner must decide which of the many authorized models should be used, what criterion should be applied to select the best fitting parameter values and the weights that should be applied when calculating this criterion for the different point estimates. Researchers have given some thought to how these selections should be made. For example, the Akaike information criterion (Akaike 1973) is often used

B.P. Marchant (✉)
British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
e-mail: benmarch@bgs.ac.uk

to select the authorized model that best fits the data. However, the formula used to calculate the AIC when using the method of moments variogram estimator is only an approximation and is very much dependent on how the fitting criterion is weighted for different lag distances (McBratney and Webster 1986). The selection of these weighting functions is a particular challenge since the uncertainty of the point estimates and hence the most appropriate choice of weights depend on the actual variogram which the user is trying to estimate. Further complications arise because the same observations feature in multiple point estimates of the experimental variogram, and some observations can feature more often than others. Therefore, these point estimates are not independent (Stein 1999). Strictly this correlation between the point estimates should be accounted for when fitting the model parameters. The correlation can result in artefacts or spikes in the experimental variogram that might be mistaken for additional variance structures (Marchant et al. 2013b).

In all of the choices listed above, different subjective decisions (or indeed carefully manipulated choices) can lead to quite different estimated variograms and in turn to quite different conclusions about the spatial variation of the soil property. Therefore, there is a need for a single objective function that quantifies how appropriately a proposed variogram model represents the spatial correlation observed in a dataset without requiring the user to make subjective decisions. Model-based geostatistics (Stein 1999; Diggle and Ribeiro 2007) uses the likelihood that the observations would arise from the proposed random function as such an objective criterion. The likelihood function is calculated using the observed data without the need for an empirical semivariogram. The variogram parameter values that maximize the likelihood function for a particular set of observations correspond to the best fitting variogram model. These parameters are referred to as the maximum likelihood estimate. Once a model has been estimated, it can be substituted into the best linear unbiased predictor (BLUP; Lark et al. 2006; Minasny and McBratney 2007) to predict the expected value of the soil property at unobserved locations and to determine the uncertainty of these predictions. The relative suitability of two proposed variograms can be assessed by the ratio of their likelihoods or the exact AIC. Hence, it is possible to compare objectively fitted models which use different authorized variogram functions. Rather than selecting a single best fitting variogram, it is also possible to identify a set of plausible variograms according to the closeness of their likelihoods to the maximized likelihood. Thus, the uncertainty in the estimate of the variogram can be accounted for by averaging predictions across this set of plausible variograms (Minasny et al. 2011).

The linear mixed model (LMM) is often used in model-based geostatistical studies. The LMM divides the variation of the observations into fixed and random effects. The fixed effects are a linear function of environmental covariates that describe the variation of the expectation of the random function across the study region. The covariates can be any property that is known exhaustively, such as the eastings or northings, the elevation and derived properties such as slope or a remotely sensed property. The random effects have zero mean everywhere, and they describe the spatially correlated fluctuations in the soil property that cannot be explained by the fixed effects.

   The model-based approach does have its disadvantages. The formula for the likelihood function for n observations includes the inverse of an $n \times n$ square matrix. This takes considerably longer to compute than the weighted difference between an empirical semivariogram and a proposed variogram function. Mardia and Marshall (1984) suggested that maximum likelihood estimation was impractical for $n > 150$. Modern computers can calculate the likelihood for thousands of observations (e.g. Marchant et al. 2011), but it is still impractical to calculate a standard likelihood function for the tens of thousands of observations that might be produced by some sensors. The use of the likelihood function also requires strong assumptions about the random function. The most common assumption being that the random effects are realized from a second-order stationary multivariate Gaussian random function. Such a restrictive set of assumptions is rarely completely appropriate for an environmental property and therefore the reliability of the resultant predictions is questionable. One active area of research is the development of model-based methodologies where these assumptions are relaxed.

   The software required to perform model-based geostatistics has been made available through several R packages such as geoR (Ribeiro and Diggle 2001) and gstat (Pebesma 2004). The software used in this chapter has been coded in MATLAB. It forms the basis of the Geostatistical Toolbox for Earth Sciences (GaTES) that before the end of 2017 will be available at http://www.bgs.ac.uk/StatisticalSoftware. The Bayesian analyses require the DREAM package (Vrugt 2016).

11.2   The Scottish Borders Dataset

We illustrate some of the model-based geostatistics approaches that have been adopted by pedometricians by applying them to a set of measurements of the concentrations of copper and cobalt in soils from the south east of Scotland (Fig. 11.1). This dataset was analysed using classical geostatistical methods in Chap. 10. The measurements were made between 1964 and 1982. At that time, there were concerns that livestock grazing in the area were deficient in copper and/or cobalt. Therefore, staff from the East of Scotland College of Agriculture measured the field-mean extractable concentrations of these elements in more than 3500 fields. Full details of the sampling protocol and laboratory methods are provided by McBratney et al. (1982). The dataset has been extensively studied using a selection of geostatistical techniques (McBratney et al. 1982; Goovaerts and Webster 1994; Webster and Oliver 2007). These authors have mapped the probabilities that the concentrations of copper and cobalt are less than acceptable thresholds (1.0 and 0.25 mg kg^{-1}, respectively) and related their results to the previously mapped soil associations.

   We consider a subset of the data. We randomly select 400 copper measurements and 500 cobalt measurements from this dataset. The remaining copper observations are kept for model validation (Fig. 11.2). We use soil information from the

Fig. 11.1 Location of the
Scottish Borders study region
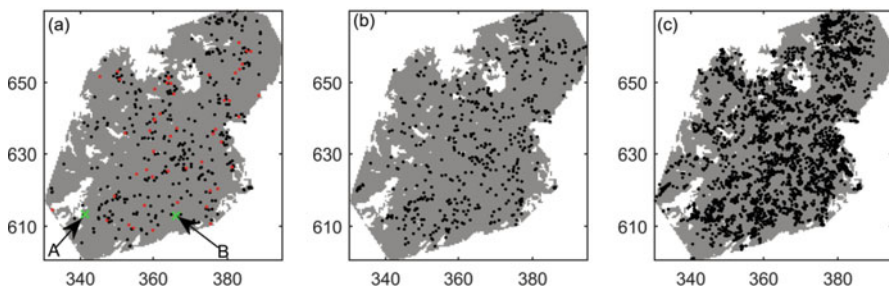(*shaded region*) within Great
Britain

Fig. 11.2 Locations of (**a**) 400 copper observations used for calibration, (**b**) 500 cobalt observations used for LMCR calibration and (**c**) 2481 copper observations used for model validation. The locations coloured *red* in (**a**) are the 50-point subset of these data, and the *green* crosses are the locations of the predictions shown in Figs. 11.8 and 11.9. Coordinates are km from the origin of the British National Grid

1:250,000 National Soil Map of Scotland (Soil Survey of Scotland Staff 1984). These data are available under licence from http://www.soils-scotland.gov.uk/data/ soils. The study area contains eight soil types. We only consider four of these, namely, mineral gleys, peaty podzols, brown earths and alluvial soils (Fig. 11.3), within which more than 95% of the soil measurements were made. Our primary
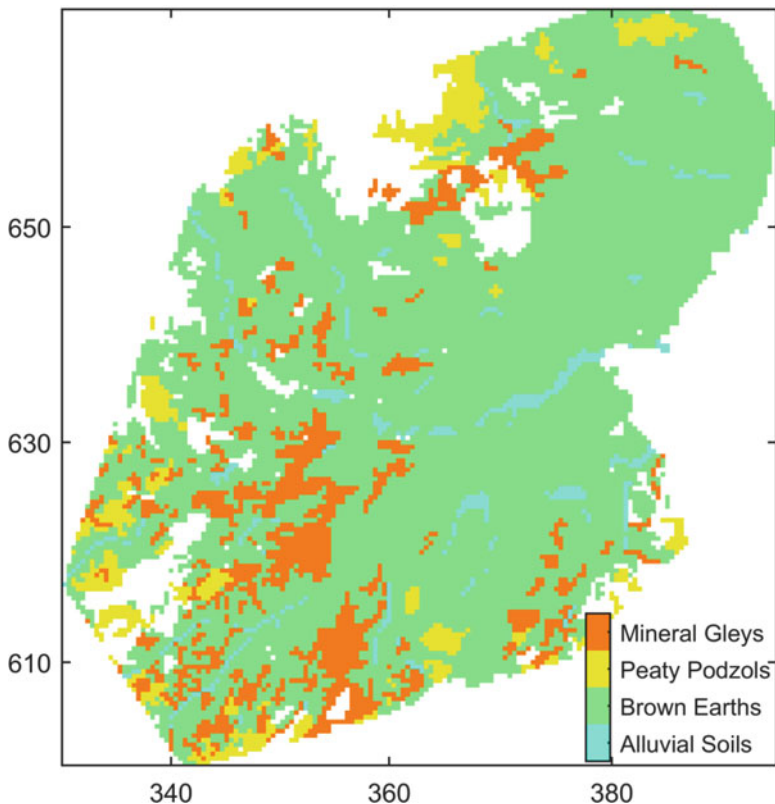
Fig. 11.3 Map of the soil types within the study region according to the 1:250,000 National Soil Map of Scotland (Used with permission from the James Hutton Institute, © Crown copyright). Coordinates are km from the origin of the British National Grid

objective is to use these data to map the concentration of copper and the probability that it is less than 1.0 mg kg^{-1}. We explore the extent to which the soil type information and the observed cobalt concentrations can explain the variation of copper concentrations and hence improve the accuracy of our maps. We validate our maps using the remaining 2481 observations of copper. We explore the implications of using fewer observations by repeating our analyses on a 50 sample subset of the 400 copper measurements (Fig. 11.2a). At the two sites marked by green crosses in Fig. 11.2a, we describe the prediction of the copper concentration in more detail. Site 'A' is 0.6 km from the nearest observation, whereas site 'B' is 1.3 km from an observation.

