INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY. CABLE CENTRATOM TRIESTE

SMR.705 - 26

COLLEGE ON SOIL PHYSICS

(6 - 24 September 1993)

"Quantitative Spatial Analysis of Soil in the Field"

R. Webster
Rothamsted Experimental Station
Harpenden, Hertfordshire AL5 2JQ
England

Quantitative Spatial Analysis of Soil in the Field

R. Webster*

*Rothamsted Experimental Station, Harpenden, Hertfordshire, AL5 2JQ, England

Advances in Soil Science, Volume 3

I. Introduction

Soil scientists have recognized variation in soil from place to place for many years. They have portrayed the variation by dividing large regions into smaller parcels each of which is relatively homogeneous, and they have classified the soil to show similarities between soil in widely separated parcels. This procedure, which may be regarded as standard soil survey practice, requires appreciation of the scale of change, the abruptness or otherwise of change, the degree of correlation among different soil properties, and of relations in the landscape. Yet that appreciation has almost always been intuitive. Good soil surveyors have needed flair. Rarely have they gained their appreciation by quantitative analysis.

The standard procedure has undoubtedly been successful. It has also had its disappointments; for example, in small regions where there are no obvious boundaries and where the properties of interest are uncorrelated with visible change. Now, because soil scientists increasingly require quantitative estimates of soil properties for regions of varying sizes and wish to plan their surveys in the most economical way, standard procedure must be augmented by more rational and quantitative methods.

In other contexts, such as advisory work, field experimentation, and irrigation planning, soil scientists have been very concerned with sampling efficiency and the measurements of variation in soil. Beckett and Webster (1971) reviewed the extensive literature on this and the numerous attempts to plan sampling strategies. At that time, the only way to increase efficiency and the precision of estimates seemed to be to stratify the soil first and then to use standard sampling theory to determine the estimation variances. This was a simple combination of classical statistics and soil classification.

The limitations of this classical approach were clear, and some alternative way of representing variation was obviously desirable. Trend surface analysis—that is, a form of multiple regression with the spatial coordinates as independent variables—enjoyed a brief spell of popularity in geology, especially for oil exploration (Harbaugh and Merriam, 1968). There were a few attempts to apply it to soil. Walker *et al.* (1968) had some success using polynomials over distances of up to 100 m on some segments of landscape but not others. In general, however, there was no obvious functional relation between geographic position and soil and no theory from which to

predict one. Soil properties behaved much more like random variables. Time series methods seemed much more promising, therefore, and Webster (1973, 1977), Webster and Cuanalo (1975), and Kozlovskii and Sorokina (1976) began to explore their value and to adapt them to the spatial distribution of soil, as had Matérn (1960) to forestry and Gandin (1965) to meteorology. As it happened, they were overtaken by events.

Matheron (1965), working in the mining context, had brought together a number of isolated results in spatial statistics into a coherent body of theory, the theory of regionalized variables. His thesis in French was followed by an English text (Matheron, 1971), but because of the highly mathematical treatment and unfamiliar setting it was still some time before earth scientists, and soil scientists in particular, saw in it what they had been seeking.

This work must be regarded as a breakthrough for soil science. Regionalized variable theory now provides the basis for describing spatial variation in soil quantitatively, for estimating soil properties and mapping them soundly, and for planning rational sampling schemes that make the best use of manpower. This review shows how each of these is accomplished using the theory with examples from the recent literature. It begins, however, with a description of a much earlier approach to the problem attributable to Youden and Mehlich (1937), one that deserved much more attention than it received and that still has a proper place in a survey scheme.

II. Nested Sampling and Analysis

Nested or hierarchical sampling schemes are commonly used in survey research to distinguish variation deriving from two or more levels of subdivision of a population. In the present context, the soil of a region might be divided into districts, farms, and fields, and in a survey one might wish to measure the variation contributed by each to the total variation in some property over the whole region. By sampling the soil within fields and identifying the farm and district to which each sampled field belongs, the variance can be partitioned by a hierarchical analysis and components estimated. Youden and Mehlich (1937) adapted this technique to measure the variance associated with different spatial scales in Broome County, New York, as follows. On each of two soil series they chose nine primary stations (stage 1) approximately 1.6 km apart. At each station they chose two substations (stage 2) 305 m apart, and at each of these they chose two sampling areas (stage 3) 30.5 m apart, in each of which two sampling points (stage 4) were located 3.05 m apart. They thus had a complete balanced design, which ensured that the analysis was straightforward. Note also the geometric progression of the spacings. In one series, the Culvers

series, a coarse-loamy, mixed, mesic Typic Fragidcherpt, a sample of topsoil (0 to 15 cm) was taken at each sampling point and its pH measured. In the other series, Sassafras, a fine-loamy, siliceous, mesic Typic Hapludult, the pH of the A and B horizons was measured on samples from each point. Youden and Mehlich then performed an analysis of variance on each set of measurements, and their paper presents the results in the usual tabular form.

The model of variation is

$$Z_{ijkl} = \mu + A_i + B_{ij} + C_{ijk} + \varepsilon_{ijkl}, \qquad [1]$$

where z_{ijkl} is the pH at point l in the kth sampling area in the jth substation in station i; μ is the mean of z in the region; A_i, B_{ij}, and C_{ijk} are random variables associated with stages 1, 2, and 3 with variances σ_1^2, σ_2^2, and σ_3^2, respectively; and ε_{ijkl} is a residual term with variance σ_4^2. Table 1 shows the general form of the analysis of variance for a four-stage balanced design with the contributions to the degrees of freedom and mean squares. In this study $n_1 = 9$ and $n_2 = n_3 = n_4 = 2$, so it is a simple matter to compute the estimated components of variance σ_1^2, σ_2^2, σ_3^2, and σ_4^2. Table 2 gives these.

The results have several features. In the Culvers series, the largest component of variance, accounting for almost 40% of the total, derives from the largest spacing, 1.6 km. The variance in stage 4 from points only 3.05 m apart accounts for 20% and is by no means the smallest. Turning to the Sassafras series, we see that the 3.05 m spacing contributes even more variance, 30.2% and 41.5% in the A and B horizons, respectively. In the A horizon, the 1.6 km spacing contributes nothing to the variance, most of which derives from spacings between 30 and 300 m.

