Earth Surface Processes and Landforms
Earth Surf. Process. Landforms **31**, 862–874 (2006)
Published online 22 March 2006 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/esp.1296

Geostatistical mapping of geomorphic variables in the presence of trend

R. M. Lark* and R. Webster
Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK

**Correspondence to:*
R. M. Lark, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK.
E-mail: murray.lark@bbsrc.ac.uk

Received 8 March 2005;
Revised 4 June 2005;
Accepted 6 September 2005

Abstract

Mapping geomorphic variables geostatistically, specifically by kriging, runs into difficulties when there is trend. The reason is that the variogram required for the kriging must be of residuals from any trend, which in turn cannot be estimated optimally by the usual method of trend surface analysis because the residuals are correlated. The difficulties can be overcome by the use of residual maximum likelihood (REML) to estimate both the trend and the variogram of the residuals simultaneously.

We summarize the theory of REML as it applies to kriging in the presence of trend. We present the equations to show how estimates of the trend are combined with kriging of residuals to give empirical best linear unbiased predictions (E-BLUPs). We then apply the method to estimate the height of the sub-Upper-Chalk surface beneath the Chiltern Hills of southeast England from 238 borehole data. The variogram of the REML residuals is substantially different from that computed by ordinary least squares (OLS) analysis. The map of the predicted surface is similar to that made from kriging with the OLS variogram. The variances, however, are substantially larger because (a) they derive from a variogram with a much larger sill and (b) they include the uncertainty of the estimate of the trend.

Keywords: geomorphology; stratigraphy; Chalk; Chiltern Hills; trend; geostatistics; REML

Introduction

For many years geomorphologists have been mapping forms of land surface and subsurface features from punctual data. Trend-surface analysis, i.e. straightforward regression on spatial coordinates as predictors, enjoyed a spell of popularity. Then it was realized that the residuals were spatially correlated, and so one of the assumptions of regression, namely independence of the residuals, did not hold. Further, if the surface was at all complex then one would have to use a high-order polynomial or some kind of periodic function to represent it. Otherwise the result might be an excessively smooth version of reality, and one could lose detail of interest.

Along came geostatistics and interpolation by kriging; they appeared to embody solutions to the outstanding problems. If there was no evident trend then variation could be modelled as the outcome of an intrinsically stationary random process and represented by a variogram, as described by Oliver and Webster (1986). Interpolation by ordinary or simple kriging would follow. Very short-range variation would be averaged out, leaving a relatively smooth surface that was of interest. In the last 20 years ordinary kriging has found application in almost all branches of environmental science where investigators wanted to map spatial variation.

If trend were present then it could be incorporated into the kriging model. Matheron (1969) originally called the technique ‘universal kriging’. As Webster and Burgess (1980) remarked, it is far from universal, and it is now more often termed ‘kriging with trend’. Stein (1999) has pointed out that universal kriging is equivalent to the best linear unbiased prediction of a variable (BLUP) when the local expectation is determined by the specified model of the trend. Because the variogram is unknown, but is estimated from the data and the estimate inserted into the BLUP, a kriging prediction is strictly the empirical BLUP or E-BLUP.

Like regression, universal kriging recognizes that the variation has two components, namely the trend and the residuals from the trend. It differs from regression in that it takes into account the dependence in the residuals, which

it treats as spatially correlated stationary random variables. So the residuals have a variogram, and the kriging systems draw their entries from this variogram.

There is a serious difficulty to be overcome, and that is obtaining a model for the variogram. Olea (1975) described in detail a structural analysis from which to obtain the experimental variogram. He listed algorithms, and it was from these that Webster and Burgess (1980) wrote their computer programs. Models can then be fitted to the experimental values. The technique relies on having data at regular intervals on grids or transects, however, and geomorphologists typically do not have such data; rather their data derive from observations irregularly distributed in the landscape and from boreholes. They are not alone in this respect; miners and petroleum engineers are usually in the same situation.

If there is some simple recognizable long-range trend then one legitimate way forward is to compute and model the variogram in the direction perpendicular to the trend. The model will represent the residual random component, on the assumption that this is isotropic, and it can be used in the universal kriging.

Another way to deal with trend is to model it first, as in trend-surface analysis, and remove it from the data. The residuals from the trend are treated as stationary correlated random variables; their variogram is computed and modelled and then used to krige. Finally, the trend is added back to the kriged estimates. Moffat *et al*. (1986) took this approach to map the height of several stratigraphic horizons in the Chalk beneath the Chiltern Hills of southern England. The method is attractive, especially where the trend is interesting in its own right, as in the example above. The estimates, both of the trend and of the random residuals, are unbiased provided that the data are unbiased in the first place.

This method has been widely used in earth sciences under the title 'regression kriging' (e.g. Knotters *et al*., 1995; Odeh *et al*., 1994, 1995). It is equivalent to universal kriging, given the choice of variogram, provided that the universal kriging uses all the data and not just those in a local window.

The disadvantages of regression kriging are twofold. First, the trend is generally estimated by ordinary least squares, which is unbiased, but does not yield estimates of minimum variance unless the sampling sites have been selected independently at random, which is unlikely in geomorphology. Ordinary least squares requires the assumption that the residuals are independent random variables. This assumption can be justified only if the sampling design ensures that the observation points are chosen independently of one another and at random. If this is not the case, and it rarely is, then methods of analysis other than ordinary least squares must be used.

The second disadvantage of regression kriging is that the estimates of the variogram obtained from residuals from the trend are biased. This is because they depend non-linearly on the trend parameters (called 'nuisance' parameters in this context) which are themselves estimated with error. As a result the variogram is underestimated, and the bias increases with increasing lag distance (Cressie, 1993).

