Journal of Soil Science, 1986, **37**, 617–639

Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates

A. B. McBRATNEY & R. WEBSTER*
*CSIRO Division of Soils, Cunningham Laboratory, 306 Carmody Road, St Lucia, Queensland, Australia, 4067 and *Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ, U.K.*

SUMMARY

The semi-variogram is central to geostatistics and the single most important tool in geostatistical applications to soil. Mathematical functions for semi-variograms must be conditional negative semi-definite, and there are only a few families of simple function that meet this demand. These include the transitive models with finite a priori variance deriving from moving average processes. The spherical and exponential schemes are the most often encountered members. The other major group is that of unbounded models in which the variance appears to increase without limit. The linear model is the most common in this group. If more complex models are needed they can be formed by combining two or more simple models.

The usual estimator of the semi-variance is often considered inefficient and to be sensitive to departures from normality in the data. It is compared with a robust estimator and shown to be generally preferable in being unbiased and having confidence intervals that are no wider. For routine analysis, fitting models to sample semi-variograms by weighted least squares approximation, with weights proportional to the expected semi-variance, is preferred to the more elaborate and computationally demanding statistical procedures of generalized least squares and maximum likelihood. The Akaike information criterion is recommended for selecting the best model from several plausible ones to describe the observed variation in soil, though for kriging it may be desirable to validate the chosen model.

Examples of models fitted to soil semi-variograms are shown and compared.

INTRODUCTION

Soil scientists have been concerned with spatial variation of soil for many years. It is only fairly recently, however, that satisfactory means have become available for describing the variation quantitatively, thanks largely to Matheron (1965, 1971) and colleagues (for example, Journel & Huijbregts, 1978) at the Paris School of Mines. These means are embodied in the theory and practice of geostatistics. As in the mining industry, where geostatistics originated, the main motivation for using the methods in soil science is the design of optimal sampling schemes and the spatial prediction of resources.

Central to geostatistics is the *semi-variogram*. It is the key to describing variation quantitatively, to understanding, and to prediction. In earlier papers (eg. Burgess & Webster, 1980*a,b;* McBratney *et al.*, 1982; McBratney & Webster, 1981*b*, 1983) we presented sample semi-variograms of soil properties in several regions and fitted mathematical functions to them. Other workers have done the same (Gajem *et al.*, 1981; Vieira *et al.*, 1981; Yost *et al.*, 1982*a,b*). In these examples the mathematical functions and fitting procedures, if mentioned, are simply stated as empirical and reasonable choices in the circumstances. Only Sisson & Wierenga (1981), Burrough (1983*a,b*) and Webster (1985) discuss the theoretical reasoning for particular forms of semi-variogram.

Some aspects of geostatistics, namely sound sampling, estimation by kriging, and isarithmic

mapping are now fairly well understood and accepted by soil scientists. Choosing an appropriate semi-variogram in any given situation, however, is still something of a mystery. It is by no means automatic. It requires good judgement based on experience and an understanding of the mathematical limitations of the function, and it is not a task that the soil scientist can delegate. This paper reviews the subject and draws on our experience to recommend suitable models for semi-variograms and procedures for estimating them.

Definitions

Consider a soil property Z, such as clay percentage or pH, that can vary continuously in geographic space. The property takes values $z(\mathbf{x}_i)$ at places $\mathbf{x}_i$, $i=1, 2\ldots$, where $\mathbf{x}$ denotes the set of spatial coordinates in one, two or three dimensions. If we consider the values of Z at two places, $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$, where $\mathbf{h}$ is a vector, known as the *lag*, denoting their separation in both distance and direction, then their variance is given by

$$\begin{aligned} s^2 &= \{z(\mathbf{x})-\bar{z}\}^2+\{z(\mathbf{x}+\mathbf{h})-\bar{z}\}^2 \\ &= \tfrac{1}{2}\{z(\mathbf{x})-z(\mathbf{x}+\mathbf{h})\}^2, \end{aligned} \tag{1}$$

where $\bar{z}$ is the mean of the two values. This value of s^2 is specific to $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$, and to make more general use of the measure we have to consider the soil as the realization of a random process with certain stationarity conditions. These are:

1. that the expected difference between values at any two places separated by $\mathbf{h}$ is zero:

$$E[z(\mathbf{x})-z(\mathbf{x}+\mathbf{h})]=0, \tag{2}$$

and

2. that the variance of the differences depends on $\mathbf{h}$ and not on $\mathbf{x}$, and is given by

$$\begin{aligned} \text{var}[z(\mathbf{x})-z(\mathbf{x}+\mathbf{h})] &= E[\{z(\mathbf{x})-z(\mathbf{x}+\mathbf{h})\}^2] \\ &= 2\gamma(\mathbf{h}). \end{aligned} \tag{3}$$

These two conditions constitute the *intrinsic hypothesis* of regionalized variable theory, and from our experience they seem widely applicable to soil.

In regionalized variable theory the quantity γ is known as the *semi-variance*, and is the expectation of s^2, the variance per site when sites are considered in pairs. The *semi-variogram* is then the function that relates γ to $\mathbf{h}$. Where the intrinsic hypothesis holds it contains all the information on the spatial variation of the soil. The intrinsic hypothesis also enables semi-variances to be estimated simply from a sample of a single realization of the process as

$$\hat{\gamma}(\mathbf{h})=\frac{1}{2m(\mathbf{h})}\sum_{i=1}^{m(\mathbf{h})}\{z(\mathbf{x}_i)-z(\mathbf{x}_i+\mathbf{h})\}^2, \tag{4}$$

where $m(\mathbf{h})$ is the number of pairs of observations separated by $\mathbf{h}$. An ordered set of these constitutes a sample semi-variogram.

GENERAL FORMS OF SEMI-VARIOGRAM

Several recent publications illustrate sample semi-variograms of soil properties. When sampling fluctuations are allowed for, most of these can be seen to have simple forms. Broadly there are two main types.

In the one type the semi-variance increases with increasing lag distance, $|\mathbf{h}|$, to some maximum at which it remains with further increase in $|\mathbf{h}|$. Such a semi-variogram is said to be *transitive*, and is much as one might expect. Fig. 1 illustrates a typical example. It shows that as the separating distance increases the soil becomes more dissimilar on average, but that there is a finite lag within which all variation is encountered. This lag, the *range* of the semi-variogram, marks the limit of

spatial dependence in the property concerned. The maximum variance is known as the *sill* of the semi-variogram. It is the a priori variance of the property.

A variable with this kind of semi-variogram is not only intrinsic, in the above sense, but also second-order stationary. It has a constant expectation for all $\mathbf{x}$:

$$E[z(\mathbf{x})] = \mu, \tag{5}$$

and a covariance defined by

$$\begin{aligned} C(\mathbf{h}) &= E[\{z(\mathbf{x}) - \mu\}\{z(\mathbf{x}+\mathbf{h}) - \mu\}] \\ &= E[z(\mathbf{x}) - z(\mathbf{x}+\mathbf{h})] - \mu^2. \end{aligned} \tag{6}$$

At zero lag equation (6) is that of the variance

$$\begin{aligned} C(0) &= E[z^2(\mathbf{x})] - \mu^2 \\ &= \sigma^2 \end{aligned} \tag{7}$$

and the semi-variogram is readily seen to be complementary to the covariance function, since

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}). \tag{8}$$

The autocorrelation $\rho(\mathbf{h})$ is also defined here since it will be needed later. It is

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(0)} = 1 - \frac{\gamma(\mathbf{h})}{C(0)}.$$

The second main type of semi-variogram appears to increase without limit, i.e. is unbounded. There is no finite a priori variance. The idea of infinite variance may seem at first somewhat strange, but it should not be wholly unfamiliar to soil scientists. Youden & Mehlich (1937) found it in their sampling. Fairfield Smith (1938) embodied it in his empirical law of the variance of crop yields, and pedologists expect to encounter new soil types as they go further afield. Figs 2 & 3 show examples of this type of variation.

Another feature of semi-variograms that must be recognized is that although by definition $\gamma(0) = 0$ the apparent limiting value of $\gamma(\mathbf{h})$ as $|\mathbf{h}|$ approaches 0 is not necessarily zero. Where this happens it is known as the *nugget effect*, and the intercept is known as the *nugget variance*. In principle it can happen only if there are discontinuities in the random function, when it corresponds to white noise. For a continuously variable material it arises from a contribution of measurement errors and variation over distances much smaller than the closest sampling interval. In what follows the nugget variance is denoted by c_0, and the difference between the nugget and the sill as c.

These then are the forms for which we seek mathematical models.

AUTHORIZED FUNCTIONS

In choosing a function for a semi-variogram allowance must be made in most instances for up to three elements: an intercept on the ordinate, a monotonically increasing section, and a sill.

Not any model that appears to fit the observed values will serve, however, for the following reason. Suppose that Z is a second-order stationary random function giving rise to the regionalized variable $z(\mathbf{x}_i)$, $i = 1, 2, \ldots n$, and that its semi-variogram and covariance function are $\gamma(\mathbf{h})$ and $C(\mathbf{h})$, respectively. Let Y be a linear combination of $z(\mathbf{x}_i)$ such that

$$Y = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i), \tag{9}$$

where λ_i, $i = 1, 2, \ldots n$ are any arbitrary weights. The quantity Y is also a regionalized variable with variance:

$$\text{var}[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C(\mathbf{x}_i, \mathbf{x}_j). \tag{10}$$

The variance of Y may be positive or zero, but not negative, and the covariance function in the right-hand side of equation (10) must ensure that this condition is satisfied. The covariance matrix must be *positive semi-definite*, that is, the determinant

$$\begin{vmatrix} C(\mathbf{x}_1, \mathbf{x}_1)\, C(\mathbf{x}_1, \mathbf{x}_2) \ldots C(\mathbf{x}_1, \mathbf{x}_n) \\ C(\mathbf{x}_2, \mathbf{x}_1)\, C(\mathbf{x}_2, \mathbf{x}_2) \ldots C(\mathbf{x}_2, \mathbf{x}_n) \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ C(\mathbf{x}_n, \mathbf{x}_1)\, C(\mathbf{x}_n, \mathbf{x}_2) \ldots C(\mathbf{x}_n, \mathbf{x}_n) \end{vmatrix}$$

and all its principal minors are positive or zero. Thus the function $C(\mathbf{h})$, if it exists, must be positive semi-definite, and only functions that meet this criterion are permissible.