The results are best appreciated by accumulating the components and plotting them against sample spacing, as in Figure 1. The variance in all three increases with increase in sample spacing, but there the similarity

Table 1. Degrees of Freedom and Mean Squares for a Four-Stage Balanced Hierarchical Analysis of Variance

Source (stage)[a]	Degrees of freedom	Components of mean square
1	$n_1 - 1$	$\sigma_4^2 + n_4\,\sigma_3^2 + n_3\,n_4\,\sigma_2^2 + n_1\,n_2\,n_3\,\sigma_1^2$
2	$n_1(n_2 - 1)$	$\sigma_4^2 + n_4\,\sigma_3^2 + n_3\,n_4\,\sigma_2^2$
3	$n_1\,n_2(n_3 - 1)$	$\sigma_4^2 + n_4\,\sigma_3^2$
4	$n_1\,n_2\,n_3(n_4 - 1)$	σ_4^2
Total	$n_1\,n_2\,n_3\,n_4 - 1$	

[a]For each stage, g, n_g is the number of subdivisions within stage $g - 1$ and σ_g^2 is its component of variance.

Table 2. Components of Variance of pH in Two Soil Series in Broome County, New York

Stage	Spacing (m)	Degrees of freedom	Culvers series 0–15 cm		Sassafras series A horizon		Sassafras series B horizon	
			Estimated component	Percentage of variance	Estimated component	Percentage of variance	Estimated component	Percentage of variance
1	1600	8	0.02819	39.7	0	0	0.00569	16.7
2	305	9	0.02340	32.9	0.04440	60.3	0.00366	10.8
3	30.5	18	0.00552	7.8	0.00698	9.5	0.01055	31.0
4	3.05	36	0.01391	19.6	0.02225	30.2	0.01409	41.5

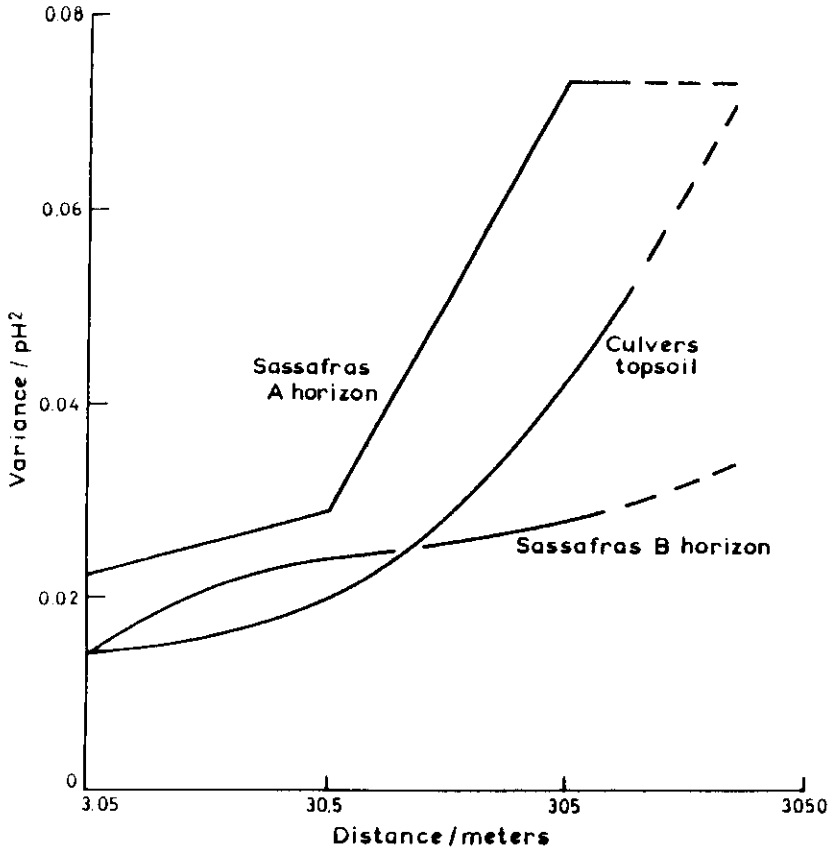

Figure 1. Graphs of accumulated variance of pH against separating distance of Broome County, New York.

ends. In the Culvers series, the variance appears to be increasing without limit, whereas in the A horizon of the Sassafras series, the variance has reached a maximum, which we now know as a *sill*. At the other end of the scale, two of the graphs appear to be leveling out to appreciable finite values, not zero, as might be expected. This we also know now to be quite general and recognize as the "nugget effect," which will be discussed in more detail later.

Two further examples illustrate the power of this technique. The first is from a study of soil variation on the 400 ha Ginninderra Experiment Station in the Australian Capital Territory (Webster and Butler, 1976). Again four sampling stages were chosen, eight primary stations at 180 m intervals with twofold subdivision at 50 m, 18 m, and 5 m, to give 64

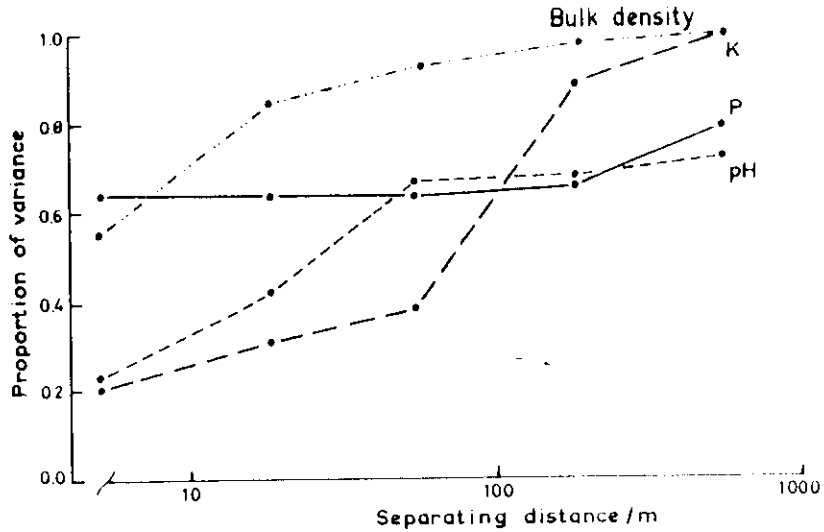

Figure 2. Graphs of accumulated variance of form topsoil properties against distance at Ginninderra, Australian Capital Territory.

sampling points at which several soil properties were measured on 10-cm diameter cores. Data for larger spacings were available, and these have been added in the analysis. The estimated components for four properties of the topsoil are accumulated and plotted against spacing in Figure 2. The most striking feature of these results is the difference in pattern between one soil property and another. For example, most of the variance in potassium content is contributed from spacings between 50 and 180 m, whereas for pH it derives from distances between 5 and 50 m. For bulk density most variance is present within 18 m, and for phosphorus it lies mainly within 5 m. The region was one for which it had been difficult to make a generally useful soil map, and these results show why.