One proposed solution to these problems is to use generalized least squares to estimate the trend parameters. The generalized least squares method itself requires a variogram for the residuals, so an iterative procedure is followed. The ordinary least squares estimates are obtained, and a variogram is fitted to the residuals. This variogram is then used in generalized least squares to re-estimate the trend parameters, and the procedure is repeated until the estimates change little or not at all (e.g. Hengl *et al*., 2004). This approach reduces the error variance of the trend parameters, but it does not remove the bias from the estimates in the variogram because these still depend on nuisance parameters (Gambolati and Galeati, 1987). This bias might not matter if data are dense because it is typically very small at short lag distances, and only data at such short distances from target points or blocks carry appreciable weight in the kriging systems.

As a final observation, even if we ignore the bias of the prediction variances of both the trend and the kriging from the residuals, the simple regression kriging procedure does not allow us to combine them into a valid prediction variance for the regression kriging estimate, although we could compute the universal kriging variance, as do Hengl *et al*. (2004).

Another way of dealing with trend is to use the intrinsic random functions of order k, also devised by Matheron (1973). These effectively filter the trend from the data by taking successive differences between neighbouring data. The method requires generalized covariances; these are calculated from the data, which must be at regular intervals on transects or grids. So, again the method is severely limited in geomorphology.

In summary, to map the spatial variation of geomorphic variables that have both pronounced spatial trend and spatially dependent residual random variation requires a method to estimate the variogram of the latter without bias, to obtain minimum-variance estimates of the trend, and to estimate the sum of the trend and the random variation at unsampled sites with known prediction variance. A practical solution to this problem is to compute the empirical best linear unbiased predictor (E-BLUP) using a variogram estimated by residual maximum likelihood (REML). We describe the principles of REML–E-BLUP in what follows, and then illustrate them with data that Moffat *et al*. (1986) previously analysed by regression kriging.

Theory

Let us start on familiar ground. Let us assume that we have a regionalized variable, a variable expressed as a function of spatial co-ordinates that is spatially dependent but has no trend. We can represent it mathematically by

$$Z(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}) \tag{1}$$

in which $Z(\mathbf{x})$ denotes the random variable at a position $\mathbf{x} \equiv \{x_1, x_2\}$ in the two lateral dimensions of a land surface, μ is the mean of $Z(\mathbf{x})$, a constant, and $\varepsilon(\mathbf{x})$ is a spatially dependent component with zero mean everywhere. We assume that ε is second-order stationary. By this we mean that, in addition to the constant mean, the covariance matrix for the set $\{\varepsilon(\mathbf{x}_1), \varepsilon(\mathbf{x}_2), \ldots, \varepsilon(\mathbf{x}_n)\}$ is identical to that for the set $\{\varepsilon(\mathbf{x}_1 + \mathbf{h}), \varepsilon(\mathbf{x}_2 + \mathbf{h}), \ldots, \varepsilon(\mathbf{x}_n + \mathbf{h})\}$ for any vector $\mathbf{h}$, the lag in distance and direction. For a second-order stationary random variable the structure of this or any other such covariance matrix depends on the variogram:

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}[\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2] \quad \text{for all } \mathbf{h} \tag{2}$$

In words, the quantity $\gamma(\mathbf{h})$ is the expected squared difference between values of Z at pairs of places $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ separated by the lag $\mathbf{h}$. It depends on $\mathbf{h}$ and only on $\mathbf{h}$. As a function of $\mathbf{h}$ it is the variogram. If variation is isotropic, i.e. the same in all directions, then we can replace $\mathbf{h}$ by a scalar, $h = |\mathbf{h}|$, so that the variogram is a function of lag distance only.

The variogram of a second-order stationary variable is bounded by a 'sill'. The extent of dependence may be limited by a finite range. Alternatively the sill may be approached asymptotically as the lag distance increases. These two situations are exemplified in two popular variogram models, the spherical and the exponential. Their formulae are as follows.

Spherical:

$$\begin{aligned}\gamma(h) &= c\left\{\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right\} \quad \text{for } h < a \\ &= c \quad \text{for } h \geq a\end{aligned} \tag{3}$$

in which c is the sill and a is the range.

Exponential:

$$\gamma(h) = c\left\{1 - \exp\left(-\frac{h}{r}\right)\right\} \tag{4}$$

in which c is again the sill and r is a distance parameter. As above, the model does not have a finite range, but an effective range is often taken as $a' = 3r$ at which point the variogram has reached $0{\cdot}95c$.

There is often an additional component of variation which is independent at lag intervals greater than or equal to the basic sampling interval. This nugget variance, denoted c_0, is added to the models. The models thus have three parameters. We shall find it convenient later to characterize the variation in ε as the vector:

$$\boldsymbol{\phi} = [c_0, c, a] \tag{5}$$

for the spherical model, for example.

Any second-order stationary process has a covariance function corresponding to its variogram:

$$C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h}) \tag{6}$$

where $C(\mathbf{0})$ is the covariance at zero lag, i.e. the *a priori* variance or sill of the variogram.

Models with trend

We now introduce trend, and we elaborate Equation 1 to incorporate it:

$$Z(\mathbf{x}) = \sum_{k=0}^{K} \beta_k f_k(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{7}$$

We have replaced the constant μ by the sum of $K + 1$ linear combinations in which the f_k are known functions of $\mathbf{x}$ and the β_k are coefficients to be determined.

Typically the trend can be modelled as a polynomial in the spatial coordinates. So, for example, a linear trend, an inclined plane, would be:

$$Z(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon(\mathbf{x}) \tag{8}$$

For compactness later we rewrite this in matrix notation:

$$Z(\mathbf{x}) = \mathbf{W}\boldsymbol{\beta} + \varepsilon(\mathbf{x}) \tag{9}$$

In the first term on the right-hand side $\mathbf{W}$ is what is known technically as a design matrix with $K + 1$ columns, the $K + 1$ terms in the trend model, and $\boldsymbol{\beta}$ is the vector containing the coefficients. We shall assume that, for any finite set of points, the random component has a covariance matrix $\mathbf{C}$, which derives from the lag intervals between the points and the covariance function with the parameters in $\boldsymbol{\phi}$ and its equivalence to the variogram, Equation 6.