As above, soil properties that do not have finite *a priori* variances have no definable covariances. Provided the intrinsic hypothesis holds, however, we can stipulate conditions for the semi-variogram. Equation (10) is equivalent to

$$\text{var}[Y] = C(0) \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} \lambda_j - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_i). \tag{11}$$

If the weights, λ_i, sum to zero then the first term on the right-hand side of equation (11) disappears, giving

$$\text{var}[Y] = -\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j). \tag{12}$$

This too must be non-negative. We can therefore state that $-\lambda(\mathbf{h})$ must be conditional positive semi-definite, the condition being that $\sum_{i=1}^{n} \lambda_i = 0$. Or, on removing the sign, the semi-variogram must be conditional negative semi-definite (CNSD). In the French literature such functions are called *authorized models*. Incidentally, the condition applies to any set of weights that sum to 0: they do not have to be optimal in any sense as they do for kriging.

The above condition is strict, and a model for a semi-variogram may not be chosen just because it looks as though it fits well. Armstrong & Jabin (1981) have drawn attention to the dangers of doing this, especially as a model that is CNSD in one dimension is not necessarily so in two or three. Equation (12) can give spurious negative values for var[Y] from unauthorized models. Armstrong & Jabin also showed how difficult it can be to discover empirically whether a function is permissible. Testing a function for acceptability involves examining its Fourier transform, and Christakos (1984) has given the conditions that the spectrum must meet for an authorized model. This in itself is not easy, and the best advice is: 'if in doubt choose from the valid models listed by Matérn (1960) and Journel & Huijbregts (1978)'.

Given the above constraint we find that there are just two broad groups of simple function for semi-variograms. They may be combined into more complex ones if the sample evidence seems to warrant it since any combination of authorized models is itself authorized. These are described below with examples from soil sampling to illustrate them.

Transitive models

This group of models derives from the notion of auto-correlation among the average values within blocks. The idea is that the random function, of which the soil property measured is a realization, depends on the amount of overlap of two blocks, i.e. a zone of transition. We consider first situations where the blocks are of equal size, and then generalize.

Linear model with sill. In one dimension the blocks are simply lines. Suppose these are of length a and that the distance between their centres is h. Their overlap is then $a - h$, which when expressed as a proportion of a is $1 - (h/a)$, provided $h \leqslant a$. The autocorrelation function is then

$$\rho(h) = 1 - \frac{h}{a}, \tag{13}$$

and the semi-variogram is

$$\gamma(h)=c_0+c\left(\frac{h}{a}\right) \qquad \text{for } 0<h\leqslant a$$

$$\gamma(h)=c_0+c \qquad \text{for } h>a$$

$$\gamma(0)=0. \tag{14}$$

Fig. 1 shows this model fitted to a one-dimensional sample semi-variogram of one of the colour variables, L^*, analysed by McBratney & Webster (1981*a*) for a transect at Tillycorthie in north east Scotland. The estimated values of c_0, c and a are given in Table 1. As can be seen the fit is reasonable. The range of the model is clearly a, the length of the blocks, and the gradient of the increasing part is c/a.

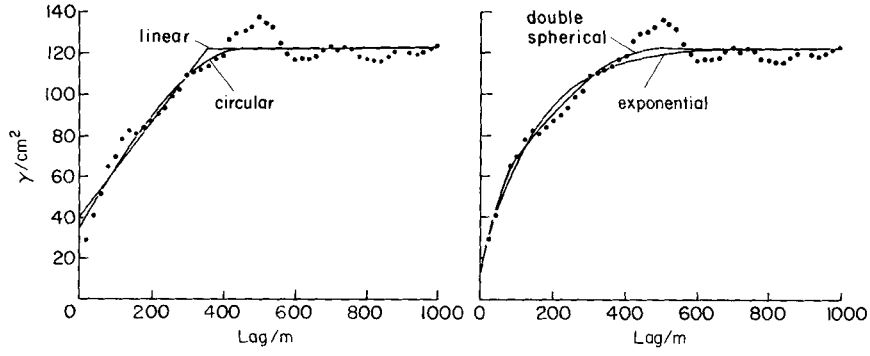

Fig. 1. Sample semi-variogram of the colour variable L^* at Tillycorthie with (left) linear, circular and (right) spherical and exponential models fitted to it.

Table 1. Parameter estimates and their standard errors (SE) fitted to the sample semi-variograms of the colour variables L^* at Tillycorthie and the Akaike Information Criterion

	c_0		c		a			
Model	Estimate	SE	Estimate	SE	Estimate	SE	*RSS*	*AIC*
Bounded Linear	15.93	1.36	47.01	1.54	317.2	9.8	192.2	163.7
Circular	14.88	1.43	48.60	1.53	365.8	15.0	175.2	161.0
Spherical	14.13	1.51	49.56	1.58	413.0	19.2	177.7	161.4
Exponential	10.02	2.24	60.14	2.12	203.6	21.6	227.5	168.8

The circular model. In two dimensions the blocks are discs of fixed diameter, a. The area of intersection of two such discs with centres h apart is

$$A=\frac{a^2}{2}\cos^{-1}\left(\frac{h}{a}\right)-\frac{h}{2}\sqrt{(a^2-h^2)} \qquad \text{for } h\leqslant a. \tag{15}$$

If this is expressed as a fraction of the area of a disc we obtain the auto-correlation function:

$$\rho(h)=\frac{2}{\pi}\left\{\cos^{-1}\left(\frac{h}{a}\right)-\frac{h}{a}\sqrt{\left(1-\frac{h^2}{a^2}\right)}\right\} \tag{16}$$

and the semi-variogram:

$$\begin{cases} \gamma(h) = c_0 + c\left\{1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a}\right) + \frac{2h}{\pi a}\sqrt{\left(1 - \frac{h^2}{a^2}\right)}\right\} & \text{for } 0 < h \leqslant a \\ \gamma(h) = c_0 + c & \text{for } h > a \\ \gamma(0) = 0. \end{cases} \tag{17}$$

As with the linear model the range is a, but the curve rises more steeply from the origin, where its gradient is $4c/\pi a$. It is defined for two dimensions, but since the circles are isotropic the lag becomes a scalar $h = |\mathbf{h}|$.

This model, known as the circular model, was used by Zubrzycki (1957) and Dalenius *et al.* (1961). It has not found much application in the earth sciences, though Webster (1985) and Oliver & Webster (1986) give examples of its fit to semi-variograms of soil pH and buried sedimentary horizons. Fig. 1 shows that it fits the semi-variogram of the colour variable L^* at Tillycorthie well. Again, Table 1 lists the estimated values of the parameters.

The spherical model. By analogy with the above two models we can derive a spherical model for semi-variograms by considering the volume of overlap of two spheres of diameter a. If h is the distance separating their centres then the volume of intersection is

$$V = \frac{\pi}{4} \cdot \frac{2a^3}{3} - a^2 h + \frac{h^3}{3} \qquad \text{for } h \leqslant a. \tag{18}$$

Dividing this by $\frac{1}{6}\pi a^3$, the spherical volume, we obtain the auto-correlation function:

$$\rho(h) = 1 - \frac{3h}{2a} + \frac{1}{2}\left(\frac{h}{a}\right)^3, \tag{19}$$

and the semi-variogram

$$\begin{cases} \gamma(h) = c_0 + c\left\{\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right\} & \text{for } 0 < h \leqslant a \\ \gamma(h) = c_0 + c & \text{for } h > a \\ \gamma(0) = 0. \end{cases} \tag{20}$$

Again a is the range of the model, which rises yet more steeply from the origin than the circular model, with gradient $3c/2a$. A spherical model is fitted to the same estimates of semi-variance of L^* on the right of Fig. 1, and the parameter estimates are given in Table 1.

The spherical model seems clearly appropriate for three-dimensional rock bodies, and it is much used in mining surveys. It is less obviously relevant for one- and two-dimensional distributions, and yet it has been found to describe such variation in soil well (e.g. Burgess & Webster, 1980a; McBratney & Webster, 1981*b*; Russo & Bresler, 1982; Webster & Nortcliff, 1984), and to be better than the linear and circular models (Webster, 1985). The generating process is evidently more complex than the moving average scheme alone.

Though we may choose the spherical model for one- and two-dimensional semi-variograms of soil we may not do the reverse; i.e. use a circular model for three dimensions or a linear model for either two or three dimensions. The circular semi-variogram is CNSD in one and two dimensions, but not in three. The linear model with sill is CNSD in only one dimension. It should not be used in two-dimensional surveys of soil, however well it might appear to fit!

If it appears that the semi-variogram curves more gradually than the spherical model but nevertheless reaches a sill at a fairly definite range then the five-dimensional analogue of the model is available (Matérn, 1960). Its definition is

$$\begin{cases} \gamma(h) = c_0 + c\left\{\dfrac{15h}{8a} - \dfrac{5}{4}\left(\dfrac{h}{a}\right)^3 + \dfrac{3}{8}\left(\dfrac{h}{a}\right)^5\right\} & \text{for } 0 < h \leqslant a \\ \gamma(h) = c_0 + c & \text{for } h > a \\ \gamma(0) = 0. \end{cases} \tag{21}$$

Its slope at $h = 0$ is $15c/8a$, and so rises more steeply than the spherical model.