11.3   Maximum Likelihood Estimation

11.3.1   The Linear Mixed Model

We denote an observation of the soil property, in our case the concentration of copper, at location x_i by $s_i = s(x_i)$ and the set of observations by $\mathbf{s} = \{s_1, s_2, \ldots, s_n\}^{\mathrm{T}}$ where T denotes the transpose of the vector. We assume that $\mathbf{s}$ is a realization of an LMM:

$$\mathbf{s} = \mathbf{M}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{11.1}$$

Here, the $\mathbf{M}\boldsymbol{\beta}$ are the fixed effects and the $\boldsymbol{\varepsilon}$ are the random effects. The design matrix $\mathbf{M}$ is of size $n \times q$ where q is the number of covariates that are included in the random effects model. The vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)^{\mathrm{T}}$ contains the fixed effects parameters or regression coefficients. Each column of $\mathbf{M}$ contains the value of a covariate at the n observed sites. If the random effects include a constant term, then each entry of the corresponding column of $\mathbf{M}$ is equal to one. A column of $\mathbf{M}$ could consist of the values of a continuous covariate such as elevation or the output from a remote sensor (e.g. Rawlins et al. 2009). Thus, the fixed effects include a term that is proportional to the covariate. If the fixed effects differ according to a categorical covariate such as soil type, then the columns of $\mathbf{M}$ include c binary covariates which indicate the presence or absence of each of the c classes at each site. Note that if the c classes account for all of the observations, then a constant term in the fixed effects would be redundant.

The vector $\boldsymbol{\varepsilon}$ contains the values of the random effects at each of the n sites. The elements of $\boldsymbol{\varepsilon}$ are realized from a zero-mean random function with a specified distribution function. The $n \times n$ covariance matrix of the random effects is denoted $\mathbf{C}(\boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ is a vector of covariance function parameters. The elements of this matrix can be determined from any of the authorized and bounded variogram functions described in Chap. 10 since

$$C(h) = \gamma(\infty) - \gamma(h), \tag{11.2}$$

where $C(h)$ is a covariance function for lag h, $\gamma(h)$ is a bounded and second-order stationary variogram and $\gamma(\infty)$ is the upper bound or total sill of the variogram. We will focus on the nested nugget and Matérn variogram (Minasny and McBratney 2005) so that

$$C(h) = \begin{cases} c_0 + c_1 \text{ if } h = 0 \\ c_1 G(h) \text{ for } h > 0' \end{cases} \tag{11.3}$$

where

$$G(h) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}h}{a}\right)^{\nu} K_{\nu}\left(\frac{2\sqrt{\nu}h}{a}\right). \tag{11.4}$$

Here, Γ is the gamma function and K_v is a modified Bessel function of the second kind of order v. The random effects model parameters are c_0 the nugget, c_1 the partial sill, a the distance parameter and v the smoothness parameter. Jointly, there are four random effect parameters, i.e. $\boldsymbol{\alpha} = (c_0, c_1, a, v)^{\mathrm{T}}$.

11.3.2   Coregionalized Soil Properties

In Sect. 10.4, we saw that it was possible to extend the classical geostatistical model of a single soil property to consider the spatial correlation between observations of two or more properties by using a linear model of coregionalization (LMCR). Marchant and Lark (2007a) demonstrated how the LMM could also be extended to include coregionalized variables. The LMCR consists of a variogram for each soil property and a series of cross-variograms describing the spatial correlation between each pair of properties. Each variogram or cross-variogram must have the same variogram structure. This means they are based on the same authorized models, and they have common spatial and, in the case of the Matérn model, smoothness parameters. The nugget and sill parameters can differ for the different soil properties. We denote these parameters by $c_0^{d,e}$ and $c_1^{d,e}$, respectively, where the d and e refer to the different soil properties. So, if $d = e$, these are parameters of a variogram, whereas if $d \neq e$, they are parameters of a cross-variogram. There are further constraints to ensure that the LMCR leads to positive definite covariance matrices. If we define matrices $\mathbf{B}_0$ and $\mathbf{B}_1$ by

$$\mathbf{B}_0 = \begin{bmatrix} c_0^{1,1} & \cdots & c_0^{1,v} \\ \vdots & \ddots & \vdots \\ c_0^{v,1} & \cdots & c_0^{v,v} \end{bmatrix} \text{ and } \mathbf{B}_1 = \begin{bmatrix} c_1^{1,1} & \cdots & c_1^{1,v} \\ \vdots & \ddots & \vdots \\ c_1^{v,1} & \cdots & c_1^{v,v} \end{bmatrix} \tag{11.5}$$

where v is the number of soil properties; then, both of these matrices must have a positive determinant.

The LMCR can be incorporated into the LMM (Eq. 11.1) by altering the random effects covariance matrix to accommodate the different variograms and cross-variograms. In this circumstance, the observation vector $\mathbf{s}$ will include observations of each of the v soil properties. It is likely that the random effects matrix $\mathbf{M}$ will require sufficient columns to accommodate different fixed effects for each soil property. If each variogram and cross-variogram consists of a nested nugget and Matérn model, then element i, j of the random effects covariance matrix would be:

$$C_{i,j} = \begin{cases} c_0^{v_i,v_j} + c_1^{v_i,v_j} & \text{if } h_{i,j} = 0 \\ c_1^{v_i,v_j} G\left(h_{i,j}\right) & \text{if } h_{i,j} > 0 \end{cases}, \tag{11.6}$$

where the ith element of $\mathbf{s}$ is an observation of property v_i, the jth element of $\mathbf{s}$ is an observation of property v_j and $h_{i,j}$ is the lag separating the two observations.

In common with the LMCR from classical geostatistics, the cross-variogram nuggets and sills must be constrained to ensure that the determinants of $\mathbf{B}_0$ and $\mathbf{B}_1$ are positive. For $v_i \neq v_j$, the $c_0^{v_i,v_j}$ parameter only influences the covariance function if there are some locations where both v_i and v_j are observed. If this is not the case, then the parameter cannot be fitted. If $v = 2$, and both properties are observed at some sites, then $\boldsymbol{\alpha} = \left(c_0^{1,1}, c_0^{1,2}, c_0^{2,2}, c_1^{1,1}, c_1^{1,2}, c_1^{2,2}, a, v \right)^{\mathrm{T}}$.

11.3.3 The Likelihood Function

If the distribution function of the random effects is assumed to be multivariate Gaussian, then the log of the likelihood function is equal to

$$\ln l(\mathbf{s}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = L(\mathbf{s}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \text{Constant} - \frac{1}{2}\ln|\mathbf{C}(\boldsymbol{\alpha})| - \frac{1}{2}(\mathbf{s}-\mathbf{M}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}(\mathbf{s}-\mathbf{M}\boldsymbol{\beta}),$$

(11.7)

where $|\ |$ denotes the determinant of a matrix. The log-likelihood is the objective function which we use to test the suitability of an LMM to represent the spatial variation of a soil property.

The assumption of Gaussian random effects is restrictive and often implausible for soil properties. For example, the histograms of observed copper and cobalt concentrations in the Scottish Borders region are highly skewed (Fig. 11.4a, b). A transformation $s^* = H(s)$ can be applied to the data so they more closely approximate a Gaussian distribution. Figure 11.4c, d show the more symmetric histograms that result when the natural log-transform, $s^* = \ln(s)$, is applied to the copper and cobalt concentrations. The Box-Cox transform might also be applied to skewed data. It generalizes the natural log-transform via a parameter λ:

$$H(s) = \begin{cases} \ln(s) & \text{if } \lambda = 0, \\ \frac{s^\lambda - 1}{\lambda} & \text{otherwise.} \end{cases}$$

(11.8)

If we assume that the observed data are multivariate Gaussian after the application of a transformation, then the formula for the log-likelihood becomes

$$L(\mathbf{s}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \text{Constant} - \frac{1}{2}\ln|\mathbf{C}(\boldsymbol{\alpha})| - \frac{1}{2}(\mathbf{s}^*-\mathbf{M}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}(\mathbf{s}^*-\mathbf{M}\boldsymbol{\beta})$$
$$+ \sum_{i=1}^{n} \ln(J\{s_i\})$$

(11.9)

where $J\{s_i\} = \frac{\mathrm{dH}}{\mathrm{ds}}$ is the derivative of the transformation function evaluated at $s = s_i$. For the Box-Cox transform,
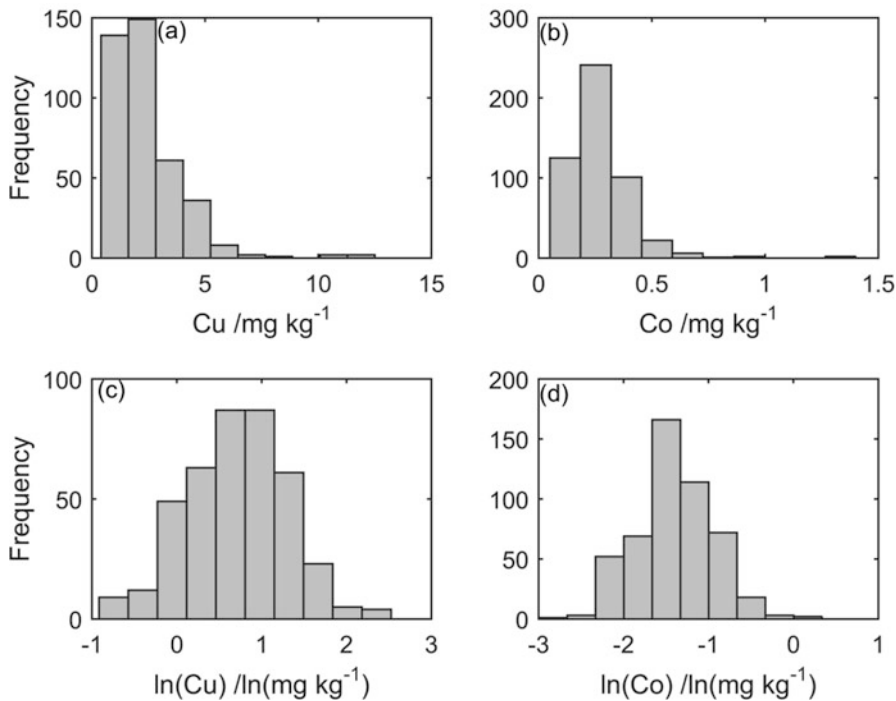
Fig. 11.4 Histograms of the observations of (**a**) 400 soil copper concentrations, (**b**) 500 soil cobalt concentrations, (**c**) the natural logarithm of the copper concentrations and (**d**) the natural logarithm of the cobalt concentrations

$$\ln\left(J\left\{s_i\right\}\right) = (\lambda - 1)\ln\left(s_i\right), \tag{11.10}$$

and the corresponding function for the natural log-transform is found by setting $\lambda = 0$.