Prediction

Our task is now to predict values of z at unsampled points from known values elsewhere. Let us denote the known values by $\mathbf{z}(\mathbf{X}_o)$ and the unknowns that we wish to predict by $\mathbf{z}(\mathbf{X}_p)$, in which the subscript 'o' signifies observation and the subscript 'p' prediction. We assume first that the variance parameters, $\boldsymbol{\phi}$, and the parameters of the trend, $\boldsymbol{\beta}$, are known. The best linear predictor (BLP), best in the sense of minimum variance, is:

$$\tilde{\mathbf{Z}}(\mathbf{X}_p) = (\mathbf{W}_p - \mathbf{C}_{po}\mathbf{C}_{oo}^{-1}\mathbf{W})\boldsymbol{\beta} + \mathbf{C}_{po}\mathbf{C}_{oo}^{-1}\mathbf{z}(\mathbf{X}) \tag{10}$$

where $\mathbf{W}_p$ is the design matrix for the prediction sites. Matrix $\mathbf{C}_{po}$ is the covariance matrix of the observed values of z with the values at the target sites (see Welham *et al*., 2004; Gilmour *et al*., 2004).

In practice we require an estimate of $\boldsymbol{\beta}$, which can most easily be obtained by ordinary least squares (OLS):

$$\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{z} \tag{11}$$

This estimate is always unbiased, but it is not of minimum variance unless the data have been obtained by independent random sampling. If we substitute $\hat{\boldsymbol{\beta}}$ into Equation 10 the resulting predictor is the best linear unbiased predictor (BLUP).

For practical purposes the variance parameters $\boldsymbol{\phi}$, required to compute the covariance matrices, must also be estimated from data. Obtaining them is not simple because we must separate the trend from the random component in our data, and the estimates will be biased if they depend non-linearly on other parameters of the model, so-called 'nuisance parameters'. The usual method of estimating them by modelling the experimental variogram obtained by the method of moments from residuals produces bias, because the residuals depend on estimates of the parameters $\boldsymbol{\beta}$.

The solution to this problem is to use residual maximum likelihood (REML), first proposed for the estimation of variance components by Patterson and Thompson (1971). For this we have to assume that the set of random variables $\boldsymbol{\varepsilon}$ is multivariate normal with covariance matrix $\mathbf{C}_{oo}$. The essential principle of REML is to obtain a new random variable, a function of the data that is independent of the nuisance parameters and that has a covariance matrix that is a known function of $\mathbf{C}_{oo}$. One can therefore estimate $\boldsymbol{\phi}$ by applying maximum likelihood to this new variable; this is the residual likelihood. In short, if for some non-singular matrix $\mathbf{L}$

$$\mathbf{L}^{\mathrm{T}}\mathbf{W} = 0$$

then we may compute

$$\mathbf{y} = \mathbf{L}^{\mathrm{T}}\mathbf{z}\ (\mathbf{X})$$

and

$$\mathbf{y} \sim N\{\mathbf{0}, \mathbf{L}^{\mathrm{T}}\mathbf{C}_{\mathrm{oo}}\mathbf{L}\}$$

For the general linear model, as used here, the log residual likelihood is (Stuart *et al*., 1999):

$$l(\phi|\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \text{constant} - \frac{1}{2}\ln|\mathbf{C}_{\mathrm{oo}}| - \frac{1}{2}|\mathbf{W}^{\mathrm{T}}\mathbf{C}_{\mathrm{oo}}^{-1}\mathbf{W}| - \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{C}_{\mathrm{oo}}^{-1}(\mathbf{I} - \mathbf{Q})\mathbf{z} \tag{12}$$

where

$$\mathbf{Q} \equiv \mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{C}_{\mathrm{oo}}^{-1}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{C}_{\mathrm{oo}}^{-1} \tag{13}$$

and $\mathbf{I}$ is an identity matrix.

Gilmour *et al*. (1995) proposed the average information (AI) algorithm for efficient maximization of the log residual likelihood. However, the algorithm is not suitable for estimating the parameters of a spherical variogram. The reason is that this variogram does not have a smooth likelihood function, and so the gradient method used in the AI algorithm may find an estimate of $\boldsymbol{\phi}$ that is only locally optimal with respect to the residual likelihood, and does not converge to the global maximum residual likelihood. Hence, Lark and Cullis (2004) used simulated annealing to find the REML estimates of spatial variance models, and this is the method that we have used here.

Once estimates of the variance parameters have been obtained, they can be used to compute an estimate of the covariance matrix $\hat{\mathbf{C}}_{\mathrm{oo}}$. With this we can then obtain the generalized least squares estimate of the parameters $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ which are of minimum variance for unbiased $\hat{\mathbf{C}}$, as follows.

$$\hat{\boldsymbol{\beta}} = (\mathbf{W}^{\mathrm{T}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{z} \tag{14}$$

Lark and Cullis (2004) describe the method in more detail.

If we substitute estimated covariance matrices $\hat{\mathbf{C}}$ and the generalized least squares estimate $\hat{\boldsymbol{\beta}}$ into Equation 10 we obtain the empirical BLUP (E-BLUP):

$$\tilde{\mathbf{Z}}(\mathbf{X}_{\mathrm{p}}) = (\mathbf{W}_{\mathrm{p}} - \hat{\mathbf{C}}_{\mathrm{po}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{W})\hat{\boldsymbol{\beta}} + \hat{\mathbf{C}}_{\mathrm{po}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{z}(\mathbf{X}) \tag{15}$$

There are two parts to this equation. The first corresponds to the trend component, and the second to the kriging. As we have noted above, the E-BLUP is directly equivalent to universal kriging, given the variogram defined by the REML estimates of ϕ (Stein, 1999).