The exponential model. If the overlapping blocks are allowed to vary in size at random then the semi-variogram becomes exponential in form. In the isotropic case its equation is

$$\begin{cases} \gamma(h) = c_0 + c\{1 - \exp(-h/r)\} & \text{for } 0 < h \\ \gamma(0) = 0. \end{cases} \tag{22}$$

Here the non-linear parameter r defines the spatial scale of the variation in a way analogous to the range of the previous models. The sill is approached asymptotically, and so there is no strict range. Nevertheless, for practical purposes the semi-variance ceases to increase beyond some point, and a commonly used rule of thumb is to regard the effective range as $a' = 3r$, which is the lag at which the sill reaches approximately $c_0 + 0.95\,c$. The gradient at the ordinate is c/r, or $3c/a'$, and so initially the curve rises more steeply than those in the first group. It has been added to Fig. 1 to show its form, though it clearly does not fit as well as the spherical and circular models do. Table 1 lists the parameter estimates.

The exponential model has appealed to statisticians for a long time because of its generality. First-order auto-regressive and Markov processes give rise to exponential semi-variograms, as do transition phenomena as mentioned earlier. The latter situation is closely related to the exponential distribution of distances between soil boundaries found by Burgess & Webster (1984). Where changes in soil across boundaries constitute the principal source of variation in a soil property and the boundaries occur as a Poisson process the exponential semi-variogram is again to be expected. If the intensity of the process is $\bar{\alpha}$, then the mean distance between boundaries is $\bar{d} = 1/\alpha$, and the semi-variogram is

$$\begin{aligned} \gamma(h) &= c\{1 - \exp(-h/\bar{d})\} \\ &= c\{1 - \exp(-\alpha h)\}. \end{aligned} \tag{23}$$

Whittle (1954, 1963) showed that a simple stochastic diffusion process also has an exponential semi-variogram in one and three dimensions. In two dimensions the process, perhaps surprisingly, leads to the semi-variogram:

$$\gamma(h) = c\left\{1 - \frac{h}{r}K_1\left(\frac{h}{r}\right)\right\},$$

in which K_1, a modified Bessel function of the second kind, replaces the exponential function. This increases more gradually than the exponential semi-variogram, and is slightly sigmoid near the origin. Whittle (1954) called the function the *elementary correlation* in two dimensions. Recently Baker (1984) has used it to model the spatial variability of mechanical properties of the soil.

Clearly, because exponential semi-variograms can have a variety of causes a knowledge of the semi-variogram alone is not enough to identify processes in particular instances.

Gaussian model. The one-dimensional semi-variogram has a sigmoid shape. There are no reports of its having been applied to soil data, and we have not encountered it in our experience. Nevertheless, because it is frequently described in the geostatistical literature it is mentioned here for completeness. Its equation is:

$$\begin{cases} \gamma(h) = c_0 + c\{1 - \exp(-h^2/r^2)\} \\ \gamma(0) = 0, \end{cases} \tag{24}$$

where r again is a non-linear parameter determining the spatial scale of the variation as in the exponential case.

Unbounded models

The unbounded semi-variograms describe situations in which the soil appears to vary increasingly without limit as the area increases. Equally, if one examines the variation over smaller and smaller distances it never seems actually to reach the origin. However small the sampling interval, if the supports are smaller than the interval there is always some variation remaining unresolved.

One starting point for understanding such variation is one-dimensional Brownian motion in which

$$z(x)-z(x+h)=\varepsilon \tag{25}$$

is a spatially independent Gaussian random variable. Its semi-variogram is

$$\gamma(h)=\frac{1}{2}h^{\theta}. \tag{26}$$

where $\theta=1$; i.e. it is linear with gradient $\frac{1}{2}$. If the sampling interval is divided by any arbitrary positive value b and the result rescaled in the ratio $b^{\theta/2}$ then the new semi-variogram will be identical to the old one. Brownian motion is thus self similar. Mandelbrot (1982) calls the outcome of such a process a *fractal* (see also Burrough, 1981, 1983*a*). The $\frac{1}{2}$ in equation (26) can be replaced with a parameter k without changing the character of the process.

In ordinary Brownian motion, with $\theta=1$, successive values of ε are wholly independent. Mandelbrot (1982) has shown, however, that if θ is increased in the range 1 to 2 a family of trails smoother than those of Brownian motion can be obtained in which successive values of ε are positively correlated. Conversely, by diminishing θ between 1 and 0 the successive values of ε are negatively correlated and the traces are noisier than those of Brownian motion. Corresponding to the traces is a family of semi-variograms, some of which are shown in Fig. 2. The figure also gives the fractal dimension or, to give it its full name, the Hausdorff–Besicovitch dimension, $D=2-(\theta/2)$.

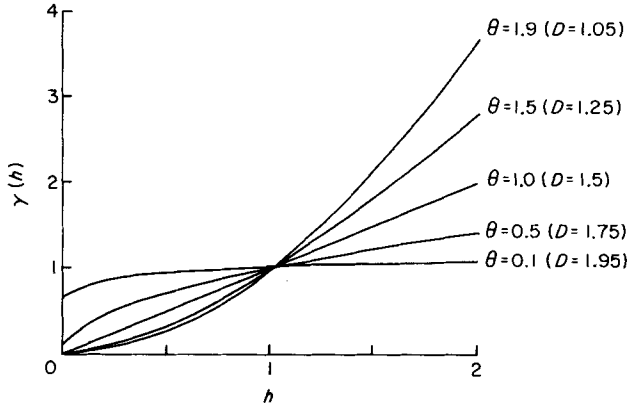

Fig. 2. Semi-variograms of fractional Brownian processes.

When $\theta=2$ the semi-variogram is a parabola and describes a trace that is smooth and differentiable. The fractal dimension, D, is 1. This no longer represents a random process, and the function $\gamma(h)=kh^2$ is not CNSD. At the other extreme $\theta=0$ represents pure noise, which is impossible for a continuously varying property. So both limiting values are precluded. The parameter θ must take values between 0 and 2.

We note that many semi-variograms appear to be pure nugget variance (e.g. Campbell, 1978;

McBratney & Webster, 1981*a*) even though the semi-variance must be zero at $h=0$. This arises because all the variance occurs within the smallest sampling interval. A function is needed therefore, that is independent of h for all positive values. This can be done formally by employing a Dirac function, $\delta(h)$, such that

$$\gamma(h)=c_0\{1-\delta(h)\}, \tag{27}$$

where $\delta(h)=1$ when $h=0$ and 0 otherwise. Equation (27) is then the formal definition of a pure nugget semi-variogram.

A semi-variogram that approaches the origin parabolically can signify changing *drift*, that is, a change in the expected value, $E[z(\mathbf{x})]$, from place to place. In these circumstances a full structural analysis in the sense of Olea (1975) is needed to find a semi-variogram that properly characterizes the stochastic process. Webster & Burgess (1980) and Moffat *et al.* (1986) encountered this situation and illustrated suitable techniques for dealing with it.

Burrough (1981, 1983*a*) examined a number of one-dimensional sequences of soil data for self similarity. The D values he obtained ranged from 1.8 to 2, so that θ lay between 1 and 0. Oliver & Webster (1986) fitted an isotropic model of this kind to the plot of soil in the Wyre Forest. Again the value of θ was <1 and the semi-variogram convex upwards. Otherwise, only models with $\theta=1$, i.e. linear models, have been applied to soil (for example Burgess & Webster, 1980*a*,*b*; Webster & Nortcliff, 1984). The model is widely used for estimating ore reserves, where it is perhaps second only in popularity to the spherical model.

Burrough's results suggested that soil varies in a way more random than Brownian motion. Burrough (1983*b*, 1984) was not happy with that explanation: it seemed more likely that the data had been generated by two or more processes nested within one another at different scales rather than by a single Brownian process. Likewise a linear model fitted to a sample semi-variogram does not necessarily represent Brownian motion or even imply that the variation is truly unbounded. It may simply be the initial near-linear segment of a spherical or circular model—see Webster & Burgess (1984) for an example. It often happens that the range of the semi-variogram is a large proportion of the distance across the sampled region, or even exceeds it. In these circumstances semi-variances near the sill are either poorly estimated or not computed at all.

Logarithmic model. A model that is of historical interest is the logarithmic or de Wijsian model with equation

$$\gamma(h)=k\log(h). \tag{28}$$

Much of the early statistical estimation in the mining industry (Krige, 1951; Matheron, 1955) was based on this model because it was easy to convert to linear form and then to fit without a computer. The recent results of Gajem *et al.* (1981) for particle size distribution are convincingly fitted by this model, Fig. 3.

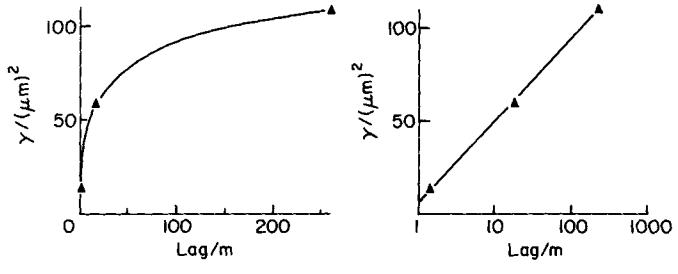

Fig. 3. Semi-variances of mean particle diameter plotted on linear (left) and logarithmic (right) scales with a de Wijsian model fitted. The values are taken from the sills and ranges determined by Gajem *et al.* (1981) from three transects with sampling intervals of 0.2 m, 2 m and 20 m.