The second example is from a study by Nortcliff (1978) in Norfolk, England. The sampling scheme has a less strict scale of distance. It was deliberately designed to distinguish major geological formations in the first stage, since it would have been foolish to have ignored such obvious features. Units in the second and third stages were chosen randomly within squares of 500 m and 100 m sides, respectively, rather than at constant spacings. The fourth and fifth stages were at 20 m and 5 m spacings. Nortcliff was especially interested in a strategy for general-purpose survey. He therefore recorded all the more obvious soil properties and transformed them to principal components before performing analysis of variance. Figure 3 presents the accumulated variances for the first six principal components. It distinguishes clearly the long-range variation of the leading

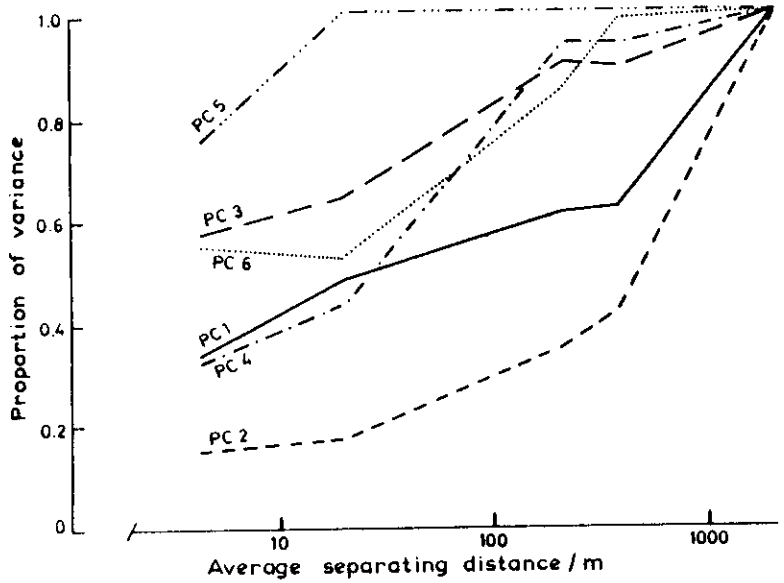

Figure 3. Graphs of accumulated variance of the six leading principal components of soil in Norfolk against distance. From Nortcliff (1978) with permission.

components from the dominantly short-range variation in the higher order ones. The correlated information seems to relate to the major geological differences and to be mappable at conventional medium mapping scales—1:25,000 or 1:50,000—whereas the uncorrelated information is not.

III. Regionalized Variable Theory

In the nested analysis, variation is assumed to comprise a number of independent components, one for each spacing or stage of the hierarchy. If the spacings chosen increase in geometrical progression, this assumption is reasonable. Regionalized variable theory, however, takes a different approach. It considers differences between pairs of values of a property at places separated by any distance and expresses these as their variances. It also takes into account direction. Suppose we have the values $z(\mathbf{x})$ and $z(\mathbf{x} + \mathbf{h})$ at $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$, respectively, where $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ are positions with one, two, or three spatial coordinates and $\mathbf{h}$ is a vector with both distance and direction, usually known as the *lag*, separating them. Then for this pair the variance per site is

$$s^2 = [z(\mathbf{x}) - \bar{z}]^2 + [z(\mathbf{x} + \mathbf{h}) - \bar{z}]^2, \qquad [2]$$

where $\bar{z}$ is the mean of the two values. Notice that s^2 is half the square of the difference:

$$s^2 = \tfrac{1}{2}[z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})]^2. \qquad [3]$$

Regionalized variable theory focused attention first on such differences and their variances. The quantity s^2 was therefore called the *semi-variance*, and the name has stuck. Nevertheless, it is the variance per site or observation. If further we have, say, m pairs of observations separated by the same lag, $\mathbf{h}$, then we can define their average

$$\bar{s}^2 = \frac{1}{2m} \sum_{i=1}^{m} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2. \qquad [4]$$

This is effectively what Youden and Mehlich did in computing the residual variance at 3.05-m spacing.

To make use of this simple notion and generalize equation [4], certain stationarity assumptions must be made. These are as follows.

1. The expected value of z at any place $\mathbf{x}$ is the mean, μ:

$$E[z(\mathbf{x})] = \mu. \qquad [5]$$

2. For any $\mathbf{h}$ the difference $[z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})]$ has a finite variance, which again is independent of $\mathbf{x}$:

$$\begin{aligned} \text{var}\,[z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})] &= E\{[z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})]^2\} \\ &= 2\,\gamma(\mathbf{h}). \end{aligned} \qquad [6]$$

These two assumptions constitute the *intrinsic hypothesis* of regionalized variable theory. They assume the following model of soil variation:

$$z(\mathbf{x}) = \mu_v + \varepsilon(\mathbf{x}), \qquad [7]$$

where $z(\mathbf{x})$ is the value of the property at position $\mathbf{x}$ within a region, μ_v is the mean value in that region, and $\varepsilon(\mathbf{x})$ is a spatially dependent random component with zero mean and variance defined by

$$\text{var}\,[\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x} + \mathbf{h})] = E\{[\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x} + \mathbf{h})]^2\} = 2\,\gamma(\mathbf{h}). \qquad [8]$$

In a large region, of course, we know that a soil property will vary from one part to another. Nevertheless, the property will commonly be locally stationary within some neighborhood V, and this condition is usually quite adequate for analysis in which $\mathbf{h}$ is limited to some maximum radius r within which the relationships apply. It is for this reason that the subscript V is used in equation [7].

Where the intrinsic hypothesis holds, we can expect the same degree of difference in the soil property at any two places $\mathbf{h}$ apart, whatever the actual values of the property are. In these circumstances, the sample value

s^2 in equation [4] is an unbiased estimate of the average semi-variance, $\gamma(\mathbf{h})$, in the population.

A. Relation to Time Series Analysis

There is a close analogy between space and time, at least for one-dimensional space, and the methods of analysis are in large measure interchangeable. Traditionally, time series analysis has built on the auto-covariance and autocorrelation. The auto-covariance of a property that varies in time is defined as

$$\begin{aligned} C(h) &= \mathrm{E}\{[z(x) - \mu]\,[z(x + h) - \mu]\} \\ &= \mathrm{E}[z(x) \cdot z(x + h)] - \mu^2, \end{aligned} \qquad [9]$$

where x and $x + h$ are points in time separated by a time lag h. The notation is changed slightly since x and h are now single valued in the one dimension. When $h = 0$, expression [9] defines the variance, i.e.,

$$\begin{aligned} C(0) &= \mathrm{E}[z^2(\mathbf{x})] - \mu^2 \\ &= \sigma^2. \end{aligned} \qquad [10]$$

The ratio $C(h)/C(0)$ is the autocorrelation, denoted by $\rho(h)$.