Once a fixed-effect design matrix, a covariance function and any transformation have been proposed, the problem of maximum likelihood estimation is reduced to finding the elements of $\widehat{\alpha}$ and $\widehat{\beta}$ vectors which lead to the largest value of the log-likelihood and hence of the likelihood. This can be achieved using a numerical optimization algorithm to search the parameter space for these parameter values. In the examples presented in this chapter, we use the standard MATLAB optimization algorithm which is the Nelder-Mead method (Nelder and Mead 1965). This is a deterministic optimizer in the sense that if it is run twice from the same starting point, the same solution will result each time. It is prone to identifying local rather than global maxima. Therefore, Lark et al. (2006) recommends the use of stochastic optimizers such as simulated annealing which permit the solution to jump away from a local maximum. In our implementation of the Nelder-Mead method, we run

the algorithm from ten different starting points in an attempt to avoid the selection of local maxima. The parameter space is constrained to ensure that negative variance parameters or negative definite covariance matrices cannot result.

The optimization problem can be simplified by noting that the log-likelihood function is maximized when

$$\boldsymbol{\beta} = \left(\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{M}\right)^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{s}. \qquad (11.11)$$

Hence, there is only a need to search the $\boldsymbol{\alpha}$ parameter space, and the corresponding optimal value of $\boldsymbol{\beta}$ can be found by using the above formula. However, it is still a computational challenge to find an optimal $\boldsymbol{\alpha}$ vector within a four- or higher-dimensional parameter space. Diggle and Ribeiro (2007) suggest reducing the problem further to a series of optimizations in a lower-dimensional parameter space. They search for the optimal values of c_0, c_1 and a for a series of fixed v values. Then they plot the maximum log-likelihood achieved (or equivalently the minimum negative log-likelihood) against the fixed v and extract best of these estimates. This plot is referred to as a profile-likelihood plot.

It is possible to fit LMMs of varying degrees of complexity by adding more terms to the fixed and random effects. If we add an extra term to a model (e.g. an extra column to the $\mathbf{M}$ matrix), then the maximized likelihood of the more complex model will be at least as large as the likelihood of the simpler model. We need a test to decide whether the improvement that is achieved by adding extra terms is worthwhile. If an LMM is too complex, there is a danger of overfitting. This means that the model is too well suited to the intricacies of the calibration data, but it performs poorly on validation data that were not used in the fitting process.

If two LMMs are nested their suitability to represent the observed data can be compared by using a likelihood ratio test (Lark 2009). By nested, we mean that by placing constraints on its parameters it is possible to transform the more complex model to the simpler model. For example, if one LMM included a Box-Cox transform of the data and another model was identical except that the natural log-transform was applied, then these two models would be nested. The more complex model could be transformed to the simpler model by setting $\lambda = 0$. We denote the parameters of the complex model by $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$ and the parameters of the simpler model by $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$. Under the null hypothesis that the additional parameters in the more complex model do not improve the fit, the test statistic:

$$D = 2L\left(\mathbf{s}|\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1\right) - 2L\left(\mathbf{s}|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\right) \qquad (11.12)$$

will be asymptotically distributed as a chi-squared distribution with r degrees of freedom. Here, r is the number of additional parameters in the more complex model. Therefore, it is possible to conduct a formal statistical test to decide whether the more complex model has a sufficiently larger likelihood than the simpler one. However, the likelihood ratio test does not always meet our needs. An LMM with a nugget and exponential variogram and an identical model except that the variogram

is pure nugget would not be properly nested. This is because the more complex model has two additional parameters, c_1 and a, but only the first of these needs to be constrained to $c_1 = 0$ to yield the simpler model. Therefore, it is unclear what the degrees of freedom should be in the formal test. Lark (2009) used simulation approaches to explore this issue.

In this chapter, we calculate the AIC (Akaike 1973) for each estimated model:

$$\text{AIC} = 2k - 2L, \tag{11.13}$$

where k is the number of parameters in the model. The preferred model is the one with the smallest AIC value. This model is thought to be the best compromise between quality of fit (i.e. the likelihood) and complexity (the number of parameters). The AIC does not require the different models to be nested.

11.3.4   The Residual Maximum Likelihood Estimator

Patterson and Thompson (1971) observed that there was a bias in variance parameters estimated by maximum likelihood. This bias occurs because the $\boldsymbol{\beta}$ parameters are estimated from the data and are therefore uncertain, whereas in the log-likelihood formula, they are treated as if they are known exactly. This problem is well known when considering the variance of independent observations. The standard formula to estimate the variance of a population with unknown mean that has been sampled at random is:

$$\text{Var}(S) = \frac{1}{n-1}\sum_{i=1}^{n}(s_i - \bar{s})^2 \tag{11.14}$$

where n is the size of the sample and $\bar{s}$ is the sample mean. Since the variance is defined as $E[\{S - E(S)\}^2]$, one might initially be surprised that the denominator of Eq. 11.14 is $n - 1$ rather than n. However, it can be easily shown that if the denominator is replaced by n because the mean of the population is estimated, the expectation of the expression would be a factor of $n/(n-1)$ times the population variance.

Patterson and Thompson (1971) devised a method for correcting the analogous bias in the maximum likelihood estimates of random effects parameters. They transformed the data into stationary increments prior to calculating the likelihood. The likelihood of these increments was independent of the fixed effects and hence the bias did not occur. The expression for the residual log-likelihood that resulted is:

$$L_R(\mathbf{s}|\boldsymbol{\alpha}) = \text{Constant} + \frac{1}{2}\ln|\mathbf{W}| + \frac{1}{2}\ln|\mathbf{C}(\boldsymbol{\alpha})| + \frac{1}{2}\mathbf{s}^{\text{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{Q}\mathbf{s}, \tag{11.15}$$

where $\mathbf{W} = \mathbf{M}^{\text{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}\mathbf{M}$ and $\mathbf{Q} = \mathbf{I} - \mathbf{M}\mathbf{W}^{-1}\mathbf{M}^{\text{T}}\mathbf{C}(\boldsymbol{\alpha})^{-1}$.

11.3.5 Estimating Linear Mixed Models for the Scottish Borders Data

We illustrate maximum likelihood estimation of an LMM using the 400 copper concentration measurements from the Scottish Borders. We initially assume that the fixed effects are constant and that a log-transform is sufficient to normalize the data. Thus, we assume that the Box-Cox parameter λ is zero. In fact, when the Box-Cox parameter was unconstrained, the increase to the log-likelihood was negligible and did not improve the AIC. We estimated the c_0, c_1 and a parameters for fixed v equal to 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0 and 2.5. The profile-likelihood plot of the minimized negative log-likelihoods that resulted is shown in Fig. 11.5a. The smallest negative log-likelihood occurred for $v = 0.1$. When v was unfixed, a smaller negative log-likelihood resulted with $v = 0.12$. The best fitting variogram is plotted in Fig. 11.5b. It sharply increases as the lag is increased from zero, reflecting the small value of the smoothness parameter. This maximum likelihood estimate of the variogram is reasonably consistent with the empirical variogram (see Fig. 11.5b).

Figure 11.6 shows the maximum likelihood estimates of the variograms and cross-variograms when the **s** vector contained the 400 observations of copper and the 500 observations of cobalt. The natural log-transform was applied to each soil property, and the fixed effects consisted of a different constant for each property. Again, there is reasonable agreement between the maximum likelihood estimate and the empirical variograms. However, some small discrepancies are evident. The maximum likelihood estimate for cobalt has a longer range than might be
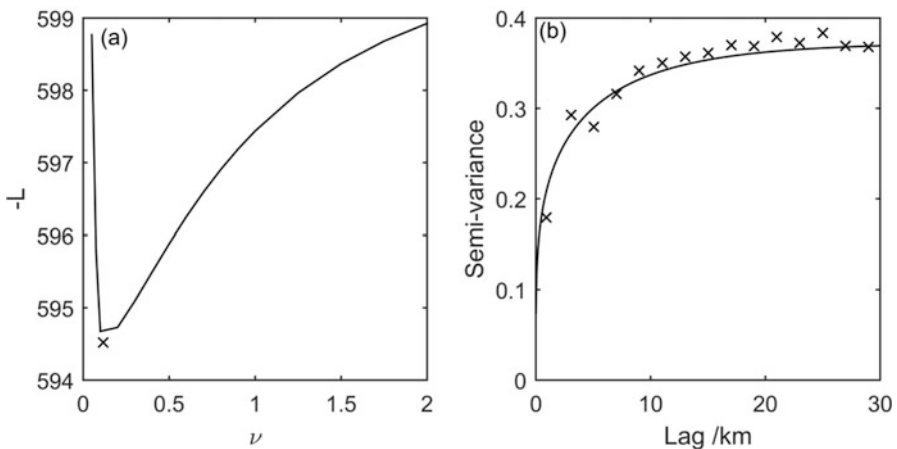

Fig. 11.5 (**a**) Plot of minimized negative log-likelihood of 400 ln copper concentrations for different fixed values of v. The *black* cross denotes the minimized negative log-likelihood when v is unfixed, (**b**) maximum likelihood estimate of the variogram (*continuous line*) of ln copper concentrations and the corresponding method of moments point estimates (*black crosses*)
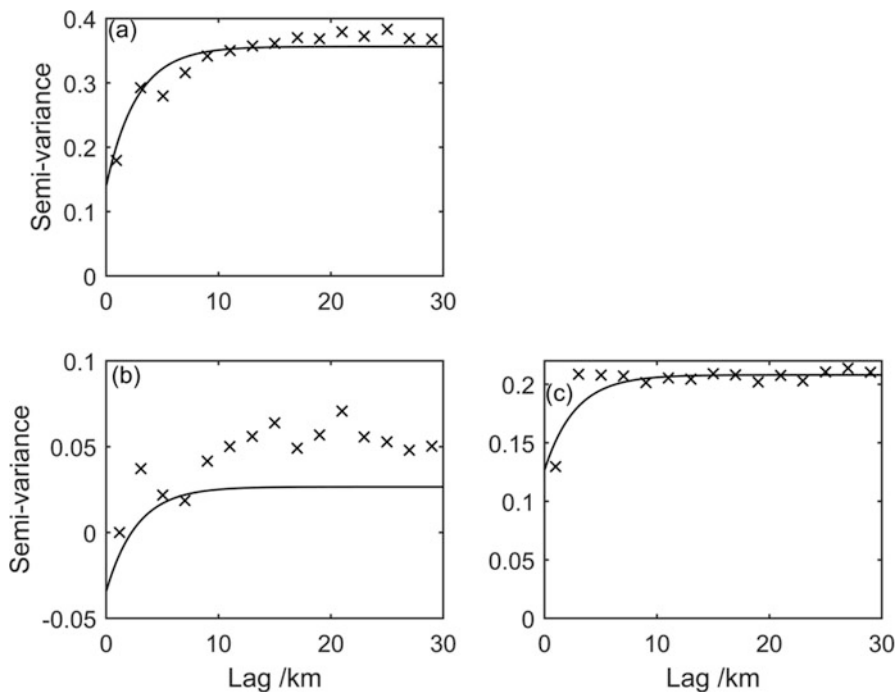
Fig. 11.6 Maximum likelihood estimate of (**a**) the LMCR ln copper auto-variogram, (**b**) the LMCR ln copper and ln cobalt cross-variogram and (**c**) the LMCR ln cobalt auto-variogram. The *black crosses* are the corresponding method of moments point estimates

fitted to the empirical variogram, and the maximum likelihood estimate of the cross-variogram appears to be consistently less than the corresponding empirical variogram. This was probably caused by the constraints placed on the parameters, such as the requirement that all the variograms had the same range.