Combinations

The semi-variograms described above are simple in form. In practice more complex forms are often encountered, and it is desirable therefore to have more complex models available to fit to them. In many instances these can be built by combining the simpler models. This is possible since any linear combination of covariances with positive coefficients is itself a covariance, and the same holds true for semi-variances. Thus if we have semi-variograms $\gamma_i(h)$, $i=1, 2, \ldots n$, then

$$\gamma(h)=\sum_{i=1}^{n}\lambda_i\gamma_i(h) \qquad \text{for all positive } \lambda_i. \tag{29}$$

In a sense two models have already been combined by incorporating the nugget effect. Thus the semi-variograms in equations (14), (17), (20), (21) and (22) are

$$\gamma(h)=\gamma_0(h)+\gamma_1(h),$$

where

$$\gamma_0(h)=c_0\{1-\delta(h)\}$$

is the semi-variogram of the nugget and γ_1 describes the spatially dependent variance.

Apart from this the most useful combinations of models used to date for describing soil variation have been those that incorporate two or three different ranges. These arise where the contributing factors have distinctly different spatial scales, and so are nested one within another.

Burrough (1983*b*) found that by combining two or three bounded linear models in this way he could describe adequately all the sample semi-variograms that he had obtained from many sets of transect data. He concluded from this research that the variation he had encountered was essentially nested.

The bounded linear model is, as above, CNSD only in one dimension. More generally combinations for two and three dimensions are desired. A nested model that has been found especially useful, both in mineral exploration and soil research, is the double spherical model. McBratney *et al.* (1982) used it to describe variation in copper and cobalt in the soil of south east Scotland. They distinguished farm effects with a range of 2 to 3 km from those of geology with a range of about 15 km. Webster & Nortcliff (1984) described the variation in iron content of the soil in small plots with ranges of 18 m and 77 m.

Fig. 4 shows a further example of this model. It is fitted to the semi-variogram of the combined thickness of the A and B horizons. The data are from a soil survey by Wong (1967) over Oxford Clay, Pleistocene gravels and recent alluvium in the Upper Thames Valley in Oxfordshire. Observations were made at 20 m intervals using a 10 cm diameter auger along a transect almost 6 km long to give 294 observations. The model includes a nugget variance, and the three components are

$$\text{(a)} \qquad \gamma_0(h)=c_0\{1-\delta(h)\}$$

$$\text{(b)} \qquad \begin{cases} \gamma_1(h)=c_1\left\{\dfrac{3h}{2a_1}-\dfrac{1}{2}\left(\dfrac{h}{a_1}\right)^3\right\} & \text{for } h\leqslant a_1 \\ \gamma_1(h)=c_1 & \text{for } h>a_1 \end{cases}$$

$$\text{(c)} \qquad \begin{cases} \gamma_2(h)=c_2\left\{\dfrac{3h}{2a_2}-\dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3\right\} & \text{for } h\leqslant a_2 \\ \gamma_2(h)=c_2 & \text{for } h>a_2 \end{cases} \tag{30}$$

Table 2 lists the estimates of the five parameters. The range, 492 m, of the third component represents the limit of spatial dependence, and equates with the average spacing between major changes in the deposits. It accords with the ranges of several other properties and of the leading principal components from 27 recorded properties in all (Webster, 1973). The cause of the shorter range component, $a_1=102$ m, is not known.

Repetitive features in soil may require models that combine transition and periodic functions.

Table 2. Parameter estimates ($\hat{p}$) and their standard errors (SE), fitted to the sample semi-variogram of soil thickness near Witney, the residual sum of squares (*RSS*) and the Akaike information criterion (*AIC*)

	c_0		c_1		a_1		c_2		a_2			
Model	$\hat{p}$	SE	$\hat{p}$	SE	$\hat{p}$	SE	$\hat{p}$	SE	$\hat{p}$	SE	*RSS*	*AIC*
Bounded linear	40.0	3.08	81.7	3.27	358	14.4					1864	382
Circular	37.3	3.13	84.7	3.22	418	19.4					1614	375
Spherical	35.0	3.11	87.0	3.22	458	20.2					1544	373
Exponential	15.5	5.21	107.7	5.04	152	12.0					1980	385
Double spherical	14.8	7.59	31.0	8.87	102	25.0	76.4	4.63	492	22.4	1188	364

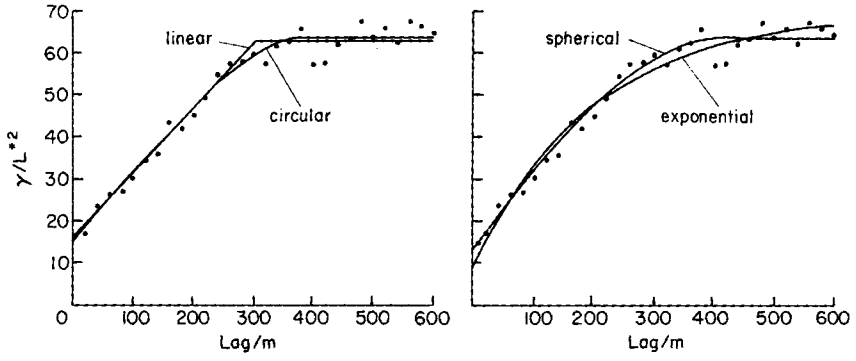

Fig. 4. Sample semi-variogram of soil thickness near Witney with (left) bounded linear, circular and (right) exponential and double spherical models fitted to it.

One of the most striking repetitive soil patterns is gilgai. Webster (1977) analysed the spatial pattern in one dimension of soil properties in gilgai terrain of the Bland Plain of New South Wales. Several properties showed strong spectral peaks corresponding to a wavelength of 33 m, and the semi-variograms have a period of about this length (Webster, 1985). Fig. 5 shows our most recent attempt to model the semi-variogram of the electrical conductivity at 30 to 40 cm. It has four components:

$$\begin{aligned}
&\text{nugget variance} && \gamma_0(h) = c_0\{1-\delta(h)\}\\
&\text{linear} && \gamma_1(h) = b_1 h\\
&\text{spherical} && \begin{cases} \gamma_2(h) = c_2\left\{\dfrac{3h}{2a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3\right\} & \text{for } h \leqslant a_2\\ \gamma_2(h) = c_2 & \text{for } h > a_2 \end{cases}\\
&\text{periodic} && \gamma_3(h) = c_3\left\{1-\cos\frac{2\pi h}{a_3}\right\}.
\end{aligned} \tag{31}$$

Table 3 lists the estimated values of the parameters and their standard errors.

The first component, the nugget variance is small, but in the absence of sampling at a spacing less than 4 m is poorly estimated. The linear component with gentle gradient suggests a long range variation across the region, which contrasts with the dominant short-range spherical component. The fourth component, which is CNSD only in one dimension, is especially interesting in that its period, $a_3 = 34.2$ m, agrees closely with that found earlier using spectral analysis (Webster, 1977).

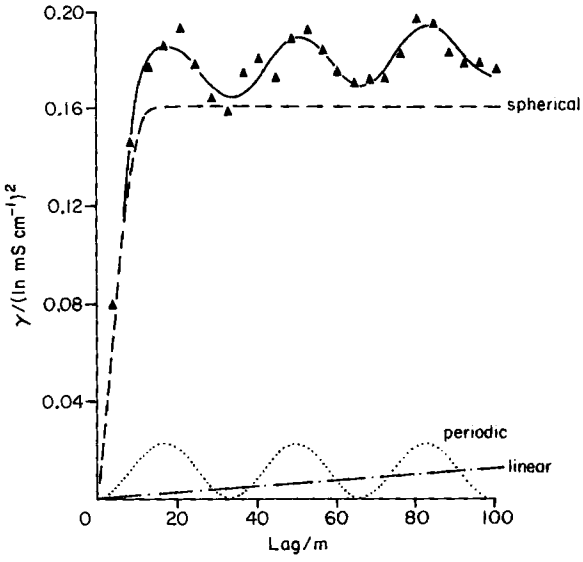

Fig. 5. Sample semi-variogram of the electrical conductivity on the Bland Plain showing the four-component fitted model as a solid line and the linear, spherical and periodic components separately. The fourth component is the nugget variance.

Table 3. Parameters of the four-component semi-variogram of electrical conductivity, (in mS cm^{-1} transformed to natural logarithms) sampled at 30–40 cm at 355 points at 4 m intervals

	Parameter	Estimate	SE
Nugget	c_0	0.00549	0.02059
Linear	b_1	0.00014	0.00004
Spherical	c_2	0.1565	0.0232
	a_2	12.63	1.749
Periodic	c_3	0.01059	0.00146
	a_3	34.20	0.281

Anisotropy

All the above models describe isotropic variation. But soil does not necessarily vary equally in all directions. There are numerous situations where the variation is anisotropic: on point bar deposits and river *levées*, on bevelled sedimentary beds, in soliflucted materials on slopes, such as those described by Evans (1972) in the Norfolk Breckland, and the elongated gilgai of Eastern Australia (Hallsworth *et al.*, 1955; Beckmann *et al.*, 1973). In these situations each direction has its own semi-variogram differing from those in other directions. If in a region the anisotropy can be accounted for by a simple linear transformation of the coordinates then it is said to show *geometric* or *affine* anisotropy, and the transformation has the following formula:

$$\Omega(\theta) = \{A^2 \cos^2(\theta - \varphi) + B^2 \sin^2(\theta - \varphi)\}^{\frac{1}{2}} \tag{32}$$

This function is a factor that can be applied either to the distance parameter of a transitive model or to the gradient of an unbounded model. The parameters A and B define the maximum and minimum ranges or gradients of the model, and φ is the angle of maximum range or gradient. The ratio $A:B$

is a measure of the anisotropy. Burgess *et al.* (1981), McBratney & Webster (1981*b*, 1983), Tabor *et al.* (1984), Webster (1985) and Moffat *et al.* (1986) all show examples of geometric anisotropy, in some instances with large anisotropy ratios. Another example is given for the thickness of peat at Lephinmore, Argyllshire, where some 1200 observations were made at 100 m intervals on a square grid. The estimated semi-variances for eight directions are shown in Fig. 6. The directional effect is clear. Table 4 shows that even though there is much short-range variation, the anisotropic linear model fits well, and certainly better than any isotropic model. Once again, the anisotropy ratio A:B of 4.4 is large.