By combining equations [6], [7], and [8] the semi-variance is seen to be simply related to the auto-covariance and autocorrelation by

$$\begin{aligned} \gamma(h) &= C(0) - C(h) \\ &= \sigma^2\,[1 - \rho\,(h)]. \end{aligned} \qquad [11]$$

They are complementary. The more similar are values at lag h, the smaller is the semi-variance and the larger are the auto-covariance and autocorrelation. The latter must lie between 1 and -1.

To make use of this approach requires the assumption of full second-order stationarity. That is, both the mean and variance must be constant. It frequently happens in earth science that spatially distributed variables appear to have no finite *a priori* variance nor covariance. In these circumstances, the semi-variogram can exist without there being a corresponding covariance, and it is for this reason that the semi-variance rather than the covariance is used predominantly in spatial analysis.

B. The Semi-Variogram and Its Estimation

Equations [6] and [8] define the semi-variance as a function of $\mathbf{h}$, the lag. This function is the *semi-variogram*, $\gamma(\mathbf{h})$. In one dimension, $\gamma(h)$ can be estimated at regular intervals by sampling along transects. Thus, given a set of values $z(x_1), z(x_2), \ldots z(x_n)$ we can estimate $\gamma(h)$, where h is any integral multiple of the sampling interval, by

$$\hat{\gamma}(h) = \frac{1}{2(n-h)} \sum_{i=1}^{n-h} [z(x_i) - z(x_i + h)]^2. \qquad [12]$$

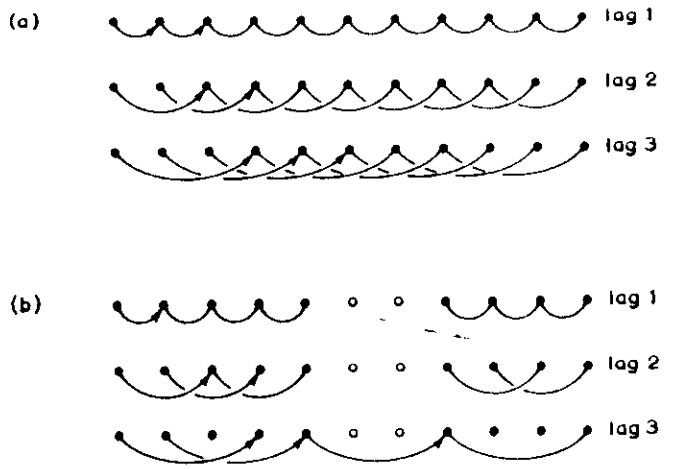

Figure 4. Comparisons for estimating semi-variances on linear transects at lags of 1, 2, and 3 sampling intervals, (a) for complete data and (b) where some observations are missing the open circles.

Figure 4a shows the comparisons involved for $h = 1$, 2, and 3. The result is an ordered set of values that constitute the *sample semi-variogram*. These can readily be plotted, and there are numerous examples now in the literature of soil science. Some are shown in Figures 6, 12, 13, 20, and 27.

This simple formula for estimating semi-variance is sensitive to extreme values of the differences $z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})$, especially as the squared difference follows a chi-squared distribution with one degree of freedom, which is highly skewed. Cressie and Hawkins (1980) investigated more robust estimators of $\gamma(\mathbf{h})$ and discovered that the fourth root of the usual squared difference,

$$y\,(\mathbf{x}) = \{|z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})|^2\}^{1/4}, \qquad [13]$$

had a distribution close to normal with negligible skew. They therefore used this formula to compute a mean $\bar{y}$ of m separate $y(\mathbf{x})$. This value must, of course, be transformed back. Cressie and Hawkins showed that the expectation

$$\mathrm{E}[\bar{y}^4/2\,\gamma(\mathbf{h})] = 0.457 + \frac{0.494}{m} + \frac{0.045}{m^2}, \qquad [14]$$

and so the required back transformation gives the estimated semi-variance as

$$\hat{\gamma}(\mathbf{h}) = \bar{y}^4 / 2(0.457 + 0.494\,m^{-1} + 0.045\,m^{-2}). \qquad [15]$$

This is a recent development in variogram estimation, and there are no accounts yet of its being used for soil. Soil scientists should be aware of it, especially when analyzing data that are themselves skewed.

C. Practical Problems

Formula [12] for the sample semi-variance assumes that there is a value for every sample point along the transect. More often than not there are deficiencies in the records that preclude such a tidy computation. The most common is that observations are lacking where the transect crosses roads, rivers, bare rock, and the like. The situation is then as illustrated in Figure 4b, and the summation is made over the actual number of pairs, $m < n - h$, that can be compared at lag h. In other instances several transects are surveyed because it is more convenient or because they cover the region more evenly than does just one. The semi-variance can again be computed by pooling the individual sums for each lag over all the transects and dividing by total m comparisons.

A more serious difficulty arises where transects have been sampled at irregular intervals. This can be overcome by choosing a set of lags, h_i, $i = 1, 2, \ldots$ at arbitrary but constant intervals, d, and assigning to each a class with limits $(i - 1)d$ and id and $h_i = (i - \frac{1}{2})d$ as its mid-point. Every pair of observations that are separated by the lag $h_i \pm d/2$ are then used to estimate $\gamma(h_i)$, and each pair contributes to one and only one estimate.

The effect of this procedure is to smooth the semi-variogram: The larger the increment the smoother will be the result. Some care is therefore needed to judge a suitable increment. If it is small there might be too few pairs of observations contributing to each semi-variance, so that the estimates are very imprecise. If, on the other hand, the increment is large then information can be lost by unnecessary smoothing. The best compromise will depend on the amount of data, the evenness of coverage, and the form of the underlying semi-variogram. Better still is to sample at regular intervals initially. The soil is accessible at most places, and there is rarely any good reason for not sampling systematically.

The confidence limits of semi-variances calculated in this way, whether from regular or irregular sampling, are uncertain. For a single transect, however, good practice is to have at least 100 sampling points and to estimate $\gamma(h)$ for lags up to no more than about one-fifth of the total run. This is a conservative guide; some workers are prepared to extend the lag to a third of the total length. If several transects are sampled then they should be long enough to provide at least 80 pairs for each estimate.