In Table 11.1, we show the negative log-likelihood and AIC values that result when different LMMs are estimated for the Scottish Borders data. The simplest models only consider copper observations and assume constant fixed effects and a pure nugget variogram. Then, the pure nugget variogram is replaced by a nugget and Matérn model. The third model also replaces the constant fixed effects by ones that vary according to the soil types displayed in Fig. 11.3. The final model considers observations of both copper and cobalt and assumes that each of these properties has constant fixed effects. The models are estimated for both the 50 and 400 observation samples of copper. Both coregionalized models include 500 cobalt observations.

For both sample sizes, the negative log-likelihood decreases upon inclusion of the Matérn variogram function and the nonconstant fixed effects. In the case of the 400 observation samples, the AIC also decreases in the same manner. However, for the 50-point sample, the addition of these extra terms to the model causes the AIC to increase. This indicates that there is insufficient evidence in the 50-point sample

Table 11.1 Number of parameters (p), minimized negative log-likelihood ($-L$) and AIC value for models with different fixed effects (**M**), variogram models, number of copper observations (n_{Cu}) and number of cobalt observations (n_{Co})

				$n_{Cu} = 50$		$n_{Cu} = 400$	
M	Variogram	n_{Co}	p	$-L$	AIC	$-L$	AIC
Constant	Nugget	0	2	-82.25	168.50	-640.26	1284.53
Constant	Matérn	0	5	-81.36	172.72	-594.52	1199.04
Soil type	Matérn	0	8	-78.20	172.39	-590.68	1197.37
Constant	Matérn	500	9	-34.38	86.76	-225.26	468.52

to indicate that copper is spatially correlated or that it varies according to soil type. For both sample sizes, the lowest AIC occurs when the 500 cobalt observations are included in the model. Note that the models of coregionalized variables were estimated by minimizing the negative log-likelihood function which included both properties. However, the negative log-likelihood that is quoted in Table 11.1 is the likelihood of the copper observations given both the estimated parameters and the cobalt observations. This means that the corresponding AIC value is comparable to those from the other three models.

11.4 Bayesian Methods and Variogram Parameter Uncertainty

The model-based methods described in this chapter are compatible with Bayesian methodologies which can be used to quantify the uncertainty of the random effects parameters (Handcock and Stein 1993). Classical statistical methodologies assume that model parameters are fixed. Generally, when applying classical or model-based geostatistics, we look for a single best fitting variogram model and take no account of variogram uncertainty. In Bayesian analyses, model parameters are treated as probabilistic variables. Our knowledge of the parameter values prior to collecting any data is expressed as a prior distribution. Then the observations of the soil property are used to update these priors and to form a posterior distribution which combines our prior knowledge with the information that could be inferred from the observations.

Minasny et al. (2011) demonstrated how Bayesian approaches could be applied to the spatial prediction of soil properties. They placed uniform priors on all of the random effects parameters and then used a Markov chain Monte Carlo (MCMC) simulation approach to sample the multivariate posterior distribution of these parameters. Rather than a single best fitting estimate, this approach led to a series of parameter vectors that were consistent with both the prior distributions and the observed data. The MCMC approach generates a chain of parameter vectors which follow a random walk through the parameter space. The chain starts at some

value of parameters, and the log-likelihood is calculated. Then the parameters are perturbed and the log-likelihood is recalculated. The new parameter values are accepted or rejected based on the difference between the likelihoods before and after the perturbation according to Metropolis-Hastings algorithm (Hastings 1970). If the log-likelihood increases, then the new parameter vector is always accepted. If the likelihood decreases, then the proposal might be accepted. The probability of acceptance decreases as the difference in likelihood increases.

Under some regularity conditions, the set of parameter vectors that result from the Metropolis-Hastings algorithm are known to converge to the posterior distribution of the parameters. However, the algorithm requires careful tuning of some internal settings within the algorithm. These settings particularly relate to the distribution from which a proposed parameter vector is sampled. If this distribution is too wide, then too many of the proposed parameter vectors will be rejected and the chain will remain at its starting point. If the proposal distribution is too narrow, then nearly all of the parameter vectors will be accepted, but the perturbations will be small, and it will take a considerable amount of time to consider the entire parameter space. Therefore Vrugt et al. (2009) developed a DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm to automate the selection of these internal settings and to produce Markov chains that converge efficiently. The DREAM algorithm simultaneously generates multiple Markov chains. The information inferred from acceptances and rejections within each chain is pooled to select efficient proposal distributions. MATLAB and R implementations of the DREAM algorithm are freely available (Vrugt 2016; Guillaume and Andrews 2012).

We used the DREAM algorithm to sample parameters of the nested nugget and Matérn model for both the 50 and 400 observations of ln copper. In each case, we used four chains and sampled a total of 101,000 parameter vectors. The bounds on the uniform prior distributions were zero and one $\{\ln(\mathrm{mg\,kg^{-1}})\}^2$ for c_0 and c_1, zero and 40 km for a and 0.01 and 2.5 for v. For comparison, the variance of the ln copper observations was 0.35. The first 1000 of the sampled vectors were discarded since the MCMC was converging to the portion of parameter space that was consistent with the observed data. This is referred to as the burn-in period. We used the R statistic (see Vrugt et al. 2009) to confirm that the chain had converged. Successive entries of the series of parameter vectors that remained were correlated because the parameter vector was either unchanged or only perturbed a short distance. Therefore every 100th entry of this series was retained. The final series contained 1000 vectors which were treated as independent samples from the posterior distribution of the parameter vector. Figure 11.7 shows the 90% confidence intervals for the variogram of each dataset. These confidence intervals stretch between the 5th and 95th percentiles of the semi-variances for each lag. It is evident that the variogram from the 50 observation sample is uncertain across all lag distances. The uncertainty is greatly reduced for the 400 observation sample.
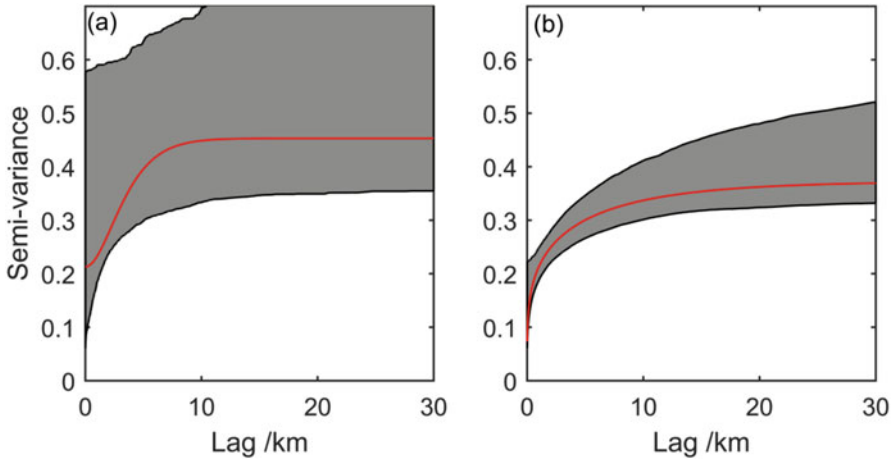
Fig. 11.7 Maximum likelihood estimates of the variogram for (**a**) 50 ln copper concentrations and (**b**) 400 ln Cu concentrations. The *grey* shading indicates the 90% confidence interval for the variogram function according to the MCMC sample of variogram parameters

11.5 Spatial Prediction and Validation of Linear Mixed Models

11.5.1 The Best Linear Unbiased Predictor

Having estimated the fixed and random effects parameters $\widehat{\alpha}$ and $\widehat{\beta}$, we can use the LMM and the observations of the soil property to predict the expected value and uncertainty of the possibly transformed soil property at a set of locations $\mathbf{x}_p$ where it has not been observed. We denote the fixed effects design matrix at these locations by $\mathbf{M}_p$, the matrix of covariances between the random effects of the soil property at the observation and prediction locations by $\mathbf{C}_{po}$ and the random effects covariance matrix at the prediction locations by $\mathbf{C}_{pp}$. These matrices are calculated using the estimated $\widehat{\alpha}$ parameters. The best linear unbiased predictor (BLUP; Lark et al. 2006; Minasny and McBratney 2007) of the expected value of the possibly transformed soil property at the unobserved locations is:

$$\widehat{\mathbf{S}}_p^* = \mathrm{E}\left[\mathbf{s}^*\left(\mathbf{x}_p\right)\right] = \mathbf{M}_p\widehat{\beta} + \mathbf{C}_{po}\mathbf{C}^{-1}\left(\mathbf{s}^* - \mathbf{M}\widehat{\beta}\right), \qquad (11.16)$$

and the corresponding prediction covariance matrix is:

$$\mathbf{V} = \left(\mathbf{M}_p - \mathbf{C}_{po}\mathbf{C}^{-1}\mathbf{M}\right)\left(\mathbf{M}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{M}\right)^{-1}\left(\mathbf{M}_p - \mathbf{C}_{po}\mathbf{C}^{-1}\mathbf{M}\right)^{\mathrm{T}} + \left(\mathbf{C}_{pp} - \mathbf{C}_{po}\mathbf{C}^{-1}\mathbf{C}_{po}^{\mathrm{T}}\right).$$
$$(11.17)$$

The first term in Eq. 11.17 accounts for the uncertainty in predicting the fixed effects, whereas the second term accounts for the uncertainty in predicting the random effects. The elements of the main diagonal of $\mathbf{V}$, which we denote V_{ii}, are the total prediction variances for each site. Since we have assumed that $\mathbf{s}^*$ is Gaussian, we have sufficient information (i.e. the mean and the variance) to calculate the probability density function (pdf) or cumulative density function (cdf) for $\mathbf{s}^*$ at each of the sites. Density functions are discussed in more detail in Sect. 14.2.2. We might calculate the pdf for N equally spaced values of the variable with spacing Δy (i.e. $y_j = j\Delta y$ for $j = 1, \ldots, N$). The formula for the Gaussian pdf is:

$$f_j \equiv f\left(y_j | \widehat{S}_i^*, V_{ii}\right) = \frac{1}{\sqrt{2\pi V_{ii}}} \exp\left\{-\frac{1}{2}\left(\frac{y_j - \widehat{S}_i^*}{\sqrt{V_{ii}}}\right)^2\right\} \qquad (11.18)$$