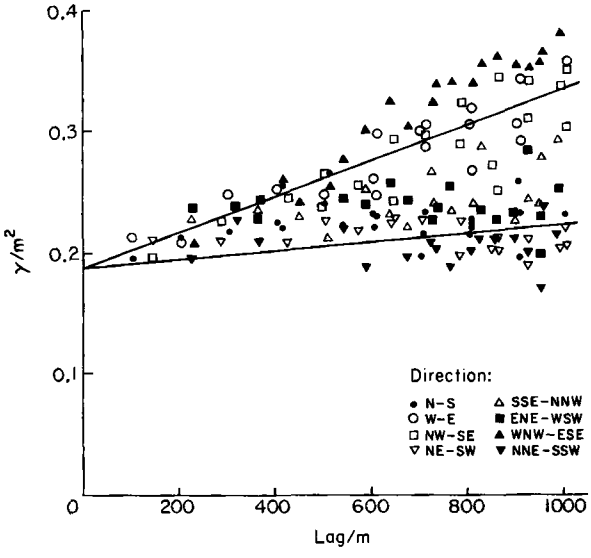

Fig. 6. Semi-variogram of peat thickness (transformed to $\log_{10}$) at Lephinmore showing two-dimensional semi-variances and the envelope of a geometrically anisotropic linear model.

Table 4. Semi-variogram modelling of Lephinmore peat thickness $\log_{10}$ (cm + 0.5) (upper) showing the residual sum-of-squares (*RSS*), number of parameters (p) and *AIC* for the fitted models and (lower) the parameter estimates for the anisotropic linear model

Model	*RSS*	p	*AIC*
Isotropic linear	112.533	2	798
Anisotropic linear	36.279	4	611
Isotropic spherical	112.284	3	792
Isotropic exponential	112.284	3	792

Model	Parameter	Estimate	SE
Anisotropic linear	c_0	0.189	0.00356
	A	0.01458	0.00074
	B	0.00332	0.00664
	φ	−0.454	0.0283

Mardia (1980) has suggested formal significance tests for the anisotropy ratio, but only by assuming a somewhat implausible model for the semi-variogram. The tests do not appear to have been used in geostatistical practice.

It may happen that a property has a semi-variogram with different sills in different directions. If so the property is said to show *zonal* anisotropy (Journel & Huijbregts, 1978). So far no one has reported such variation in soil in the horizontal plane. Almost certainly it occurs when the vertical dimension is added, but again there are no reports of three-dimensional analysis or even analyses of soil variation in the vertical plane.

ESTIMATING THE SEMI-VARIOGRAM

We have already given the general formula for estimating semi-variances at discrete values of $\mathbf{h}$, equation (4). The estimate is unbiased. For N observations at regular intervals along a transect the formula becomes

$$\hat{\gamma}(h)=\frac{1}{2(N-h)}\sum_{i=1}^{n-h}\{z(i)-z(i+h)\}^2, \tag{33}$$

where h is now an integral number of sampling intervals. For a $m\times n$ rectangular grid in two dimensions the semi-variogram is itself two-dimensional with estimates given by

$$\hat{\gamma}(p,q)=\frac{1}{2(m-p)(n-q)}\sum_{i=1}^{m-p}\sum_{j=1}^{n-q}\{z(i,j)-z(i+p,j+q)\}^2$$

$$\hat{\gamma}(p,-q)=\frac{1}{2(m-p)(n-q)}\sum_{i=1}^{m-p}\sum_{j=q+1}^{n}\{z(i,j)-z(i+p,j-q)\}^2 \tag{34}$$

where p and q are the lags in the two dimensions. Equation (34) gives the estimates for the right-hand half of the semi-variogram. The function is symmetric about its centre, and so the other half is obtained by a simple half rotation. In two dimensions it usually happens that the region of interest has an irregular outline. The lengths of the rows and columns are not then constant. The summation is performed over the available data, and the denominators must be replaced by the actual numbers of pairs in the sums. The estimates of a two-dimensional semi-variogram for the electrical resistivity of the soil at Bekesbourne, Kent, (see Webster & Burgess, 1980) computed using equation (34) is shown as an isarithmic chart in Fig. 7. The strong anisotropy appears clear from the elliptical form of the isarithms. The semi-variances in Fig. 6 were also calculated by equation (34) but were then allocated to the eight directional classes for illustration.

In two dimensions it is usually convenient to express the lag in terms of polar rather than Cartesian coordinates, as we did in discussing anisotropy. This is especially so where sampling has been irregular. In such cases the pairs of observations must be grouped into classes of both lag distance ($h\pm\Delta h/2$, where $h=|\mathbf{h}|$) and direction ($\theta\pm\Delta\theta/2$); see David (1977) and Webster (1985). The estimated semi- variogram is then to some extent smoothed, or 'regularized'. Experience has shown that if Δh *and* $\Delta\theta$ are small, i.e. Δh equal to half the average distance between neighbouring samples and $\Delta\theta<45°$, then the regularization is small enough to be ignored.

Data may have been recorded at irregular intervals along a transect. Pairs may again be grouped into classes of distance with intervals $h\pm\Delta h$, and the resulting semi-variogram will again be regularized.

Robust estimation

In addition to freedom from bias, estimates are needed that are efficient, i.e. have small variances. Unfortunately there is no easy way of obtaining confidence limits analytically for a semi-variogram, and this is a matter that is currently taxing geostatisticians. Several simulation studies have been made, and they show that confidence limits are disturbingly wide. There has been especial concern over semi-variograms from data that are strongly skewed or heavier than normal in both tails. This

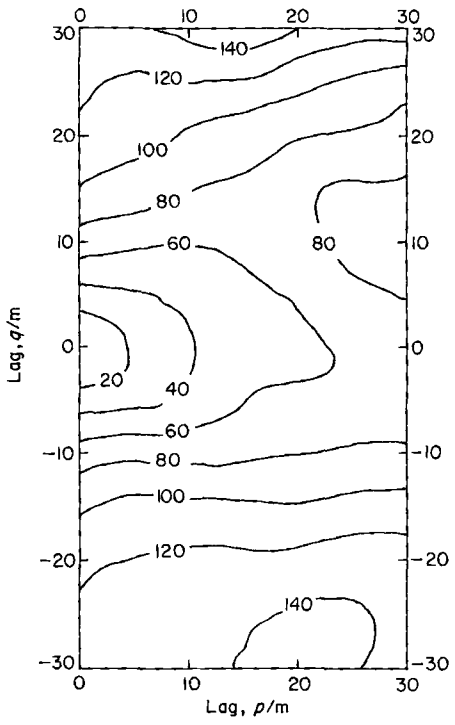

Fig. 7. Isarithmic chart of two-dimensional semi-variogram of the electrical resistivity of the soil at Bekesbourne. Isarithms are at intervals of 20 (ohm m)^2.

has led to the search for estimators that are both efficient and robust. One that has received some acclaim is due to Cressie & Hawkins (1980), and we therefore devote a section to it.

Semi-variances are ideally computed from normally distributed variates. Where data depart seriously from normality then some form of transformation should be considered. There are several ways in which non-normal distributions can be normalized; details can be found in many textbooks. Even after transformation a variate may be heavier in the tails of its distribution than normal, and it was for this kind of distribution that Cressie & Hawkins (1980) devised their estimator. They discovered that the fourth root of the usual squared difference

$$y(\mathbf{x})=[\{z(\mathbf{x})-z(\mathbf{x}+\mathbf{h})\}^2]^{1/4} \tag{35}$$

had a distribution close to normal with negligible skew. They therefore computed a mean $\bar{y}$ of all the individual $y(\mathbf{x})$. For a regular transect of N observations this gives

$$\bar{y}(h)=\frac{1}{(N-h)}\sum_{i=1}^{N-h}[\{z(z(i)-z(i+h)\}^2]^{1/4}. \tag{36}$$

The mean $\bar{y}$ must be transformed back, and Cressie & Hawkins show that the required transformation is:

$$\hat{\gamma}(h)=y^{-4}/2\left\{0.457+\frac{0.484}{N-h}+\frac{0.045}{(N-h)^2}\right\} \tag{37}$$

To illustrate how wide confidence limits on semi-variograms are and the possible merit of the robust estimator we simulated series with known population semi-variograms. Six different discrete first-order autoregressive processes were taken, with the general form

$$Y_x = \alpha Y_{x-1} + \varepsilon_x. \tag{38}$$

From each we simulated 19 series of length 200. The six processes were

(a)	$Y_x = \varepsilon_x$	$\varepsilon_x \sim N(0,1)$
(b)	$Y_x = \varepsilon_x$	$\varepsilon_x \sim N(0,1) + 10\%\ N(0,5)$
(c)	$Y_x = 0.8\ Y_{x-1} + \varepsilon_x$	$\varepsilon_x \sim N(0,1)$
(d)	$Y_x = 0.8\ Y_{x-1} + \varepsilon_x$	$\varepsilon_x \sim N(0,1) + 10\%\ N(0,5)$
(e)	$Y_x = 0.8\ Y_{x-1} + \varepsilon_x$	$\varepsilon_x \sim N(0,1) + 5\%\ N(2,1)$
(f)	$Y_x = 0.2\ Y_{x-1} + \varepsilon_x$	$\varepsilon_x \sim N(0,1)$