The covariance matrix of the prediction errors is given by:

$$(\mathbf{W}_{\mathrm{p}} - \hat{\mathbf{C}}_{\mathrm{po}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{W})\mathbf{U}^{-1}(\mathbf{W}_{\mathrm{p}} - \hat{\mathbf{C}}_{\mathrm{po}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{W})^{\mathrm{T}} + (\hat{\mathbf{C}}^{\mathrm{pp}})^{-1} \tag{16}$$

where $\mathbf{U} = \mathbf{W}^{\mathrm{T}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\mathbf{W}$ and $\hat{\mathbf{C}}^{\mathrm{pp}} = (\hat{\mathbf{C}}_{\mathrm{pp}} - \hat{\mathbf{C}}_{\mathrm{po}}\hat{\mathbf{C}}_{\mathrm{oo}}^{-1}\hat{\mathbf{C}}_{\mathrm{po}}^{\mathrm{T}})^{-1}$. The matrix $\hat{\mathbf{C}}_{\mathrm{pp}}$ is the covariance matrix of the target sites, determined from the REML estimates of the variance parameters. The estimation variance of the prediction at any site may be extracted from this matrix.

Welham *et al*. (2004) present an example of the application of E-BLUP to a geostatistical problem. We now describe the application of E-BLUP to the local estimation of a geomorphological variable with a strong spatial trend.

Case Study: the Chalk Surfaces

To illustrate the technique we return to the study of Moffat *et al*. (1986) on the Chalk surfaces beneath the Chilterns. Wooldridge (1923) and Wooldridge and Linton (1955) interpreted surveys of the landform and surface deposits, and proposed a sequence of essentially erosional events on a simple tilted sub-Tertiary surface to account for the current landscape of the region. Moffat *et al*. (1986), with many more data to hand from boreholes, challenged this view. In particular, they had 238 heights of the sub-Upper-Chalk surface. A major source of variation in these was the regional dip. Moffat *et al*. modelled this as a quadratic trend surface. They removed it from the data and treated the residuals as the outcome of a stationary random process. They computed the sample variogram of these, fitted a model to it, and with the model they interpolated the residual surface on a fine grid by ordinary punctual kriging. In doing so they revealed a series of local structures running across the regional dip with their hollows approximately in line with the present-day river valleys. They concluded that these flexures had determined the positions of the rivers.

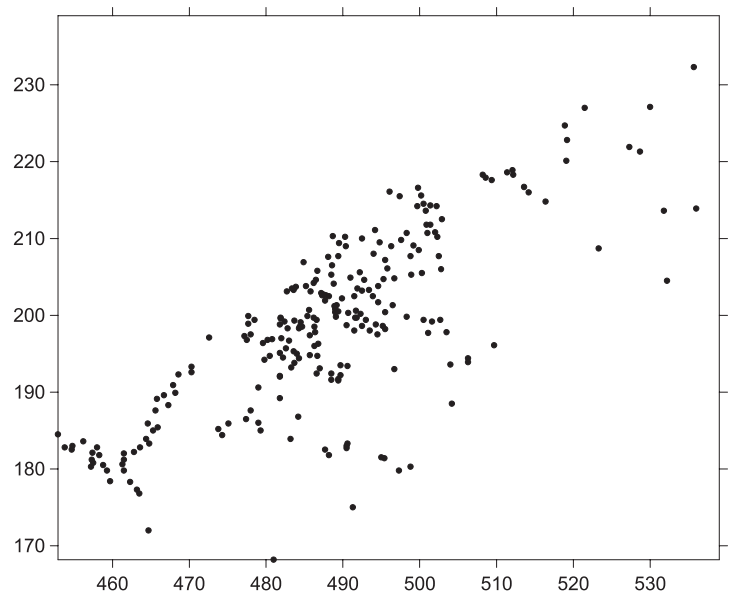

Figure 1. Positions of boreholes. Coordinates are in metres according to the National Grid of the British Ordnance Survey; the bottom left corner of the map is in 100-km square SU.

Table I. Statistical summaries of the raw data (heights, in metres, of the sub-Upper-Chalk surface above sea level) and of the residuals from linear and quadratic trend surfaces

	Raw height	Residuals	
		from linear	from quadratic
Mean	121·7	0	0
Median	137·2	2·9	−0·8
Minimum	−155·1	−108·7	−58·4
Maximum	228·6	60·4	51·6
Variance	4335·6	673·6	262·0
Standard deviation	65·8	26·0	16·2
Skewness	−1·28	−0·55	0·24

We repeat some of the results of Moffat *et al*. (1986) and then show how a modern analysis by REML improves on it. We first summarize the data.

The data and their summary

The positions of the boreholes from which the data derive are shown in Figure 1. The coordinates are those of the British National Grid. The Chalk scarp marks the northwesterly limit of the data. The first column of Table I summarizes the data, which are heights of the sub-Upper-Chalk surface above sea level.

Exploratory analysis of trend

The regional dip, as is well known, is approximately from northwest to southeast; it has a gradient of about 10 m km^{-1}. Moffat *et al*. (1986) modelled it initially as an inclined plane. In mathematical terms this is:

$$z = b_0 + b_1 x_1 + b_2 x_2 \tag{17}$$

Table II. Coefficients of linear and quadratic trend surfaces, Equations 17 and 18, fitted to the sub-Upper-Chalk surface for which the original heights were in metres

	Function	
Coefficient	**Linear**	**Quadratic**
b_0 (intercept)	65·2	−26·3
b_1	−5·30	−4·05
b_2	8·03	13·51
b_3		−0·01
b_4		−0·09

where x_1 and x_2 are the eastings and northings, and b_0, b_1 and b_2 are coefficients that were fitted by ordinary least squares (OLS). We repeated the fitting, and we list the coefficients in Table II. The fitted plane accounts for 84 per cent of the variance.