If the density is zero (to numerical precision) for $y_i < y_i$ and $y_i > y_N$, then the area under the curve f will be one and the f_j will sum to $1/\Delta y$. The cdf can then be deduced from the f_j:

$$F_j \equiv \mathrm{Prob}\left(s_i < y_j\right) = \Delta y \sum_{k=1}^{j} f_k, \qquad (11.19)$$

In Fig. 11.8, we show these pdf and cdf for ln copper concentration at site 'A' based on the maximum likelihood estimate of the LMM for 400 copper observations. The area of the grey-shaded region is equal to the probability that ln copper concentration is negative (i.e. that the concentration of copper is less than $1 \mathrm{\ mg\ kg}^{-1}$). This probability can be more easily extracted from the value of the cdf when ln copper is equal to zero (Fig. 11.8b).
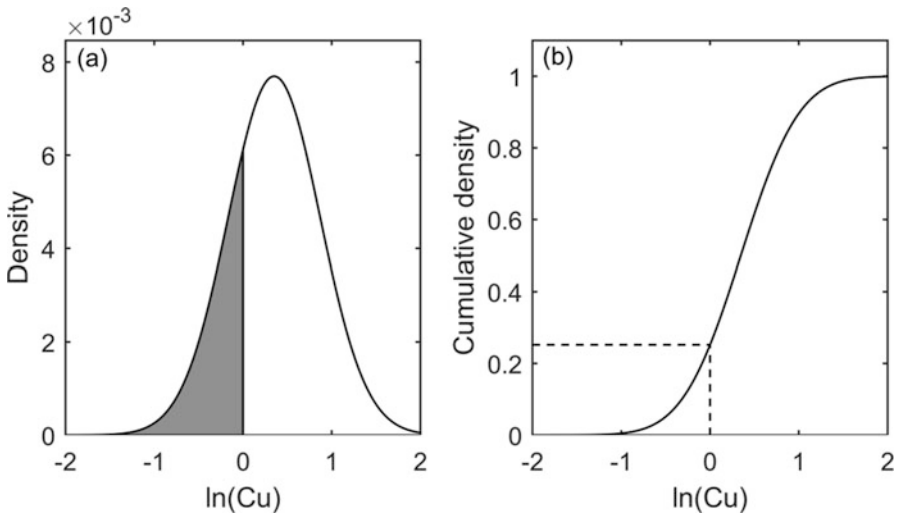

Fig. 11.8 Predicted (a) pdf and (b) cdf for ln copper at location 'A'

If we have a MCMC sample of variogram parameter vectors rather than a single estimate, then we might calculate the pdf for each of these variograms. Then we could calculate a pdf that accounted for variogram uncertainty by averaging these individual pdfs and calculate the cdf using Eq. 11.19. This is an example of the Monte Carlo error propagation method described in Sect. 14.4.2. In Fig. 11.9b, we show the cdf of ln copper at site 'A' based on the 400 observation sample. The grey-shaded region is the 90% confidence interval for this cdf using the MCMC sample of variogram parameters to account for variogram uncertainty. It is apparent that variogram uncertainty does not have a large effect on the cdf. However, when the pdfs are based on the MCMC for 50 observations, a larger effect of variogram uncertainty is evident (Fig. 11.9a). Recall that site 'A' is only 0.6 km from the nearest observation. We will see in Sect. 11.6 of this chapter that such a prediction is sensitive to uncertainty in the variogram parameters. When we repeat the exercise at site 'B' which is 1.3 km from an observation, the effects of variogram uncertainty are small using the MCMC samples based on both 50 and 400 observations.

The BLUP encompasses many of the kriging algorithms described in Chap. 10. For example, when the fixed effects are constant, it performs the role of the ordinary kriging estimator; when covariates are included in the fixed effects, it performs the role of the regression or universal kriging predictor; and when multiple soil properties are included in the observation vector, it performs the role of the co-kriging estimator. Equations 11.16 and 11.17 lead to predictions on the same support as each observation. If we wish to predict the soil property across a block that has a larger support than each observation, then $\mathbf{C}_{po}$ and $\mathbf{C}_{pp}$ should be replaced by $\overline{\mathbf{C}}_{po}$, the covariances between the observations and the block averages, and $\overline{\mathbf{C}}_{pp}$ the covariances between the block averages.

We previously noted that when multiple properties are included in the observation vector of an LMM, the nugget parameters for the cross-variograms can only be estimated if there are co-located observations of the two properties. This parameter will be required in the BLUP if we wish to predict one soil property at the exact location where another one was observed or if we wish to know the covariance between the predictions of the two properties at the same site. However, this parameter is not required to produce maps of each soil property on a regular grid that does intersect the observation locations or to calculate the prediction variances for each property.

If a transformation has been applied to the observations then it will be necessary to back-transform the predictions before they can be interpreted. If we simply calculate the inverse of the transformation for $\widehat{S}^*_{p(i)}$, the prediction of the mean of the transformed property at the ith prediction location, the result is the prediction of the median of the untransformed property. The 0.5 quantile of the transformed property has been converted to the 0.5 quantile of the untransformed property. It is possible to back-transform every quantile of the cdf in this manner. It is generally more difficult to determine the mean of a back-transformed prediction. Instead, the back-transformed mean can be approximated through simulation of the transformed variable. If the mean and variance of the transformed prediction for a
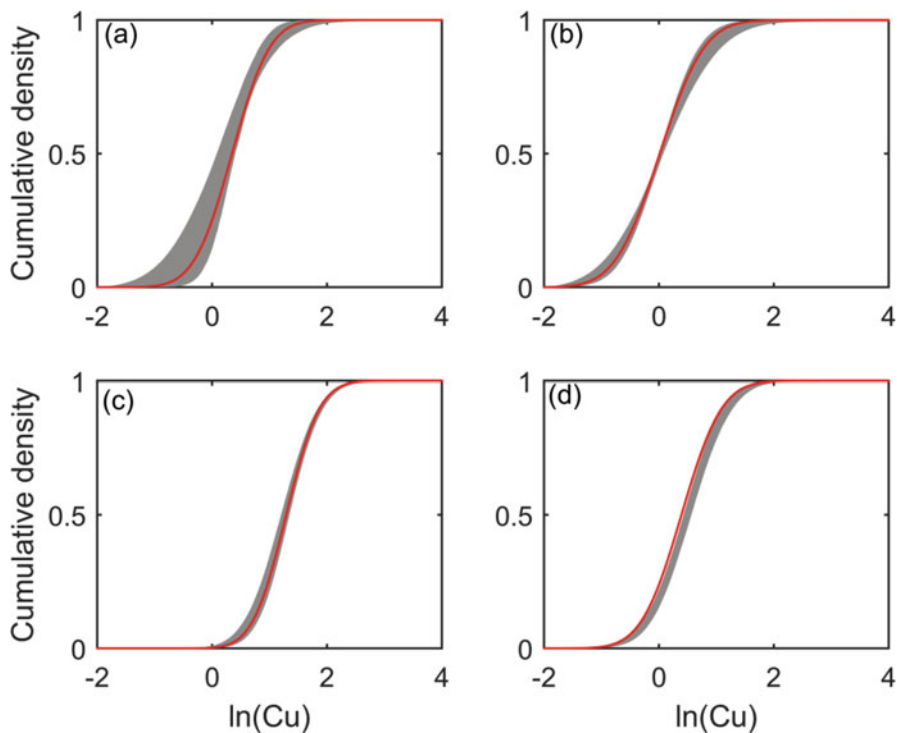
Fig. 11.9 Predicted cdf (*red line*) and 90% confidence interval accounting for variogram uncertainty (*shaded area*) of ln copper concentration at (**a**) site 'A' conditional on the 400 observation samples and the MCMC sample of variogram parameters calibrated on the 50 observation subsamples, (**b**) site 'A' conditional on the 400 observation sample and the MCMC sample of variogram parameters calibrated on the same 400 observations, (**c**) site 'B' conditional on the 400 observation sample and the MCMC sample of variogram parameters calibrated on the 50 observation subsample and (**d**) site 'A' conditional on the 400 observation sample and the MCMC sample of variogram parameters calibrated on the same 400 observations

site are $\widehat{S}^*_{p(i)}$ and $V^*_{p(i)}$, respectively, then one might simulate 1000 realizations of the Gaussian random variable with this mean and variance, apply the inverse transform to each realization and then calculate the mean (or other statistics) of these back-transformed predictions.

11.5.2   Validation of the LMM

It is important to validate an LMM to confirm that the predictions are as accurate as we believe them to be. Close inspection of validation results might also reveal patterns in the model errors that indicate that an additional covariate is required in

the fixed effects or a further generalization is required in the random effects. Ideally, validation should be conducted using a set of data that were not used to calibrate the LMM. However, in some instances, data are sparse, and then there is little choice but to carry out cross-validation. In leave-one-out cross-validation, the model is fitted to all of the measurements, and then one datum, s_i say, is removed, and the remaining data and the BLUP are used to predict the removed observation. The process is repeated for all n observations.

When conducting either cross-validation or validation, we wish to look at both the accuracy of the predictions and the appropriateness of the prediction variances. We can assess the accuracy of predictions by looking at quantities such as the mean error (ME),

$$\mathrm{ME} = \frac{1}{n}\sum\nolimits_{i=1}^{n} \left\{ s_i - \widehat{S}_i \right\}, \tag{11.20}$$

and the root mean squared error (RMSE),

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum\nolimits_{i=1}^{n} \left\{ s_i - \widehat{S}_i \right\}^2 \right]^{\frac{1}{2}}. \tag{11.21}$$

The appropriateness of the prediction variances are often explored by calculating the standardized squared prediction errors at each site:

$$\theta_i = \frac{\left\{ s_i - \widehat{S}_i \right\}^2}{V_i}, \tag{11.22}$$

where V_i is the prediction variance for $\widehat{S}_i$. If, as we expect, the errors follow a Gaussian distribution, then the θ_i will be realized from a chi-squared distribution with one degree of freedom. Pedometricians often calculate the mean $\overline{\theta}$ and median $\tilde{\theta}$ of the θ_i and compare them to their expected values of 1.0 and 0.45 (e.g. Minasny and McBratney 2007; Marchant et al. 2009). If they are properly applied, the ML and REML estimators tend to ensure that $\overline{\theta}$ is close to 1.0. However, although the average standardized error is close to its expected values, the set of θ_i values might not be consistent with the chi-squared distribution. Deviations from this distribution are often indicated by values of $\tilde{\theta}$ that are far from 0.45. If $\tilde{\theta}$ is considerably less than 0.45, this might indicate that the LMM should have a more highly skewed distribution function.