D. Two Dimensions

In survey we are usually interested in the variation in a plane rather than in a single direction along a transect. The semi-variogram is then a two-dimensional function. A sample semi-variogram can be calculated quite straightforwardly if we have measured the soil at regular intervals on a two-dimensional grid. Suppose the grid has m rows and n columns. We estimate semi-variances as follows:

$$\hat{\gamma}(p, q) = \frac{1}{2(m-p)(n-q)} \sum_{i=1}^{m-p} \sum_{j=1}^{n-q} [z(i,j) - z(i + p, j + q)]^2 \qquad [16]$$

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

where p and q are the lags in the two dimensions. These equations enable half of the semi-variogram to be computed for lags from $-q$ to q and from 0 to p. The semi-variogram is symmetrical about its center, so that if the full set of semi-variances is needed the remainder are readily obtained as

$$\hat{\gamma}(-p, q) = \hat{\gamma}(p, -q)$$

and

$$\hat{\gamma}(-p, -q) = \hat{\gamma}(p, q).$$

As in the one-dimensional case there will often be missing data, and, perhaps more likely, the region of interest will have an irregular shape. So again the quantity $(m - p)(n - q)$ in the denominators of equations [16] must be replaced by the actual number of paired comparisons in each sum.

The least tidy situation occurs where data are irregularly scattered in two dimensions. Every pair of observations is then separated by potentially different distances and directions. This difficulty is usually overcome by grouping the separations both in distance and direction. A series of lag distances and directions is chosen, usually to form regular progressions as in the one-dimensional case described above. A range in each is chosen, again usually equal to the class interval between successive lags, and applied so that the nominal lag lies at the center of the range. Each squared difference then contributes to the semi-variance for the lag class into which it falls by virtue of its actual separation. Figure 5 shows the geometry of the grouping. The nominal lag is represented by the line OL of length l and the direction θ. The range in distance is w and that in direction α.

1. Representing Two-Dimensional Semi-Variograms

Sample semi-variograms in one dimension are readily drawn as in Figures 12, 20, and 27. If desired, the plotted points may be joined by straight lines, since the functions are continuous.

Displaying two-dimensional semi-variograms is not so straightforward since there are three dimensions involved: the two spatial dimensions plus that of the semi-variance. One way is to present the variogram as an

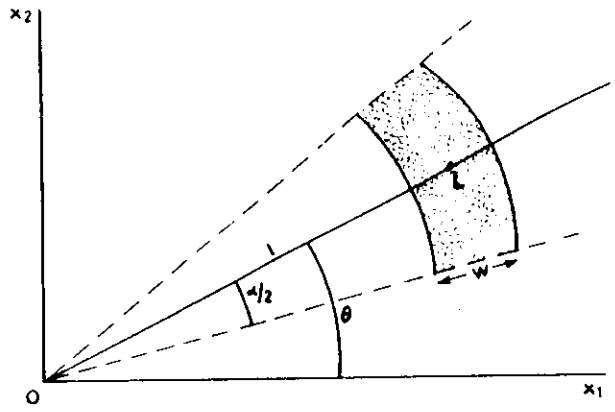

Figure 5. Grouping of lags by distance and direction. The striped area shows the extent of one group (see text).

isarithmic ("contour") map. Alternatively a perspective block diagram can be made of the interpolated surface. With a little more ingenuity stereograms can be drawn. As we shall see later, however, the forms of semi-variogram in two dimensions are usually simple geometric extensions of one-dimensional forms, and a cylindrical projection of the sample values will often serve quite well.

To illustrate these possibilities we take as an example the study made by Burgess and Webster (1980a, 1980b) of the soil in Cae Ruel, one of the field at the Welsh Plant Breeding Station, Plas Gogerddan. The topsoil, 0 to 15 cm, had been sampled at 15.2 m intervals on a square grid. Several properties, including the stone content, were measured on bulked samples of 10 cores of 2.5-cm diameter chosen randomly within the 15.2 × 15.2 m squares around each grid node. There were approximately 450 observations in all.

In the original analysis semi-variograms were computed for four directions only: along the rows and columns of the grid and parallel to the two diagonals. Figure 6 shows the result. Using equations [16], however, a full two-dimensional semi-variogram can be computed, and as above, this can then be represented as an isarithmic map, Figure 7, block diagram, Figure 8, and cylindrical projection, Figure 9. The last shows a spread of points that expands away from the origin in a way similar to that in Figure 6. The semi-variances have been grouped into classes according to their angular separation. Each class, of which there are eight between 0 and 180 degrees, subtends 22.5 degrees, and is given a unique symbol. The significance of the oblique lines enveloping the points will be dealt with later.

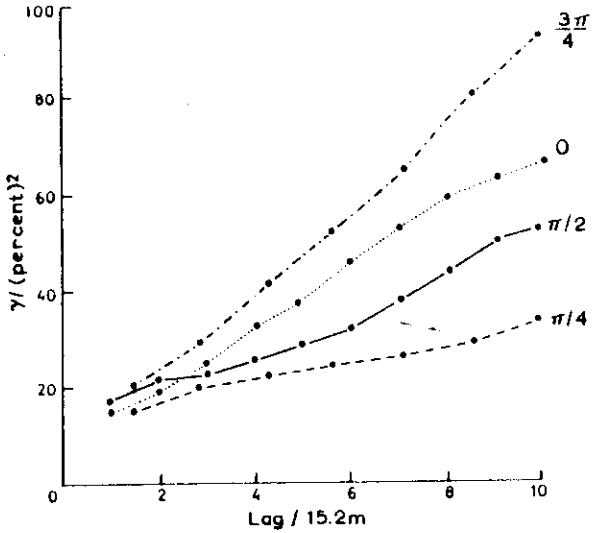

Figure 6. Sample semi-variograms of stone content in the topsoil at Plas Gogerddan in four principal directions.

2. Anisotropy

We note here that the semi-variogram of the stone content is not the same in all directions. Its gradient varies substantially from about 2 in direction $\pi/4$ to about 8 in the perpendicular direction. Thus, the stone content in this field varies anisotropically. This is not an unusual result. There are many situations in which anisotropic variation is recognized in soil survey. For example, where the land surface bevels a sequence of sediments the soil at points along the strike is more likely to be similar than that the same distance away in the direction of dip. On a point bar deposit or river levee we usually observe greater similarity in the soil if we travel parallel to the river than if we move at right angles to it. More generally it is common experience that variation is encountered more often when crossing the drainage lines of a region than when traversing along the contours. Thus an analysis of two-dimensional variation must allow for anisotropy.