When ε_x is a standard normal deviate, i.e. $\varepsilon_x \sim N(0,1)$, Y_x in equation (38) has a variance

$$\text{var}[Y_x] = \frac{1-\alpha^{2x-1}}{1-\alpha^2}, \tag{39}$$

and autocorrelation function

$$\rho(h) = \alpha^h. \tag{40}$$

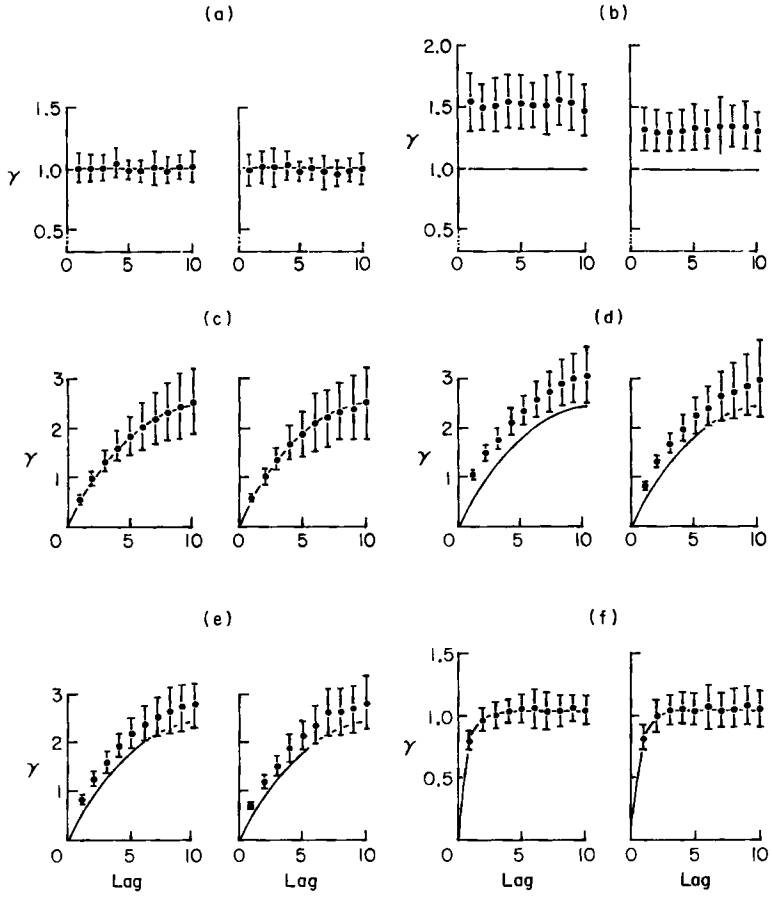

Fig. 8. Semi-variograms of six auto-regressive processes and results of simulation from them using the normal and a robust estimation. The characteristics of the processes are listed as (a) to (f) in the text. For each pair of diagrams the normal estimator is shown on the left and the robust estimator on the right.

The semi-variance is

$$\gamma(h) = \text{var}[Y_x]\{1-\rho(h)\}, \tag{41}$$

and for large x

$$\gamma(h) = \frac{1-\alpha^h}{1-\alpha^2} \tag{42}$$

The results of simulations from the above processes are shown in Fig. 8. The diagrams on the left show the results using the usual estimator and those on the right using the robust estimator. The population semi-variogram is shown as a solid line. The estimates are shown as dots or crosses with error bars depicting ± 1 standard deviations of the 19 estimates.

Fig. 8 shows the result for simulation (a). The mean estimates closely approximate the population semi-variogram, and the robust estimator appears to be consistent with the usual estimator. Simulation (b) shows the effect of filling out the tails of the distribution; the observed semi-variogram for the usual estimator is much higher and has larger error than the robust one. The robust estimator is lower but is still above the level of the population semi-variogram.

Simulations (c) and (d) give similar results to (a) and (b). It is interesting to note that the standard deviation of the estimates rises much more quickly with the first-order autoregressive process with $\alpha = 0.8$ than with $\alpha = 0$. An intermediate example with $\alpha = 0.2$ was simulated (f) and, as was suspected, the standard deviations of the estimates are smaller than in simulation (c) and they rise less quickly.

Simulation (e) is meant to describe a real soil process with normal probability density function plus a heavy upper tail. The process in mind is the distribution of the logarithms of the available potassium concentration, log K, where the measured K is especially concentrated in the faeces and urine of grazing animals. This interesting spatial process is described with some preliminary statistical analyses by Petersen *et al.* (1956*a*,*b*) and MacLusky (1960). MacLusky shows that the area of pasture covered by faeces rises linearly from the origin to 7% at 1000 cow days per hectare, and the example here corresponds to 700 cow days per hectare. This is an example of a quasi-point process superimposed on a continuous process. The reason for including this example is to make the point that it may be reasonable to use robust estimation only when one suspects that the heaviness in the tails of the distribution is not related to the process in which we are interested. In a situation such as that above we might wish to exclude the large values of K associated with the faeces to obtain the semi-variogram of the continuous underlying process.

Finally in this section we emphasize the wide confidence intervals on semi-variograms. The lengths of the bars in Fig. 8 represent approximately 66% confidence intervals for semi-variograms computed to only lag 10 from 200 transect data.

FITTING MODELS

In describing soil variation by a semi-variogram there remains the task of choosing a suitable function and fitting it to the observed estimates of $\gamma(\mathbf{h})$. Some geostatisticians fit models by eye, but the practice is unreliable, and a statistically based procedure is much to be preferred. Fitting is usually performed by least squares approximation.

Least squares estimation

Ordinary least squares estimates of the semi-variogram parameters can be obtained by minimizing

$$L(\tau;\gamma_k) = \sum_{j=1}^{k} \{\gamma(\mathbf{h}_j) - \gamma(\mathbf{h}_j;\tau)\}^2. \tag{43}$$

Here τ represents the $p \times 1$ vector of p parameter estimates and γ_k represents the $k \times 1$ vector of semi-variance estimates, each estimate being denoted by $\gamma(\mathbf{h}_j)$. The quantity $\gamma(\mathbf{h}_j;\tau)$ is the jth two-dimensional semi-variance expected from the fitted model and depends on the parameter estimates

τ. This procedure assumes that the residuals in equation (43) are normally distributed and independent of one another, and that the estimated semi-variances all have the same variance.

The method is often criticized, mainly on the grounds of these assumptions. It is not clear whether the errors are normally distributed. Complete independence of the residuals is unlikely though the degree of dependence is unknown. Further and more importantly the variances depend on the values of the parameters.

Most of the above objections can be overcome by *generalized* least squares estimation, achieved by minimizing $G(\tau;\gamma_k)$ in the following equation:

$$G(\tau;\gamma_k)=\{\gamma_k-\gamma(\tau)\}^T \Psi_k^{-1} \{\gamma_k-\gamma(\tau)\}, \tag{44}$$

where γ_k is as above, $\gamma(\tau)$ is the $k \times 1$ vector of estimated semi-variances that depend on the vector of parameter estimates, $(\gamma_k-\gamma)\tau)\}$ is the $k \times 1$ vector of residuals, and Ψ_k denotes the $k \times k$ matrix of variances and covariances between the estimates of $\gamma(\mathbf{h}_j;\tau)$. In most applications to soil the semi-variogram must be estimated in two dimensions to allow for anisotropy, so that k will be at least 100. Equation (44) must be solved iteratively. It requires the inverse of Ψ_k at each iteration, and as the time to compute the inverse increases in proportion to k^3 the solution of the equation can be seen as a formidable task.

A compromise between these two methods, which gives asymptotically efficient estimates but does require the inversion of a large matrix, is afforded by weighted least squares estimation. In this the model is fitted by minimizing

$$W(\tau;\gamma_k)=\sum_{j=1}^{k} \{\gamma(\mathbf{h}_j)-\gamma(\mathbf{h}_j;\tau)\}^2/\text{var}\,[\gamma(\mathbf{h}_j)], \tag{45}$$

where $\text{var}[\gamma(\mathbf{h}_j)]$ is the variance of the estimate of the semi-variance. Thus the weight is inversely proportional to the variance. Unfortunately this variance is unknown except in a few simple cases (see Davis & Borgman, 1979, 1982). In much of our work we have used weights equal to the number of pairs of observations, $m(\mathbf{h}_j)$, making up each estimate on the assumption that these are simply inversely related to the variances of the estimates. Cressie (1985) has shown theoretically that a better procedure is to replace $\text{var}[\gamma(\mathbf{h}_j]$ by $\{\gamma(\mathbf{h}_j;\tau)\}^2/m(\mathbf{h}_j)$ thereby giving considerably more weight to small expected semi-variances. A comparison of these two weighting schemes showed that the differences in the fitted models were small, but it did confirm our own observation that simple weighting by $m(\mathbf{h}_j)$ gives too little weight to estimates at the shortest lags. Laslett *et al.* (in press) successfully used Cressie's weighting scheme. G. M. Laslett (personal communication) has suggested a further improvement in the iterative approach. The weights are given by

$$\frac{m(\mathbf{h}_j)\,\gamma(\mathbf{h}_j)}{\{\gamma(\mathbf{h}_j;\tau)\}^3}$$

and are updated on each iteration as the $\gamma(\mathbf{h}_j;\tau)$ are re-estimated. Theoretically this form of weighting should give better convergence, and in practice we have found it to do so.