It can also be useful to consider the entire cdf of the standardized errors. If the errors are Gaussian, then the

$$p_i = \Phi_{0,1}^{-1}\left(\frac{s_i - \widehat{S}_i}{\sqrt{V_i}} \right), \tag{11.23}$$

should be uniformly distributed between zero and one where $\Phi_{0,1}^{-1}$ is the inverse cdf for a Gaussian distribution of zero mean and unit variance. We want to confirm that the p_i is consistent with such a uniform distribution. This can be achieved through the inspection of predictive QQ plots (Thyer et al. 2009). These are plots of the n theoretical quantiles of a uniform distribution against the sorted p_i values. If the standardized errors are distributed as we expect, then the QQ plot should be a straight line between the origin and (1,1). If all the points lie above (or, alternatively, below) the 1:1 line, then the soil property is consistently under-/over-predicted. If the points lie below (above) the 1:1 line for small theoretical values of p and above (below) the 1:1 line for large theoretical values of p, then the predictive uncertainty of the MM is under-/overestimated. Alternatively, Goovaerts (2001) uses accuracy plots rather than QQ plots. In these plots, the [0,1] interval is divided into a series of bins bounded by the $(1 - p)/2$ and $(1 + p)/2$ quantiles for p between 0 and 1. The two plots differ in that each bin of the accuracy plot is symmetric about 0.5. Therefore it is not possible to consider the upper and lower tail of the distribution separately. In contrast, with QQ plots, we can see how well the left-hand tail of the distribution is approximated by looking close to the origin of the plot, and we can examine how the right-hand tail is approximated by looking close to (1,1).

11.5.3   Predicting Copper Concentrations in the Scottish Borders

Figure 11.10 shows maps of the expectation of ln copper across the study region for models listed in Table 11.1. The model with a pure nugget variogram is not included since the predictions are constant. The map based on 50 copper observations and with constant fixed effects is much less variable than the other predictions. More features of the copper variation are evident when the soil type is added to the fixed effects. However, there are obvious and possibly unrealistic discontinuities in the predictions at the boundaries between different soil types. The model including cobalt observations retains the detail of the variable fixed-effect model, but there are no discontinuities. Further detail is added to all of the maps when 400 rather than 50 copper observations are used. The hotspots of copper that are evident occur close to urban centres.

The most striking feature in the validation results (Tables 11.2 and 11.3) is the difference in performance for the models based on 50 observations of copper and those based on 400 observations. When 400 observations are used, the cross-validation and validation RMSEs are very similar and smaller than those for the 50-point sample. For the 50-observation sample, there is also a greater difference between the cross-validation and validation results indicating that these models might well have been overfitted. For both sample sizes, the largest RMSEs occur for the pure nugget model, but there is little difference in the RMSEs for the other three LMMs. The prediction variances for the LMMs calibrated on 400 observations
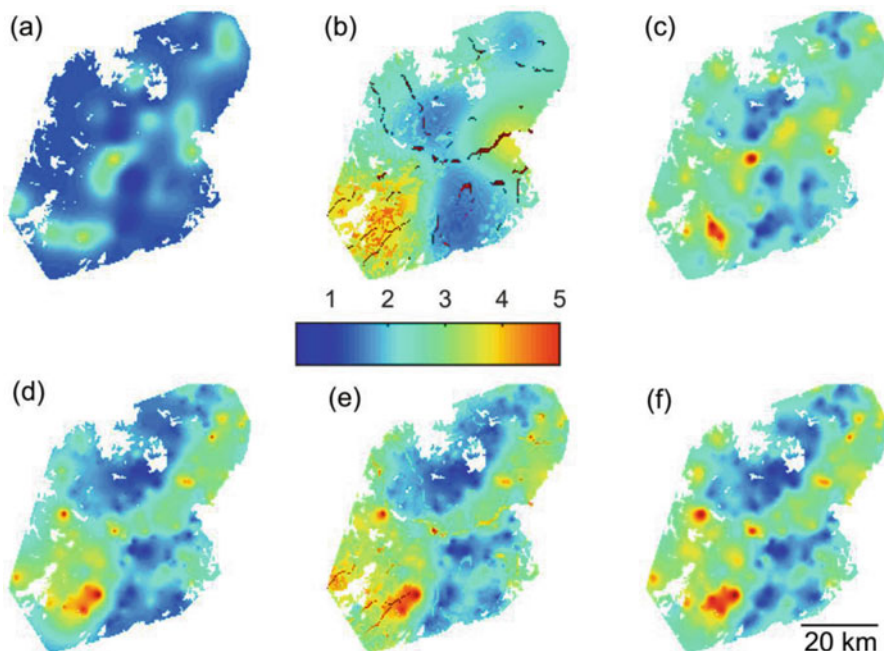
Fig. 11.10 Median prediction of copper concentration (mg/kg) across the study region from (**a**) BLUP using 50 Cu calibration data with constant fixed effects, (**b**) BLUP using 50 Cu calibration data with fixed effects varying according to soil type, (**c**) BLUP using 50 Cu calibration data and 500 Co calibration data represented by an LMCR with constant fixed effects, (**d**) BLUP using 400 Cu calibration data with constant fixed effects, (**e**) BLUP using 400 Cu calibration data with fixed effects varying according to soil type and (**f**) BLUP using 400 Cu observations and 500 Co observations represented by an LMCR with constant fixed effects

also appear to be more accurate than those based on the 50 observation sample. The mean standardized prediction errors upon cross-validation and validation for all models calibrated on 400 copper observations are between 0.98 and 1.00, and the median standardized prediction errors range between 0.38 and 0.47. Hence both quantities are close to their expected values of 1.0 and 0.45. In the case of the 50 observation samples, the ranges of these statistics are wider, stretching between 0.78 and 1.34 for the mean and 0.32 and 0.67 for the median. The cross-validation standardized prediction errors are closer to their expected values than the corresponding validation values. Again, this is an indication of overfitting when only using 50 observations. There is also evidence of overfitting in the QQ plot for the LMM with constant fixed effects and calibrated on 50 observations (Fig. 11.11a). The validation plot deviates a large distance from the 1:1 line. All of the other QQ plots more closely follow the 1:1 line although some improvement in using the larger sample size is evident.

Table 11.2 Mean error (ME) and root mean squared error (RMSE) upon leave-one-out cross-validation and validation of models for ln copper with different fixed effects (**M**), variogram models, number of copper observations (n_{Cu}) and number of cobalt observations (n_{Co})

				Calibration		Validation	
M	Variogram	n_{Cu}	n_{Co}	ME	RMSE	ME	RMSE
Constant	Nugget	50	0	0.000	0.68	0.092	0.60
Constant	Matérn	50	0	0.009	0.66	0.045	0.74
Soil type	Matérn	50	0	0.011	0.64	0.023	0.58
Constant	Matérn	50	500	−0.001	0.65	0.048	0.54
Constant	Nugget	400	0	0.000	0.60	0.028	0.59
Constant	Matérn	400	0	0.002	0.50	0.063	0.51
Soil type	Matérn	400	0	0.002	0.50	0.011	0.50
Constant	Matérn	400	500	0.002	0.50	0.013	0.50

Table 11.3 Mean standardized prediction errors $(\overline{\theta})$, median standardized prediction errors $(\tilde{\theta})$ and mean error in QQ plot $(\overline{Q_e})$ upon leave-one-out cross-validation and validation of models for ln copper with different fixed effects (**M**), variogram models, number of copper observations (n_{Cu}) and number of cobalt observations (n_{Co})

				Calibration			Validation		
M	Variogram	n_{Cu}	n_{Co}	$(\overline{\theta})$	$(\tilde{\theta})$	$\overline{Q_e}$	$(\overline{\theta})$	$(\tilde{\theta})$	$\overline{Q_e}$
Constant	Nugget	50	0	1.02	0.61	0.000	0.80	0.38	0.040
Constant	Matérn	50	0	1.03	0.49	−0.007	1.34	0.67	0.197
Soil type	Matérn	50	0	1.08	0.62	−0.022	0.90	0.38	0.012
Constant	Matérn	50	500	1.03	0.33	−0.003	0.78	0.32	0.019
Constant	Nugget	400	0	1.00	0.47	−0.002	0.99	0.45	0.012
Constant	Matérn	400	0	1.00	0.42	0.003	0.99	0.39	0.028
Soil type	Matérn	400	0	1.00	0.41	0.002	1.00	0.38	−0.001
Constant	Matérn	400	500	1.00	0.44	0.003	0.98	0.39	0.000