Moffat *et al*. (1986) noticed that the plane did not fit well in the northeast of the region, where the dip appeared to be greater than elsewhere. They therefore added a quadratic component, and we have done the same, to give the equation:

$$z = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_2^2 \tag{18}$$

in which there are the additional terms in x_1^2 and x_2^2. With these two additional terms 94 per cent of the variance was accounted for. Note that there is no term in x_1x_2; this is because it was aliased with the other terms and so could not be estimated separately. Again, Table II lists the coefficients.

Exploratory geostatistical analysis: experimental variograms of the OLS residuals

Moffat *et al*. (1986) computed semivariances of their original data and of the linear and quadratic residuals, primarily so that they could choose a function with which to interpolate surfaces by kriging. We have repeated the computations with the standard formula:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2 \tag{19}$$

in which $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i + \mathbf{h})$ are the heights or the residuals at positions $\mathbf{x}_i$ and $\mathbf{x}_i + \mathbf{h}$ separated by the lag $\mathbf{h}$, and $m(\mathbf{h})$ is the number of paired comparisons at that lag. Figure 2 displays the results, the experimental variograms. The symbols indicate the eight directions in degrees counterclockwise from east.

Figure 2a is the experimental variogram of the raw heights. The semivariances in directions 22·5° and 45° increase over lag distances of 17 to 18 km and then stabilize beyond. These semivariances are approximately in the direction of the strike. In all other directions the experimental points lie on upwardly curving lines. The most marked are those in directions 112·5° and 135°, i.e. perpendicular to the strike – in the direction of the dip, which is the direction of the trend, of course. Such upward curvature is characteristic of long-range trend and of a non-stationary underlying process.

Figure 2b is the experimental variogram of the residuals from the fitted inclined plane, Equation 17 with parameters listed in Table II. The upward curvature evident in Figure 2a has disappeared, but there are systematic differences among the different directions, especially beyond about 12 km. Moffat *et al*. (1986) fitted a linear isotropic model to this variogram, but they were dissatisfied with it, partly because of the behaviour beyond 12 km.

Computing the variogram of the residuals from the quadratic surface with parameters listed in Table II gives Figure 2c. There remains a lot of scatter among the estimates, but the systematic differences between directions have largely disappeared. The residual variation seems isotropic, and we have treated it as such. We have fitted two isotropic models, linear and spherical. The first, shown by the dashed line, is fitted to 13 km. Its equation is

$$\gamma(h) = 63{\cdot}5 + 20{\cdot}2h \tag{20}$$

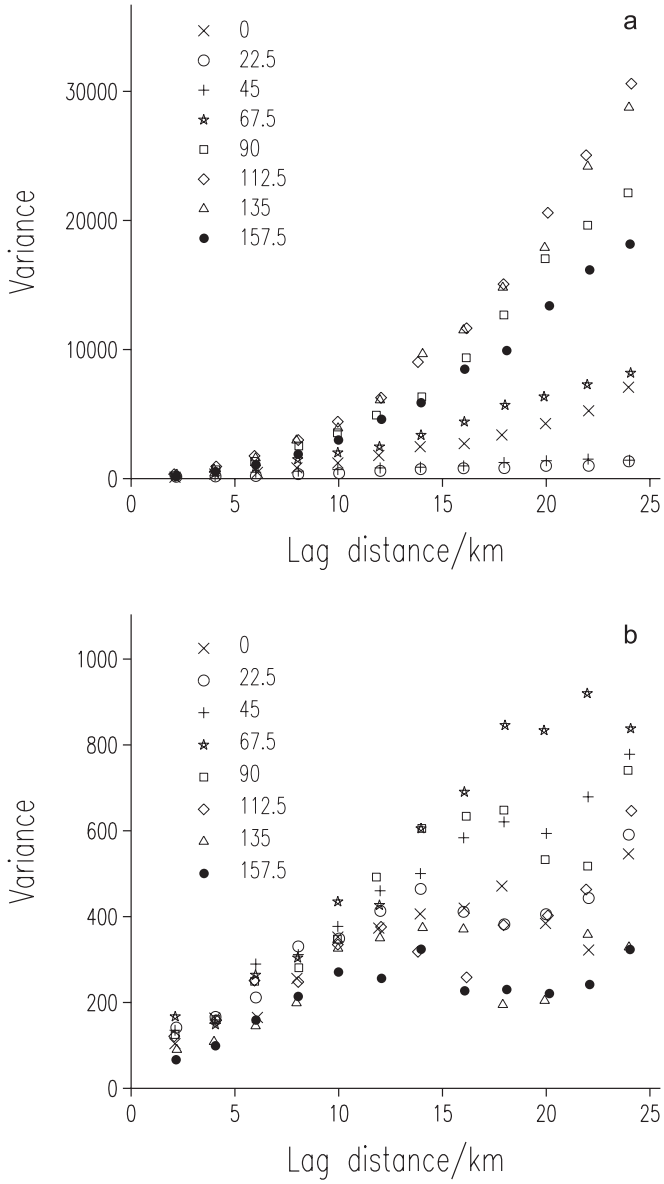

Figure 2. Variograms of (a) the raw data on heights, (b) the residuals from the OLS linear trend surface and (c) the residuals from the OLS quadratic trend surface.

Clearly the equation does not fit the experimental values beyond 13 km. The spherical function, described above, describes the whole experimental set better. The values of the three parameters are $c_0 = 30{\cdot}1$, $c = 261{\cdot}5$ and $a = 13{\cdot}5$ km.