Maximum likelihood estimation

Maximum likelihood estimation is generally regarded as the most efficient means of estimation. Mardia & Marshall (1984) have provided the statistical theory for full maximum likelihood estimation of the covariogram, i.e. for the case where the random function $Z(\mathbf{x})$ is Gaussian and weakly stationary. They state that maximum likelihood estimates of μ, the mean, and υ, the p covariogram parameters, can be obtained by maximizing

$$L_N(\boldsymbol{\beta};\mathbf{z}_N)= -0.5\,\ln|\Xi_N|-0.5(\mathbf{z}_N-\mathbf{1}\mu)^T\Xi_N^{-1}(\mathbf{z}_N-\mathbf{1}\mu), \tag{46}$$

where $\boldsymbol{\beta}$ is a $1+p \times 1$ combined parameter vector, i.e. $\boldsymbol{\beta}=[\mu, \upsilon]$, $\mathbf{z}_N=[z(\mathbf{x}_1), z(\mathbf{x}_2), \ldots, z(\mathbf{x}_N)]^T$ is the $N \times 1$ vector of observations at the N sampling sites, Ξ_N is the variance–covariance matrix among all sites, and $\mathbf{1}$ is a $N \times 1$ vector of ones. The maximum of L_N is found by iteration in which Ξ_N is updated

and inverted at each iteration. Since the time required for inversion is proportional to N^3, this procedure can take even longer than generalized least squares estimation. This technique is practicable for fairly small samples, say up to 150. Most of the surveys from which we have worked have had more than 150 data, and some have had several thousands. Mardia & Marshall's (1984) results of simulation may be of interest to those who wish to try the technique on small sets of data. They showed that for 36 observations on a 6×6 square grid estimates of the range, a, and the structural variance, c, of the spherical model, equation (20), are biased negatively. This is not so for 100 observations on a 10×10 grid. Mardia & Marshall's (1984) results cast doubt on the accuracy of semi-variogram estimates obtained from small samples (N less than 100).

Recommendation

Other methods, some quite unsound, have been proposed, but there is no agreement on a generally best (in terms of computational and statistical efficiency and reality of assumptions) way of model fitting. In our experience the weighted least-squares method has proved reliable and computationally efficient, and we can recommend it either with weights suggested by Cressie (1985) or those proposed by G. M. Laslett (personal communication) for the iterative procedure. Many of the models are non-linear in the parameters, and so a good computer program is needed for the fitting. We use MLP, the Maximum Likelihood Program, written by our colleague G. J. S. Ross (1980) for the purpose.* Full maximum likelihood estimation is worth considering for sample sizes in the range 100–150, however.

If the semi-variogram appears to approach the origin parabolically then local drift should be suspected. Local drifts in soil data seem very exceptional. Where they occur a different modelling approach is required, and this has been described by Olea (1975) and Webster & Burgess (1980).

MODEL SELECTION

Choosing from among the reasonable models, which may consist of combinations of simple models, with or without a nugget variance and with or without anistropy, can be difficult. An investigator rarely has any theoretical pedogenetic grounds for choosing a particular model. Rather he must rely on statistical considerations, which depend on his purpose.

If the investigator wishes to describe the main characteristics of the spatial variation then he must strike a balance between the goodness of fit and parsimony in his model. He is already committed to simplification, and he will want to choose the simplest model, i.e. the one with the fewest parameters, that adequately represents the observed semi-variograms. Further, he should beware of attempting an accurate description of an apparently complex semi-variogram if this involves fitting many parameters since they are likely to be poorly estimated.

Akaike information criterion

This compromise between goodness of fit and parsimony can be expressed in terms of Akaike's (1973) information criterion, *AIC* for short. It is defined as

$$A = -2 \ln (\text{maximum likelihood}) + 2 \times (\text{number of parameters})$$

and estimated by

$$\hat{A} = n \ln (R) + 2p, \tag{47}$$

where n is the number of observations, p is the number of estimated parameters and R is the residual sum of squares of deviations from the fitted model. The model having the smallest *AIC* is the best. Yost *et al.* (1982*a*) used this criterion for selecting models to describe soil variation in Hawaii, and we give *AIC* values where we compare models in the examples above.

*Specimen MLP scripts are available on request from the authors.

Using this criterion we can now see that in the first example, Fig. 1 and Table 1, the circular model describes the observed semi-variogram best, though the spherical model is only slightly less good. The evident poorer fit of the exponential model is expressed in the larger *AIC*. In this instance we could have judged the goodness of fit equally well from the residual sums of squares since in each case we were fitting a model with three parameters. For the Witney transect, however, the double spherical model with five parameters is clearly the best fit: the sum of the squares of the residuals is quite the smallest. The *AIC* is also the smallest, so despite its being the most complex of the models examined one may feel that the added complexity justifies its use.

Cross validation

Most spatial analysis to date in soil science has had the aim of prediction or estimation by kriging. Here the aim is to choose a model that minimizes the errors of prediction, given a particular sampling configuration with a limited range of sampling intensity. To make kriging worthwhile there must be several sampling points within a radius equal to the range of the semi-variogram of the estimation point or block. Within this distance the nearer the points are to the estimation point the more weight they carry. It is important therefore that the semi-variogram be well estimated at short lags, and the investigator should choose a model that fits well in this region of the semi-variogram. Webster & Burgess (1984) show how one may choose different semi-variograms when the blocks to be estimated have different areas.

In these circumstances the goodness of the model can be checked by cross validation, as follows. Each value in the set of N data from which the semi-variogram was computed is removed in turn from the set, and the value at that point estimated from the remainder by kriging (see e.g. Burgess & Webster, 1980*a*, for the punctual procedure). Its estimation variance is given by

$$\sigma_E^2 = E[\{z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)\}]^2$$
$$= 2\sum_{i=1}^{N-1} \lambda_i \gamma(\mathbf{x}_0 - \mathbf{x}_i) - \sum_{j=1}^{N-1}\sum_{i=1}^{N-1} \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j), \quad (48)$$

where $z(\mathbf{x}_0)$ is the value of the soil property at $\mathbf{x}_0$ and $\hat{z}(\mathbf{x}_0)$ is its estimate, γ is the semi-variogram, $(\mathbf{x}_0 - \mathbf{x}_i)$ and $(\mathbf{x}_i - \mathbf{x}_j)$ are vectors in $\mathbf{h}$, and λ_i, λ_j are weights. The square root, σ_E, is the kriging error.

By removing the datum at $\mathbf{x}_i$ and estimating its value we can obtain an actual deviation, $s(\mathbf{x}_i)$, which is

$$s(\mathbf{x}_i) = z(\mathbf{x}_i) - \hat{z}_c(\mathbf{x}_i), \quad (49)$$

where $\hat{z}_c(\mathbf{x}_i)$ denotes the estimate of $z(\mathbf{x}_i)$ by the procedure. The value of s can then be compared with the estimated kriging error to form a deviation ratio

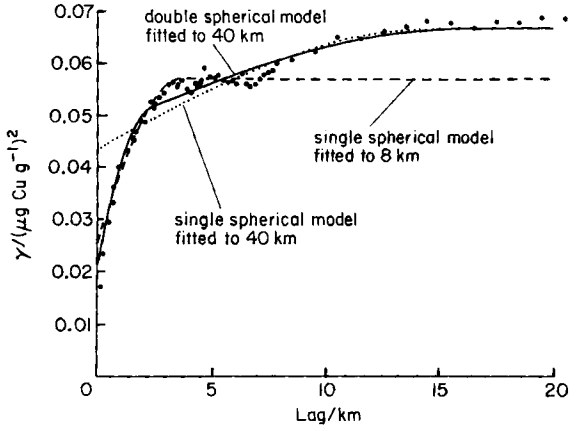

Fig. 9. Semi-variogram of $\log_{10}$ copper concentration in the top-soils of south-east Scotland with single spherical models (dashed and dotted lines) and double spherical model (solid line) fitted.

$$r(\mathbf{x}_i) = s(\mathbf{x}_i)/\sigma_E(\mathbf{x}_i). \quad (50)$$

For N data there will be N values of r, which should be normally distributed with mean 0 and unit variance. The observed mean, $\bar{r}$, and variance, $s^2(r)$, express the goodness of the model for prediction: the closer they are to 0 and 1, respectively, the better the model.

This is illustrated with an example in which choice of a suitable model for kriging was by no means easy. In the original study by McBratney *et al.* (1982) the available copper content in the topsoil (0–15 cm) of some 3500 fields in the Border Region of Scotland had been measured on bulked samples. The data were transformed to their common logarithms to normalize their distribution, and the sample semi-variogram of the transformed values is shown in Fig. 9.

Initially a single spherical model was fitted to 320 estimated semi-variances to 40 km lag. This gave a reasonable representation overall but a poor fit near the ordinate, and was not therefore well suited for kriging. A spherical model to 8 km fitted well near the ordinate, and since in most instances in this survey kriging weights at lags larger than 8 km would be negligible this model could be regarded as a practical solution. However, it described the whole range of the semi-variogram poorly. The solution finally adopted was to combine the two in a double spherical model. The estimated parameters are given in Table 5.

Table 5. Spherical models for the semi-variograms of $\log_{10}$ copper in the Border Region of Scotland

Model fitted	Parameter					*AIC*	$n-p$	$\bar{r}$	$s^2(r)$
	c_0	c_1	a_1	c_2	a_2				
To 8 km	0.0257	0.0312	3.58			508	313	−0.002	0.933
To 40 km	0.0435	0.0231	14.50			1560	317	−0.007	0.826
To 40 km	0.0213	0.0257	2.26	0.0196	15.46	2782	567	−0.002	0.947
By eye	0	0.066	10.0					0.007	9.358

Table 5 shows the means and variances of the deviation ratios when these models were used for cross validation. Of the three the single spherical to 40 km is the least good and the double spherical the best. Cross validation justified the choice of the more complex model. The last entry in Table 5 gives the mean and variance for a spherical model that was fitted by eye—a practice that is still recommended by some geostatistical practitioners. The results are clearly poor.

CONCLUSION

The above account describes the simpler functions available for modelling semi-variograms of soil properties in two dimensions. It recommends procedures for choosing and fitting models. Geostatistical theory may in some ways have already outpaced its application to soil science, but there is much interest in the subject and we hope that this paper provides the kind of guidance that soil scientists are seeking.

ACKNOWLEDGEMENTS

We thank Mrs Joyce Munden for her help with the computing and for preparing the diagrams for publication, Mr I. F. T. Wong for the data from his transect survey near Witney, and Mr J. S. Bibby for the survey data from Lephinmore.