11.6 Optimal Sample Design

It is often costly to obtain soil samples from a study area and then analyse them in the laboratory to determine the properties of interest such as the concentrations of cobalt and copper in the Scottish Borders survey. Therefore, it can be important to optimize the locations where the samples are extracted from so that an adequate spatial model or map can be produced for the minimum cost. Many spatial surveys have been conducted using a regular grid design (see examples in Webster and Oliver 2007) since this ensures that the observations are fairly evenly distributed across the study region and that the kriging variances are not unnecessarily large at any particular locations. Also, it is relatively easy to apply the method of moments variogram estimator to a grid-based sample since the variogram lag bins can be selected according to the grid spacing. Some authors (e.g. Cattle et al. 2002) have
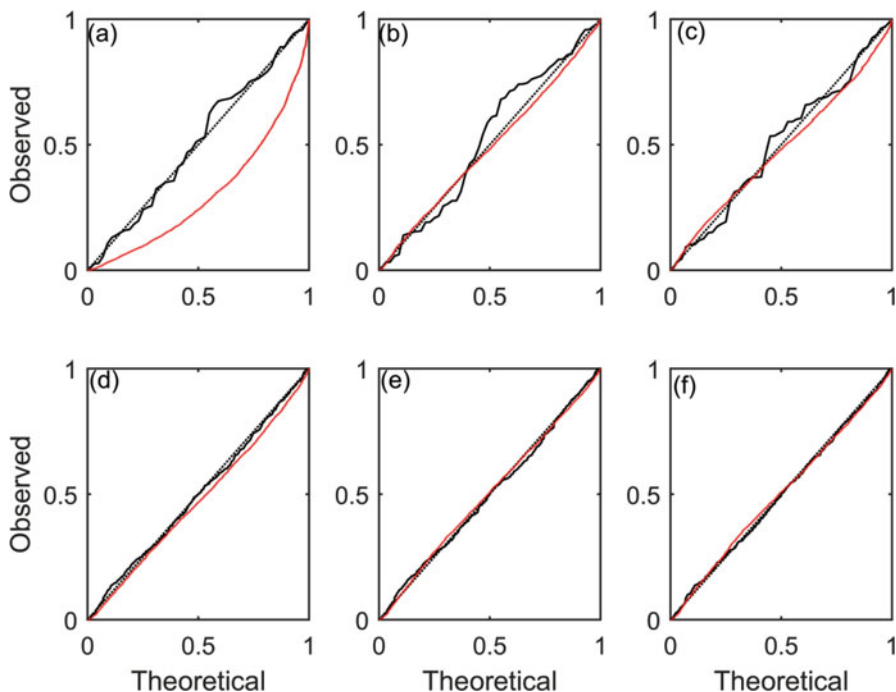
Fig. 11.11 QQ plots resulting from leave-one-out cross-validation of the calibration copper data (*black line*) and validation of the validation copper data (*red line*). The predictors and observations for plots (**a**–**f**) are identical to those in Fig. 11.10

included close pairs of observations in their survey designs since these ensure that the variogram can be reliably estimated for short lag distances. However, the decision as to how much of the sampling effort should be allocated to even coverage of the study region and how much to estimating the spatial model over short distances is often made in a subjective manner.

The kriging or prediction variance (Eq. 11.17) can form the basis of a more objective criterion for the efficient design of spatial surveys. If the random effects are second-order stationary and Gaussian, then the prediction variance does not depend on the observed data. If the variogram model is assumed to be known, then prior to making any measurements, it is possible to use Eq. 11.17 to assess how effective a survey with a specified design will be. Alternatively, we can select the configuration of a specified number of sampling locations that lead to the smallest prediction variance.

Van Groenigen et al. (1999) suggested an optimization algorithm known as spatial simulated annealing (SSA) to perform this task. The n samples are initially positioned randomly across the study region and the corresponding kriging variance (or any other suitable objective function) is calculated. Then the position of one of these samples is perturbed a random distance in a random direction. The kriging

variance is recalculated. If it has decreased, then the perturbation is accepted. If the kriging variance has increased, then the perturbation might be accepted. In common with the DREAM algorithm, described in Sect. 11.4 of this chapter, the probability of acceptance decreases with the magnitude of the increase in the objective function according to the Metropolis-Hastings algorithm (Hastings 1970). However, in contrast to the DREAM algorithm, the probability of acceptances also decreases as the optimization algorithm proceeds. The potential to accept perturbations that increase the kriging variance is included to ensure that the optimizer does not converge to a local rather than global minimum. The gradual decrease in the probability of such an increase ensures that a minimum is eventually reached. If a perturbation is rejected, then the sample is returned to its previous location. The SSA algorithm continues, perturbing each point in turn until the objective function settles to a minimum.

We illustrate the application of the SSA algorithm by optimizing 50 location sample schemes for the Scottish Borders study area in Fig. 11.12. In all of the plots within this figure, we consider a soil property where the random effects are realized from a Gaussian random function with a nested nugget and Matérn covariance function with $c_0 = 0.5$, $c_1 = 0.5$, $a = 10$ km and $v = 0.5$. The kriging variance is the objective function for the designs shown in Fig. 11.12a, b. In plot (a), the fixed effects are assumed to be constant. The optimized sample locations are spread evenly across the region. The locations are randomly allocated to the four different soil types (Fig. 11.12d) with the majority of samples being situated in the most prevalent brown earth class. When the fixed effects are assumed to vary according to soil type, the optimized samples are still fairly evenly distributed across the region. However, the number of locations that are situated in the less prevalent soil classes increases (Fig 11.12e). This ensures that a reasonably accurate estimate of the fixed effects can be calculated for each soil type. When Brus and Heuvelink (2007) optimized sample schemes for universal kriging of a soil property with an underlying trend that was proportional to a continuous covariate, they found that the soil was more likely to be sampled at sites where this covariate was particularly large or particularly small. This ensured that the gradient of the trend function could be reasonably accurately estimated.

In addition to spatial prediction, the set of observed samples should also be suitable for estimating the spatial model. Therefore, Marchant and Lark (2007b) and Zhu and Stein (2006) expanded the objective function to account for uncertainty in estimating the spatial model or variogram. These authors noted that for a linear predictor such as the BLUP,

$$\widehat{S}_{p(k)} = \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{s}, \tag{11.24}$$

where $\widehat{S}_{p(k)}$ is the prediction of the expectation of S at the ith prediction site, $\boldsymbol{\lambda}$ is a length n vector of weights and $\mathbf{s}$ is a length n vector of observations; the extra contribution to the prediction variance resulting from variogram uncertainty could be approximated by a Taylor series:
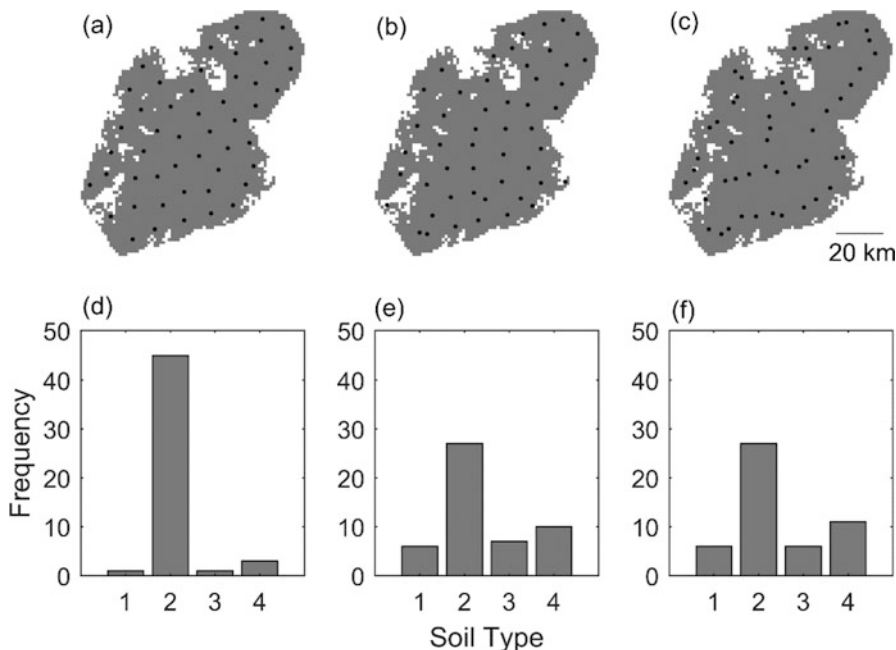
Fig. 11.12 Optimized sampling locations where the objective function is (**a**) the kriging variance with no fixed effects, (**b**) the kriging variance with fixed effects varying according to soil type and (**c**) the kriging variance plus the prediction variance due to variogram uncertainty with fixed effects varying according to soil type. Plots (**d–f**) show the distribution of sampling locations amongst soil types for the design above. The soil types are (1) alluvial soils, (2) brown earths, (3) peaty podzols and (4) mineral gleys

$$\tau^2 = \mathrm{E}\left[\left\{\widehat{S}_{\mathrm{p}(k)}(\widehat{\boldsymbol{\alpha}}) - \widehat{S}_{\mathrm{p}(k)}(\boldsymbol{\alpha})\right\}^2\right] = \sum_{i=1}^{r}\sum_{j=1}^{r}\mathrm{Cov}\left(\alpha_i, \alpha_j\right)\frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial \alpha_i}\mathbf{C}\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_j},$$

$$(11.25)$$

where τ^2 is the expected squared difference between the prediction at the ith site using the actual variogram parameters $\boldsymbol{\alpha}$, and the prediction at this site using the estimated parameters, $\widehat{\boldsymbol{\alpha}}$; r is the number of variogram parameters and $\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_i}$ is the length n vector containing partial derivatives of the $\boldsymbol{\lambda}$ weights matrix with respect to the ith variogram parameter. The covariances between the variogram parameters can be approximated using the Fisher information matrix $\mathbf{F}$ (Marchant and Lark 2004):

$$\mathrm{Cov}\left(\alpha_i, \alpha_j\right) \approx \left[\mathbf{F}^{-1}\left(\boldsymbol{\alpha}\right)\right]_{ij}; \ \left[\mathbf{F}\left(\boldsymbol{\alpha}\right)\right]_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \alpha_j}\right] \quad (11.26)$$

Here, $[]_{ij}$ denotes element i, j of the matrix inside the brackets, $\frac{\partial \mathbf{C}}{\partial \alpha_i}$ is the $n \times n$ matrix of partial derivatives of the covariance matrix $\mathbf{C}$ with respect to α_i and Tr

denotes the trace or sum of elements on the main diagonal of the matrix that follows. The $\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_i}$ can be calculated using a numerical approximation (e.g. Zhu and Stein 2006) although Marchant and Lark (2007b) noted the standard linear algebra relationship that if

$$\mathbf{L} = \mathbf{A}^{-1}\mathbf{b}, \tag{11.27}$$

then

$$\frac{\partial \mathbf{L}}{\partial \alpha_i} = \mathbf{A}^{-1}\left(\frac{\partial \mathbf{b}}{\partial \alpha_i} - \frac{\partial \mathbf{A}}{\partial \alpha_i}\mathbf{A}^{-1}\mathbf{b}\right). \tag{11.28}$$

Thus, this equation can be used to exactly calculate the $\frac{\partial \boldsymbol{\lambda}}{\partial \alpha_i}$ matrices for the universal kriging predictor formulation of the BLUP which is written in the form of Eq. 11.27 with:

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\varphi} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{M} \\ \mathbf{M}^{\mathrm{T}} & \mathbf{0}_{q,q} \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} \mathbf{C}_{\mathrm{p}(k)\mathrm{o}}^{\mathrm{T}} \\ \mathbf{M}_{\mathrm{p}(k)} \end{bmatrix} \tag{11.29}$$

where φ is the length q vector of Lagrange multipliers, $\mathbf{0}_{q,q}$ is a $q \times q$ matrix of zeros and $\mathbf{C}_{\mathrm{p}(i)\mathrm{o}}$ is the ith row of matrix $\mathbf{C}_{\mathrm{po}}$. For most authorized covariance functions, the elements of the $\frac{\partial \mathbf{C}}{\partial \alpha_i}$ that are required to calculate Eq. 11.26 and Eq. 11.28 can be determined exactly. However, in the case of the Matérn function, numerical differentiation remains the most practical method to calculate $\frac{\partial \mathbf{C}}{\partial \nu}$.