Moffat *et al*. used the linear variogram, with slightly different values from those above, to krige the residuals on to a fine grid. They contoured the results to produce a map of the residuals (Figure 4 in Moffat *et al*., 1986).

Analysis with REML

We estimated the quadratic trend surface in Equation 18 using REML to estimate the variance parameters for the error $\varepsilon(\mathbf{x})$, then computing estimates of the trend parameters by generalized least squares as in Equation 14. Minimization of the negative log residual likelihood was achieved with the simulated annealing method described by Lark and Cullis (2004). The terms $b_0, b_1, \ldots, b_4$ in Equation 18 are the elements of $\boldsymbol{\beta}$, and the design matrix $\mathbf{W}$ contains in its first column a dummy variable (1) that corresponds to the constant b_0, and in the next four columns the values of x_1,

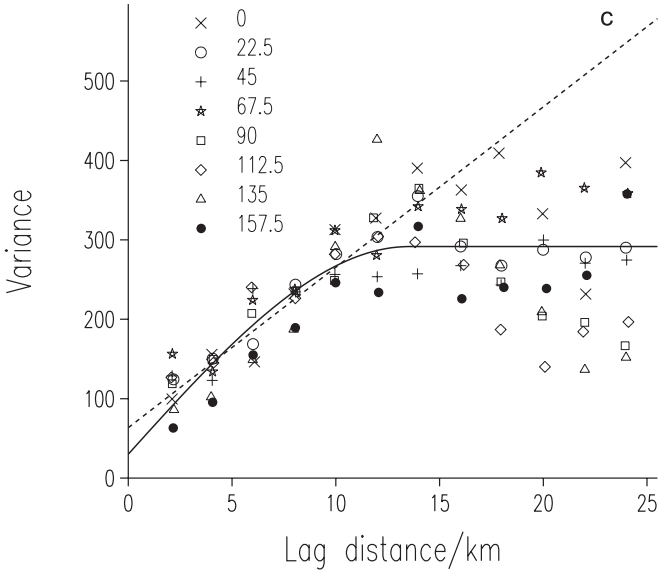

Figure 2. *Continued*

x_2, x_1^2 and x_2^2 that correspond to each of our n observations. We considered both spherical and exponential variograms for $\varepsilon(\mathbf{x})$, and selected the model for which the minimized negative log residual likelihood was smallest. This was the spherical model, and the estimated variance parameters and those of the trend surface are listed in Table III.

In the original analysis by Moffat *et al*. (1986) the decision to use a quadratic trend model rather than just a linear trend surface was based on examination of the residuals and their experimental variogram. In the REML framework this evaluation of more complex model terms can be aided by a formal statistical inference. We computed a Wald statistic (Stuart *et al*., 1999; Lark and Cullis, 2004) to test the null hypothesis that the two quadratic coefficients b_3 and b_4 are zero. This Wald statistic is tested against the F distribution with two degrees of freedom in the numerator. The number of degrees of freedom in the denominator was computed by the procedure of Kenward and Roger (1997), which allows for correlation of the error terms and computes an adjustment for the Wald statistic. The test showed that we could reject the null hypothesis $p = 0{\cdot}036$).

The results in Table III show that more than 93 per cent of the variance in the random component, $\varepsilon(\mathbf{x})$, is spatially dependent; less than 7 per cent of it is nugget variance. It has a larger variance and a longer range than models fitted to the experimental variogram of the OLS residuals from the quadratic trend (see Figure 3, in which the REML variogram is plotted on the same axes as the isotropic experimental variogram of the OLS residuals). The REML

Table III. Results from REML estimation of a quadratic trend surface fitted to the sub-Upper-Chalk surface for which the original heights were in metres

Parameter	Estimate
Spherical model	
c_0	66·6
c	957·5
a	42·1
Trend coefficient	
b_0	−0·43
b_1	−4·6
b_2	11·9
b_3	−0·004
b_4	−0·7

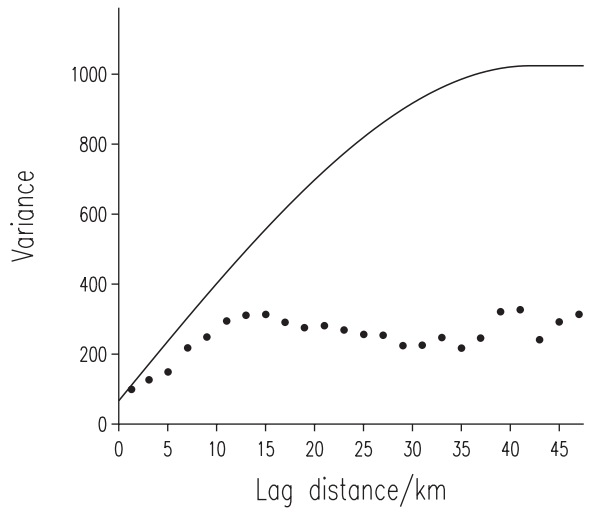

Figure 3. Isotropic experimental variogram of the residuals from the REML quadratic trend surface (symbols) and the REML variogram model (continuous line).

variogram and the experimental variogram are similar at short lags, but diverge at long lags. This is what we expect from the theory, the bias increases with lag, reflecting the non-linear dependence of the experimental variogram on the nuisance parameters in β.