REFERENCES

AKAIKE, H. 1973. Information theory and an extension of maximum likelihood principle. In: *Second International Symposium on Information Theory*. (Eds B. N. Petrov and F. Coaki) pp. 267–281. Akademia Kiado, Budapest.

ARMSTRONG, M. & JABIN, R. 1981. Variogram models must be positive definite. *Mathematical Geology* **13,** 455–459.

BAKER, R. 1984. Modelling soil variability as a random field. *Mathematical Geology* **16,** 435–448.

BECKMANN, G.G., HUBBLE, G.D., & THOMPSON, C.H. 1973. Australian landform example no. 22: linear gilgai. *Australian Geographer* **12,** 363–366.

BURGESS, T.M. & WEBSTER, R. 1980*a*. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *Journal of Soil Science* **31,** 315–331.

BURGESS, T.M. & WEBSTER, R. 1980*b*. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *Journal of Soil Science* **31,** 333–341.

BURGESS, T.M. & WEBSTER, R. 1984. Optimal sampling strategies for mapping soil types. I. Distribution of boundary spacings. *Journal of Soil Science* **35,** 641–654.

BURGESS, T.M., WEBSTER, R. & MCBRATNEY, A.B. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science* **32,** 643–659.

BURROUGH, P.A. 1981 Fractal dimensions of landscapes and other environmental data. *Nature (London)* **294,** 240–242.

BURROUGH, P.A. 1983*a*. Multiscale sources of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation. *Journal of Soil Science* **34,** 577–597.

BURROUGH, P.A. 1983*b*. Multiscale sources of spatial variation in soil. II. A non-Brownian fractal model and its application in soil survey. *Journal of Soil Science* **34,** 599–620.

BURROUGH, P.A. 1984. The application of fractal ideas to geographical phenomena. *Bulletin of the Institute of Mathematics and its Applications* **20,** 36–42.

CAMPBELL, J.B. 1978. Spatial variation of sand content and pH within single contiguous delineations of two soil mapping units. *Soil Science Society of America Journal* **42,** 460–464.

CHRISTAKOS, G. 1984. On the problem of permissible covariance and variogram models. *Water Resources Research* **20,** 251–265.

CRESSIE, N. 1985. Fitting variogram models by weighted least squares. *Mathematical Geology* **17,** 563–586.

CRESSIE, N. & HAWKINS, D.M. 1980. Robust estimation of the variogram. I. *Mathematical Geology* **12,** 115–125.

DALENIUS, T., HAJEK, J. & ZUBRZYCKI, S. 1961. On plane sampling and related geometrical problems. *Proceedings of the 4th Berkeley Symposium on Probability and Statistics* **1,** 125–150.

DAVID, M. 1977. *Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam.

DAVIS, B.M. & BORGMAN, L.E. 1979. Some exact sampling distributions for variogram estimators. *Mathematical Geology* **11,** 643–653.

DAVIS, B.M. & BORGMAN, L.E. 1982. A note on the asymptotic distribution of the sample variogram. *Mathematical Geology* **14,** 189–193.

EVANS, R. 1972. Air photographs for soil survey in lowland England: soil patterns. *Photogrammetric Record* **7,** 302–322.

FAIRFIELD SMITH, H. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science, Cambridge* **28,** 1–23.

GAJEM, Y.M., WARRICK, A.W. & MYERS, D.E. 1981. Spatial dependence of physical properties of a typic torrifluvent soil. *Soil Science Society of America Journal* **45,** 709–715.

HALLSWORTH, E.G., ROBERTSON, G.K. & GIBBONS, F.R. 1955. Studies in pedogenesis in New South Wales. VII. The 'gilgai' soils. *Journal of Soil Science* **6,** 1–31.

JOURNEL, A.G. & HUIJBREGTS, C.J. 1978. *Mining Geostatistics*. Academic Press, London.

KRIGE, D.G. 1951. A stastistical approach to some mine valuations and allied problems at the Witwatersrand. M.Sc. thesis, University of Witwatersrand.

LASLETT, G.M., MCBRATNEY, A.B., PAHL, P.J. & HUTCHINSON, M.F. 1986. Comparison of several spatial prediction methods for soil pH. *Journal of Soil Science* (in press).

MCBRATNEY, A.B. & WEBSTER, R. 1981*a*. Spatial dependence and classification of the soil along a transect in north-east Scotland. *Geoderma* **26,** 63–82.

MCBRATNEY, A.B. & WEBSTER, R. 1981*b*. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. II. Program and examples. *Computers and Geosciences* **7,** 335–365.

MCBRATNEY, A.B. & WEBSTER, R. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science* **34,** 137–162.

MCBRATNEY, A.B. & WEBSTER, R., MCLAREN, R.G. & SPIERS, R.B. 1982. Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland. *Agronomie* **2,** 969–982.

MACLUSKY, D.A. 1960. Some estimates of the areas of pasture fouled by excreta of dairy cows. *Journal of the British Grassland Society* **15,** 181–188.

MARDIA, K.V. 1980. Some statistical inference problems in kriging. II. Theory. *Sciences de la Terre, Série Informatique Géologique* **15,** 113–131.

MARDIA, K.V. & MARSHALL, R.J. 1984. Maximum likelihood models for residual covariance in spatial regression. *Biometrika* **71,** 135–146.

MANDELBROT, B.B. 1982. *The Fractal Geometry of Nature*. Freeman, San Francisco.

MATÉRN, B. 1960. Spatial variation. *Meddelanden från Statens Skogforskningsinstitut* **49,** 1–144.

MATHERON, G. 1955. Application des méthodes statistiques à l'estimation des gisements. *Annales des Mines*, Décembre, 50–75.

MATHERON, G. 1965. *Les variables régionalisées et leur estimation*. Masson, Paris.

MATHERON, G. 1971. The theory of regionalized variables and its applications. *Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, No. 5, Paris.

MOFFAT, A.J., CATT, J.A., WEBSTER, R. & BROWN, E.H. 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.

OLEA, R. A. 1975. *Optimum mapping techniques using regionalized variable theory*. Series on Spatial Analysis No. 2. Kansas Geological Survey, Lawrence.

OLIVER, M.A. & WEBSTER, R. 1986. Semi-variograms for modelling spatial pattern of geomorphic and soil properties. *Earth Surface Processes and Landforms* **11** (in press).

PETERSON, R.G., LUCAS, H.L. & WOODHOUSE, W.W. 1956*a*. The distribution of excreta of freely grazing cattle and its effect on pasture fertility. I. Excretal distribution. *Agronomy Journal* **48**, 440–444.

PETERSON, R.G., WOODHOUSE, W.W. & LUCAS, H.L. 1956*b*. The distribution of excreta by freely grazing cattle and its effect on pasture fertility. II. Effect of returned excreta and the residual concentration of some fertilizer elements. *Agronomy Journal* **48**, 444–449.

ROSS, G.J.S. 1980. *MLP Maximum Likelihood Program*. Rothamsted Experimental Station, Harpenden.

RUSSO, D. & BRESLER, E. 1982. Soil hydraulic properties as stochastic processes. II. Errors of estimates in a heterogenous field. *Soil Science Society of America Journal* **46**, 20–26.

SISSON, J.B. & WIERENGA, P.J. 1981. Spatial variability of steady-state infiltration rates as a stochastic process. *Soil Science Society of America Journal* **45**, 699–704.

TABOR, J.A., WARRICK, A.W., PENNINGTON, D.A. & MYERS, D.E. 1984. Spatial variability of nitrate in irrigated cotton: I. Petioles. *Soil Science Society of America Journal* **48**, 602–607.

VIEIRA, S.R., NIELSEN, D.R. & BIGGAR, J.W. 1981. Spatial variability of field-measured infiltration rate. *Soil Science Society of America Journal* **45**, 1040–1048.

WEBSTER, R. 1973. Automatic soil boundary location from transect data. *Mathematical Geology* **5**, 27–37.

WEBSTER, R. 1977. Spectral analysis of gilgai soil. *Australian Journal of Soil Research* **15**, 191–204.

WEBSTER, R. 1985. Quantitative spatial analysis of soil in the field. *Advances in Soil Science* **3**, 1–70.

WEBSTER, R. & BURGESS, T.M. 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. *Journal of Soil Science* **31**, 505–524.

WEBSTER, R. & BURGESS, T.M. 1984. Sampling and bulking strategies for estimating soil properties of small regions. *Journal of Soil Science* **35**, 127–140.

WEBSTER, R. & NORTCLIFF, S. 1984. Improved estimation of micronutrients in hectare plots of the Sonning series. *Journal of Soil Science* **35**, 667–672.

WHITTLE, P. 1954. On stationary processes in the plane. *Biometrika* **41**, 434–449.

WHITTLE, P. 1963. Stochastic processes in several dimensions. *Bulletin of the 34th Session of the International Statistical Institute, The Hague*, pp. 974–985.

WONG, I.F.T. 1967. A numerical test of soil boundaries interpreted on aerial photographs. Diploma Dissertation, Oxford University.

YOST, R.S., UEHARA, G. & FOX, R.L. 1982*a*. Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. *Soil Science Society of America Journal* **46**, 1028–1032.

YOST, R.S., UEHARA, G. & FOX, R.L. 1982*b*. Geostatistical analysis of soil chemical properties of large land areas. II. Kriging. *Soil Science Society of America Journal* **46**, 1033–1037.

YOUDEN, W.J. & MEHLICH, A. 1937. Selection of efficient methods for soil sampling. *Contributions of the Boyce Thompson Institute for Plant Research* **9**, 59–70.

ZUBRZYCKI, S. 1957. On estimating gangue parameters. *Zastosowania Matemakyi* **3**, 105–153 (in Polish).

(Received 15 July 1985)