Figure 11.12c shows an optimized 50-location design that results when the objective function is the prediction variance (assuming that the fixed effects vary according to soil type) plus the τ^2. In this case, the sampling locations are less evenly spread across the study region. Short transects of close locations are evident which are suitable for estimating the spatial covariance function over small lags. The prediction variance and τ^2 for this design are mapped in Fig. 11.13a, b, respectively. The prediction variance is smallest close to sampling locations and increases for locations where there are no nearby samples and that are at the boundary of the region. The additional component of uncertainty because of the estimation of the spatial model is largest close to sample locations and decreases at locations where there are no nearby samples.

As previously discussed in Sect. 10.2, there is one obvious flaw in this strategy to optimize sample designs. In reality, the spatial model is not known prior to sampling, and therefore, it is not possible to calculate the prediction variance. A spatial model must be assumed, perhaps using the results of previous surveys of the soil property at similar locations or based on a reconnaissance survey (e.g. Marchant and Lark 2006). Alternatively, a prior distribution might be assumed for each covariance function parameter and the objective function averaged across these distributions (Diggle and Lophaven 2006). Also, the prediction variance cannot be calculated
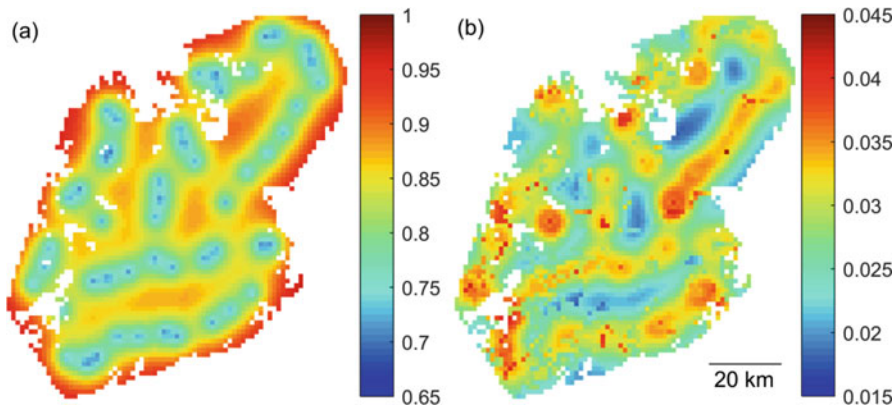
Fig. 11.13 (**a**) The kriging variance with fixed effects varying according to soil type for the optimized sample design in Fig. 11.12c. (**b**) The contribution to the prediction variance due to variogram uncertainty for the same sample design

prior to sampling if the property of interest has a skewed distribution, since the prediction variances vary according to the observed data. Marchant et al. (2013a) demonstrated how simulation could be used to optimise a multi-phase survey in these circumstances.

11.7   Conclusions

The use of model-based rather than classical geostatistical methods removes many of the subjective decisions that are required when performing geostatistical analyses. The LMM is flexible enough to incorporate linear relationships between the soil property of interest and available covariates and to simultaneously represent the spatial variation of multiple coregionalized soil properties. The log-likelihood is an objective function that can be used to compare proposed model structures and parameters. Since a multivariate distribution function is specified in the model, it is possible to predict the complete pdf or cdf of the soil property at an unsampled location. From these predicted functions, one can easily determine the probability that the soil property exceeds a critical threshold (e.g. Li et al. 2016).

The primary disadvantages of the model-based methods are the time required to compute the log-likelihood and the requirement to specify the multivariate distribution function from which the soil observations were realized. Maximum likelihood estimation of an LMM for more than 1000 observations is likely to take several hours rather than the seconds or minutes required to estimate a variogram by the method of moments. However, we have seen that the uncertainty in LMM parameters estimated from a 400 observation sample have little effect on the uncertainty of the final predictions. Therefore, when estimating an LMM for a large number of observations, it is reasonable to subsample the data. The complete dataset

should be used when using the BLUP. Stein et al. (2004) suggested an approximate maximum likelihood estimator. This uses all of the data but is faster to compute because it ignores some of the covariances between observations. We have described how the standard assumptions that the random effects of an LMM are realized from a multivariate Gaussian random function can be relaxed. A Box-Cox or natural log-transform can be applied to skewed observations so that their histogram more closely approximates that of a Gaussian distribution. Alternatively, it is possible to write the log-likelihood function in a different form that is compatible with any marginal distribution function (Marchant et al. 2011). The assumption that the expected value of the random variable is constant can be relaxed via the fixed effects of the LMM. Some authors have also explored strategies to permit the variability of the random variable to be related to a covariate (e.g. Lark 2009; Marchant et al. 2009; Haskard et al. 2010).

The model-based methods described in this chapter are also compatible with the modelling of space-time variation (e.g. Heuvelink and van Egmond 2010). However, space-time models do require more flexible models for the covariance function because the pattern of temporal variation is likely to be quite distinct from the temporal variation. De Cesare et al. (2001) review the covariance functions that are commonly used for this purpose. Such models might also be used to represent the spatial variation of soil properties in three dimensions when the vertical variation is quite different to the horizontal variation (e.g. Li et al. 2016). The same models could be estimated by classical methods. However, these models contain multiple variogram structures, and it is challenging to appropriately select the different lag bins and fitting weights required for each of these.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akadémiai Kiadó, Budapest, pp 267–281

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

Cattle JA, McBratney AB, Minasny B (2002) Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. J Environ Qual 31:1576–1588

De Cesare L, Myers DE, Posa D (2001) Estimating and modelling space-time correlation structures. Stat Probab Lett 51:9–14

Diggle PJ, Lophaven S (2006) Bayesian geostatistical design. Scand J Stat 33:53–64

Diggle PJ, Ribeiro PJ (2007) Model-based geostatistics. Springer, New York

Goovaerts P (2001) Geostatistical modelling of uncertainty in soil science. Geoderma 103:3–26

Goovaerts P, Webster R (1994) Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur J Soil Sci 45:79–95

Guillaume J, Andrews F (2012) Dream: DiffeRential evolution adaptive metropolis. R package version 0.4–2. http://CRAN.R-project.org/package=dream

Handcock MS, Stein ML (1993) A Bayesian-analysis of kriging. Technometrics 35:403–410

Haskard KA, Welham SJ, Lark RM (2010) Spectral tempering to model non-stationary covariance of nitrous oxide emissions from soil using continuous or categorical explanatory variables at a landscape scale. Geoderma 159:358–370

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Heuvelink GBM, van Egmond FM (2010) Space-time geostatistics for precision agriculture: a case study of NDVI mapping for a Dutch potato field. In: Oliver MA (ed) Geostatistical applications for precision agriculture. Springer, Dordrecht, pp 117–137

Lark RM (2009) Kriging a soil variable with a simple non-stationary variance model. J Agric Biol Environ Stat 14:301–321

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

Li HY, Marchant BP, Webster R (2016) Modelling the electrical conductivity of soil in the Yangtze delta in three dimensions. Geoderma 269:119–125

Marchant BP, Lark RM (2004) Estimating variogram uncertainty. Math Geol 36:867–898

Marchant BP, Lark RM (2006) Adaptive sampling and reconnaissance surveys for geostatistical mapping of the soil. Eur J Soil Sci 57:831–845

Marchant BP, Lark RM (2007a) Estimation of linear models of coregionalization by residual maximum likelihood. Eur J Soil Sci 58:1506–1513

Marchant BP, Lark RM (2007b) Optimized sample schemes for geostatistical surveys. Math Geol 39:113–134

Marchant BP, Newman S, Corstanje R, Reddy KR, Osborne TZ, Lark RM (2009) Spatial monitoring of a non-stationary soil property: phosphorus in a Florida water conservation area. Eur J Soil Sci 60:757–769

Marchant BP, Saby NPA, Jolivet CC, Arrouays D, Lark RM (2011) Spatial prediction of soil properties with copulas. Geoderma 162:327–334

Marchant BP, McBratney AB, Lark RM, Minasny B (2013a) Optimized multi-phase sampling for soil remediation surveys. Spat Stat 4:1–13

Marchant BP, Rossel V, Webster R (2013b) Fluctuations in method-of-moments variograms caused by clustered sampling and their elimination by declustering and residual maximum likelihood estimation. Eur J Soil Sci 64:401–409

Mardia KV, Marshall RJ (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 72:135–146

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, McLaren RG, Spiers RB (1982) Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland. Agronomie 2:969–982

Minasny B, McBratney AB (2005) The Matérn function as a general model for soil variograms. Geoderma 128:192–207

Minasny B, McBratney AB (2007) Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140:324–336

Minasny B, Vrugt JA, McBratney AB (2011) Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation. Geoderma 163:150–162

Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313

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

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30:683–691

Rawlins BG, Marchant BP, Smyth D, Scheib C, Lark RM, Jordan C (2009) Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. Eur J Soil Sci 60:44–54

Ribeiro Jr PJ, Diggle PJ (2001) GeoR: a package for geostatistical analysis. R-News 1(2), ISSN 1609-3631

Soil Survey of Scotland Staff (1984) Organisation and methods of the 1:250 000 soil survey of Scotland, Handbook 8. The Macaulay Institute for Soil Research, Aberdeen

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

Stein ML, Chi Z, Welty LJ (2004) Approximating likelihoods for large spatial data sets. J R Stat Soc B 66:275–296

Thyer M, Renard B, Kavetski D, Kuczera G, Franks S, Srikanthan S (2009) Critical evaluation of parameter consistency and predictive modelling in hydrological modelling: a case study using Bayesian total error analysis. Water Resour Res 45:W00B14

Van Groenigen JW, Siderius W, Stein A (1999) Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87:239–259

Vrugt JA (2016) Markov chain Monte Carlo simulation using DREAM software package: theory, concepts, and Matlab implementation. Environ Model Softw 75:273–316

Vrugt JA, ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D (2009) Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int J Nonlinear Sci Numer Simul 10:273–290

Webster R (1994) The development of pedometrics. Geoderma 62:1–15

Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester

Zhu Z, Stein ML (2006) Spatial sampling design for prediction with estimated parameters. J Agric Biol Environ Stat 11:24–44