We then computed the E-BLUP of the sub-Upper-Chalk surface; this is displayed in Figure 4. Figure 5 shows the corresponding surface for the random term $\varepsilon(\mathbf{x})$. Both of these are broadly similar to the results of Moffat *et al.* (1986), which we should expect because we know that the OLS estimates of β are unbiased. A particular advantage of using the REML–E-BLUP is that we can compute Figure 6, a map of the prediction variances of the E-BLUP, which shows how our confidence in the predicted height of the surface varies spatially. This map could not be produced by

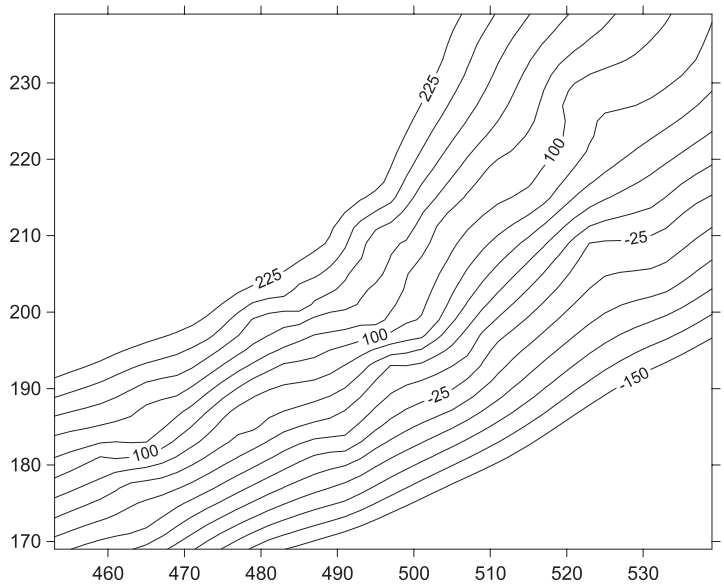

Figure 4. Contours of the E-BLUP estimate of the sub-Upper-Chalk surface. Units are metres relative to sea level.

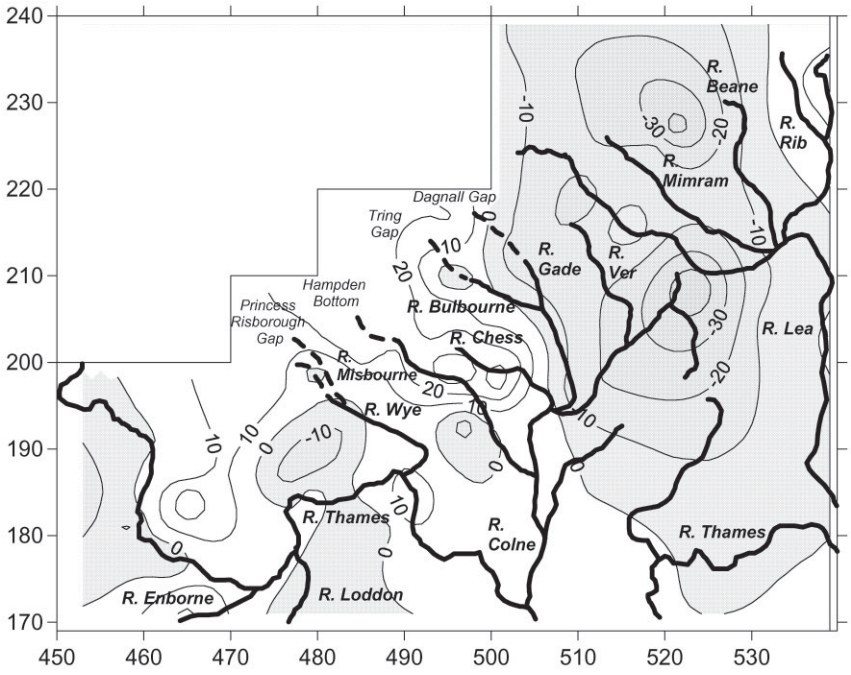

Figure 5. Contours of the random term $\varepsilon(\mathbf{x})$; units are metres, with the local rivers shown. Broken lines correspond to upper reaches of Chiltern valleys that coincide with synclinal areas (after Moffat *et al.*, 1986).

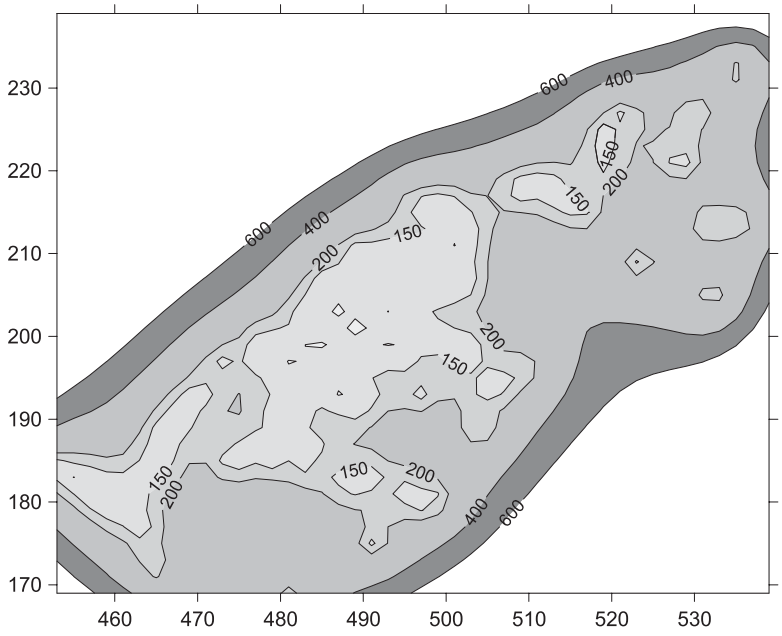

Figure 6. Estimation variance of the E-BLUP estimate of the sub-Upper-Chalk surface. Units are metres.

regression kriging, because of bias in the estimates of the variogram in this approach and because it separates the errors of deterministic and random components of the prediction, both of which contribute to the final uncertainty. The variances in Figure 6 contain contributions from both sources.

We make two comments on the actual error variances. First, the variance tends to be smaller near to observation points than it is further away. This is a general and well known characteristic of kriging. Second, the variances tend to be small near the geographic centroid of the data because that is where the prediction variance of the trend surface is smallest.