E. Size of Support

In material such as soil in which there is spatial dependence, the amount of variation present within a single core or block depends on the size of the block: The bigger its area or volume the more variation it embraces, and the less there is outside it. This affects the observed semi-variogram, and so a semi-variogram depends on the size, shape, and orientation of the individual samples on which measurements are made. Size, shape, and

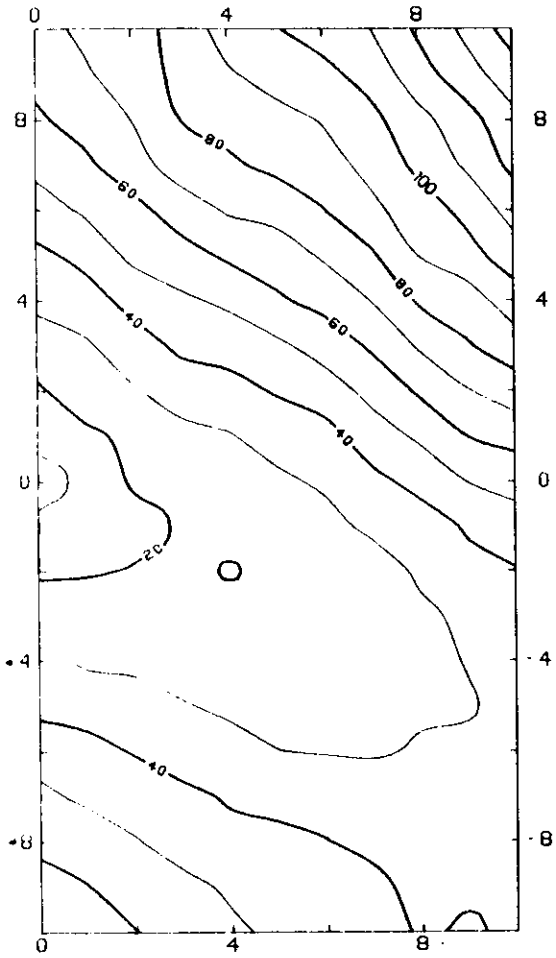

Figure 7. Isarithmic representation and sample semi-variogram of stone content at Plas Gogerddan. Border scales are in sampling intervals.

orientation constitute the *support* of the data. In general, measurements made on core samples encompass less variation than those made on larger volumes from pits. They in turn contain less than those made on bulked samples. Indeed, the aim of bulking is to diminish the variation between measurements by encompassing more within them. This increase in the support is known in geostatistics as *regularization*. Analysts should therefore realize that their results refer specifically to the particular support on which the measurements were made, and they should state the support when reporting their results.

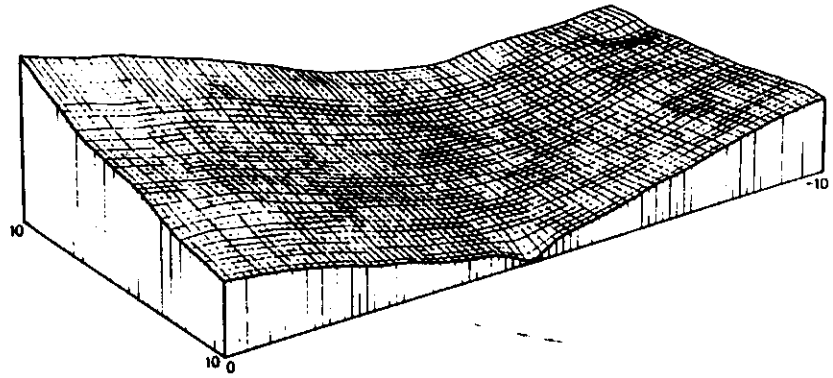

Figure 8. Perspective representation of sample semi-variogram of stone content at Plas Gogerddan viewed from a position above and to the left in Figure 7.

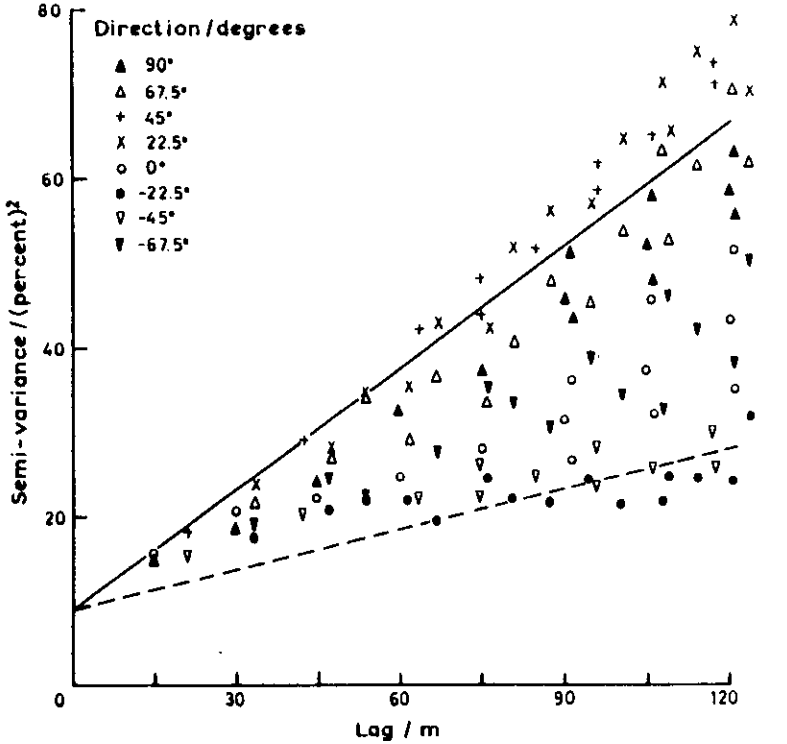

Figure 9. Cylindrical projection of two-dimensional sample semi-variogram of stone content with separate symbols for eight directions. The oblique lines form the envelope of the fitted linear model (see text).

The semi-variogram for one support can be related, at least theoretically, to that for another. In particular, there is often a practical interest in regularized semi-variograms. Suppose that $\gamma(\mathbf{h})$ is the semi-variogram for the support on which measurements were made. We shall consider this to be the punctual semi-variogram. Suppose also that we wish to determine a regularized semi-variogram, $\gamma_B(\mathbf{h})$ for larger blocks of size B. It can be shown that for a given lag $\mathbf{h}$,

$$\gamma_B(\mathbf{h}) = \bar{\gamma}(B, B_\mathbf{h}) - \bar{\gamma}(B, B), \qquad [17]$$

where $\bar{\gamma}(B, B_\mathbf{h})$ denotes the average semi-variance between two blocks of size B separated by the vector $\mathbf{h}$, and $\bar{\gamma}(B, B)$ is the average semi-variance within a block B; i.e., the within-block variance. If the distance $\mathbf{h}$ is much larger than the distance across the block B, then $\bar{\gamma}(B, B_\mathbf{h})$ approximately equals $\gamma(\mathbf{h})$, the point semi-variance. From equation [17] we obtain the useful approximation

$$\gamma_B(\mathbf{h}) \simeq \gamma(\mathbf{h}) - \bar{\gamma}(B, B). \qquad [18]$$

Thus for $|\mathbf{h}| >> \sqrt{B}$ the regularized semi-variogram is derived from the punctual semi-variogram simply by subtracting the within-block variance.