Discussion and Conclusions

Since the original analysis of these data by Moffat *et al*. (1986), various other workers have used essentially the same procedure to study variables which exhibit both a trend and spatially dependent random variation. These regression kriging methods suffer from bias in the variogram of the random component, and, in most cases when sampling is not random, the estimates of the trend parameters are not of minimum variance. In this case our E-BLUP of the trend and the variation about the trend does not differ markedly from the predictions made by Moffat *et al*. As we noted above, this is not altogether surprising because the ordinary least squares and REML-based estimates of the trend parameters are both unbiased, and, given the density of the observations, the kriging estimates are dominated by the variogram at short lags where the bias is smallest. In consequence the interpretation that Moffat *et al*. (1986) offered of the interpolated surface is unaffected.

The geomorphologist might therefore wonder whether the additional effort of the REML–E-BLUP procedure is worthwhile. We argue strongly that it is, and for five reasons.

1. Whereas the local density of observations in this study compensated for the relative inefficiencies and bias of the standard regression kriging method, geomorphologists often want to work with sparser observations.
2. The confidence in kriging estimates at any location is measured by the kriging variance. If the variogram is biased then the kriging variances will be biased, and practitioners might have undue confidence in their predictions as a result.
3. Further, the standard regression kriging procedure, as used by Moffat *et al*. (1986) and by most other workers since, does not provide an estimation variance for the sum of the trend and the kriged prediction of the random variation. An estimate can be obtained by further computation (of the universal kriging variance), but it is obtained directly from the E-BLUP as we present it here.
4. We might use the variogram of the random variation from a study such as the one reported here to plan further sampling, either to supplement the data in this region or for a study of the same formation in another region. McBratney *et al*. (1981) described the principles, and a sampling scheme can be designed to ensure that kriging predictions are of adequate precision without oversampling. If the biased variogram, computed from residuals, were used then the sampling required to achieve a particular precision might be underestimated.
5. In practice one might not be interested directly in the variable that was measured, but require values of it for input to some model. In the present context one might want to enter the heights of a particular stratigraphic horizon, and its local variations, into a model of the behaviour of an aquifer. A kriged surface is well known to be smoother than the regionalized variable it predicts, and so it is not suitable as input to a model unless its effects in the model are linear. For this reason a common practice is to simulate sets of values to put into models (e.g. Journel and Xu, 1994). The simulations are drawn at random from the set of realizations of the stationary random function represented by the variogram such that the values at the observation sites coincide with the data. If one were to use the biased variogram obtained from the trend surface residuals then the variations of the simulated values would differ from those obtained with the REML estimate of the variogram, and it would lead to bias in the model predictions.

To conclude, we commend the REML–E-BLUP to geomorphologists as a method to predict the spatial variations when spatial trend is a significant component. The method provides efficient predictions and unbiased estimates of the error variances to quantify their uncertainty. Estimation by REML yields a variogram that is free from the bias in the variogram estimated from trend-surface residuals, and so it is preferable to plan further sampling and for geostatistical simulations.

References

Cressie NAC, 1993. *Statistics for spatial data*. John Wiley & Sons Ltd: New York.

Gambolati G, Galeati G 1987. Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels – comment. *Mathematical Geology* **19**: 249–257.

Gilmour AR, Thompson R, Cullis BR. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**: 1440–1450.

Gilmour A, Cullis B, Welham S, Gogel B, Thompson R. 2004. An efficient computing strategy for prediction in mixed linear models. *Computational Statistics & Data Analysis* **44**: 571–586.

Hengl T, Heuvelink GBM, Stein A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **120**: 75–93.

Journel AG, Xu WL. 1994. Posterior identification of histograms conditional to local data. *Mathematical Geology* **26**: 323–359.

Kenward MG, Roger JH. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**: 983–997.

Knotters M, Brus D, Voshaar J. 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma* **67**: 227–246.

Lark RM, Cullis BR. 2004. Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science* **55**: 799–813.

Matheron G. 1969. *Le krigeage universel*. Cahiers du Centre de Morphologie Mathématique, Ecole de Mines de Paris: Fontainebleau.

Matheron G. 1973. The intrinsic random functions and their applications. *Advances in Applied Probability* **5**: 439–468.

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. *Computers and Geosciences* **7**: 331–334.

Moffat AJ, Catt JA, Webster R, Brown EH. 1986. A re-examination of the evidence for a Plio-Pleistocene marine transgression on the Chiltern Hills. I. Structures and surfaces. *Earth Surface Processes and Landforms* **11**: 95–106.

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.

Odeh IOA, McBratney AB, Chittleborough DJ. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* **67**: 215–226.

Olea RA. 1975. *Optimal Mapping Techniques using Regionalized Variable Theory*. Series on Spatial Analysis no 2. Kansas Geological Survey: Lawrence, Kansas.

Oliver MA, Webster R. 1986. Semi-variograms for modelling the spatial pattern of land form and soil properties. *Earth Surface Processes and Landforms* **11**: 491–504.

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

Stein ML. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer: New York.

Stuart A, Ord JK, Arnold S. 1999. *Kendall's Advanced Theory of Statistics*, Volume 2A: *Classical Inference and the Linear Model* (6th edition). Arnold: London.

Webster R, Burgess TM. 1980. Optimum interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. *Journal of Soil Science* **31**: 505–524.

Welham S, Cullis B, Gogel B, Gilmour A, Thompson R. 2004. Prediction in linear mixed models. *Australian and New Zealand Journal of Statistics* **44**: 571–586.

Wooldridge SW. 1923. Minor structures of the London Basin. *Proceedings of the Geologists' Association of London* **34**: 175–193.

Wooldridge SW, Linton DL. 1955. *Structure, Surface and Drainage in South-East England* (2nd edition). G. Philip and Son: London.