This is especially pertinent for bulking. If the semi-variogram of a soil property measured on separate cores is known, then that for samples bulked over larger areas, the regularized semi-variogram, can be determined readily from it.

IV. Semi-Variogram Models

Soil varies continuously in space, at least at most practical scales, and so semi-variograms of soil properties are continuous functions. The sample semi-variograms, however, consist simply of ordered sets of discrete values. These are estimates and as such are subject to error. They can be joined by straight lines or curves to give intermediate values, but the result is inevitably irregular. Nevertheless, a quick look at most well-estimated semi-variograms of soil will show that generally they approximate simple forms, and that it should be possible to fit simple functions to them. Later we shall wish to use the semi-variogram for estimation, and so its form is of both practical and scientific interest.

Figure 10, for one dimension, shows the principal features of semi-variograms of soil. In most instances it is found that $\gamma(h)$ increases from the smallest measured lag. In Figures 6 to 9 this increase in $\gamma(h)$ for stone content appears to have no limit: The soil appears to have no finite variance. This is represented by the solid line in Figure 10a. More often the semi-variance reaches a maximum at which it levels out, and such semi-variograms are said to be *transitive*. This maximum is known as the *sill*,

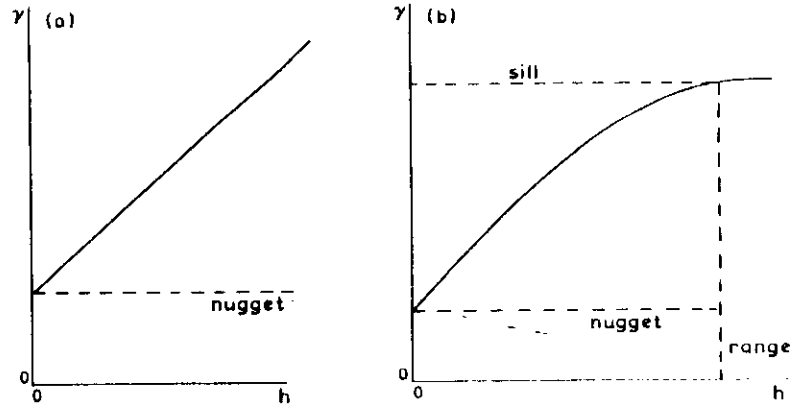

Figure 10. Elements of one-dimensional semi-variogram.

and its value is the *a priori* variance of the variable. The lag at which the sill is reached is known as the *range*. If $\gamma(h)$ approaches its maximum asymptotically, then for practical purposes the range may be chosen arbitrarily where $\gamma(h)$ is sufficiently close to its apparent sill.

The range is of considerable importance. It marks the limit of spatial dependence. The soil at places closer together than the range is related; at places further apart it is not, unless there is periodicity. The range may be interpreted as indicating the average distance across distinct soil types in some instances (Webster, 1973; Webster and Cuanalo, 1975; McBratney and Webster, 1981a). Experience to date suggests that at the common scales for detailed soil survey the range will usually be a few hundred meters. It does, however, depend on the size of the area sampled, and if the scale of the survey is changed substantially then the range is likely to do so too. The range also represents the limiting distance within which interpolation is worth attempting, and this will be discussed later.

By definition the semi-variance at zero lag is itself zero. But as can be seen from Figures 6, 12, 13, and 15, smooth curves approximating the sample semi-variances in these examples are unlikely to pass through the origin. All appear instead to approach positive finite intercepts on the ordinate at $h = 0$. This intercept is known as the *nugget variance*, and the phenomenon, which is widely recognized, is known as the *nugget effect*. The terms derive from gold mining. In gold-bearing rocks and sediments nuggets are sparse and small. Their diameters are very much smaller, by at least one order of magnitude, than the spacing between sampling cores. Most cores contain none; a small proportion contain one each, rarely more. Thus, the inclusion of a gold nugget in a drill core is regarded as a purely random event.

Few soil properties of common interest are quite like this, yet in practice

their semi-variograms usually have distinct nugget variances. Indeed, in some investigations all the variation appears as nugget even when sampling as closely as 10 or 20 m. Campbell's (1978) results for the particle size fractions on both loess and till in Kansas were of this kind, as were several properties measured by McBratney and Webster (1981a) in northeast Scotland, e.g., Figure 12. Such results almost invariably mean that there is a source of spatial variation with a range much smaller than the smallest sampling interval. The shape of the semi-variogram in this range can be resolved only by much denser sampling.

A. Positive Definite Functions

Thus we see that in choosing a model to represent a sample semi-variogram we must allow for at least three elements in most instances: an intercept, an increasing section of potentially varying shape, and a sill. In two dimensions there must also be provision for anisotropy.

Not any model that appears to fit the observed values will serve, however, for the following reason. Suppose that $Z(\mathbf{x})$ 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 covariance function is $C(\mathbf{h})$ and $\gamma(\mathbf{h})$ its semi-variogram. And consider the linear combination

$$Y = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i), \qquad [19]$$

where λ_i, $i = 1, 2, \ldots, n$, are any arbitrary weights. This quantity is a random variable with

$$\text{var}\,[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C(\mathbf{x}_i, \mathbf{x}_j). \qquad [20]$$

The variance of Y may be positive or zero, but it may not be negative, and the covariance function on the right-hand side of equation [20] must ensure that this condition is met. The covariance, $C(\mathbf{h})$, must be a positive definite function, and only functions that meet this criterion are acceptable.

As we have seen, there are situations where soil properties do not have definable covariances because their variances increase apparently without limit. Provided the intrinsic hypothesis holds, however, we can make use of the following relation. Equation [20] can be rewritten as

$$\text{var}\,[Y] = C(0) \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} \lambda_j - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i, \mathbf{x}_j). \qquad [21]$$

The first term on the right-hand side of this equation can be eliminated provided

$$\sum_{i=1}^{n} \lambda_i = 0, \qquad [22]$$

giving

$$\text{var}\,[Y] = -\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j). \qquad [23]$$

This too must be non-negative, and we are left with the constraint on the semi-variogram that it must be a positive definite function with the added condition that the weights in equation [23] sum to 0.

Although we may state the above, it is not easy to test whether any particular model is conditional positive definite, and it can be difficult to create "data" that will demonstrate that a model is inappropriate (Armstrong and Jabin, 1981). The standard approach involves examining the Fourier transform of the semi-variogram or covariogram, and Christakos (1984) has listed the conditions that the spectrum must meet for a model to be acceptable. Dunn (1983) has shown that a simpler test can be applied if a model is required only for lags at which there are observed data. This, however, will be too restrictive in many instances.

The consequences of fitting a function that is not positive definite are not always disastrous. The investigator does not necessarily encounter negative variances, and so remains unaware of the dangers. The need to avoid such models is not always appreciated, especially because a function that is positive definite in one dimension is not necessarily so in two or three dimensions.

B. Safe Models

This section defines models that can be recommended for semi-variograms of soil properties. They are defined for one dimension but are safe in the sense that they are conditional positive definite in two and three dimensions. In the French literature they are often referred to as *authorized models*.

1. Linear Models

The simplest model that can be fitted in one dimension is clearly linear. It has slope w and may have an intercept or nugget variance c_0. Its formula is

$$\begin{aligned} \gamma(h) &= c_0 + wh \qquad \text{for } h > 0 \\ \gamma(0) &= 0. \end{aligned} \qquad [24]$$

Notice that it has no sill. Figure 11 shows an example. A linear model is also fitted to the first 15 sample semi-variances of pH in Figure 13.

In the limit w can be zero (Figure 12). The semi-variogram is then said

to show a pure nugget effect. There is no spatial dependence at the scale of investigation since all of the variance occurs within the smallest sampling interval.

2. Spherical Models

A model that has been found to fit not only many semi-variograms of soil properties but also those of mineral deposits of many kinds is the spherical model. Its definition is

$$\gamma(h) = c_0 + c\left[\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right] \qquad \text{for } 0 < h \leq a$$

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

$$\gamma(0) = 0. \qquad [25]$$

Its characteristics are illustrated in Figure 13, in which a is the range, c_0 +

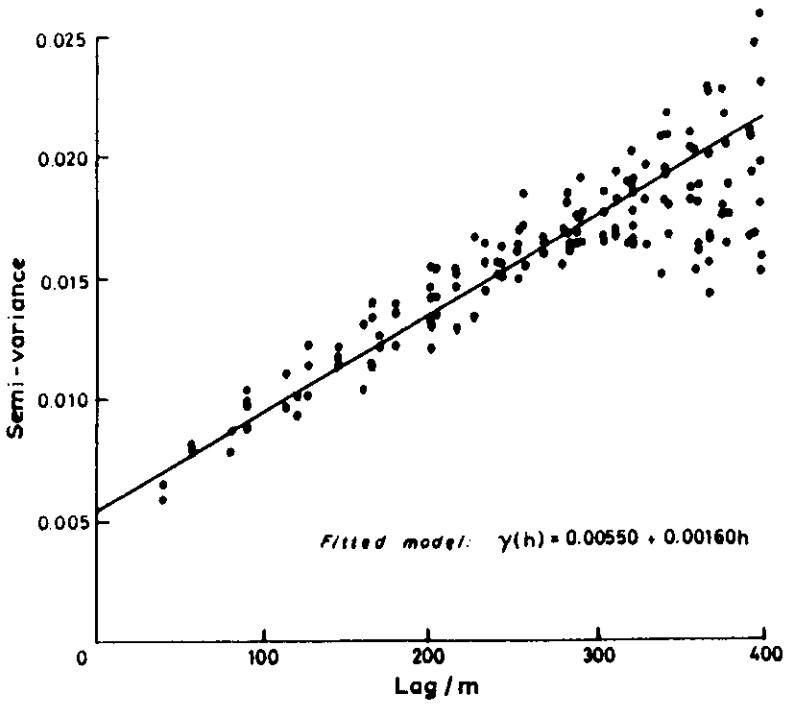

Figure 11. Semi-variogram of exchangeable potassium in topsoil, measured in μg/g and transformed to common logarithms, at Broom's Barn with linear model fitted (see Webster, 1981).

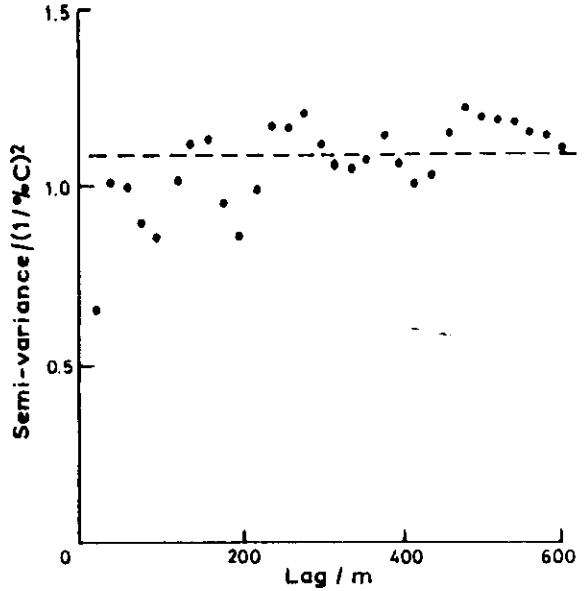

Figure 12. Semi-variogram of organic carbon in the subsoil, measured as percentage and transformed to reciprocal, at Tillycorthie, Aberdeenshire (see McBratney and Webster, 1981a).

c is the sill, and c_0 is the nugget variance as before. Its tangent at $h = 0$ cuts the sill at ⅓ a. Figure 13 shows the sample semi-variogram of pH over 420 m well fitted by the spherical model.

Theoretically, the spherical model derives from a moving average random process. It is the three-dimensional form of a quite general n-dimensional model (Matérn, 1960). The idea is that in three dimensions a random function, of which the property measured is a realization, depends on the volume of the intersection of two spheres of equal radius. If a is the diameter of the spheres and h the distance between their centers then this volume, V, is

$$V = \frac{\pi}{4}\left(\frac{2a^3}{3} - a^2h + \frac{h^3}{3}\right) \qquad \text{for } h \leq a. \qquad [26]$$

If this is expressed as a fraction of the volume of a sphere by dividing by $\frac{4}{3}\pi a^3$ we obtain the autocorrelation function,

$$\rho(h) = 1 - \frac{3}{2}\frac{h}{a} + \frac{1}{2}\frac{h^3}{a^3}, \qquad [27]$$

from which the semi-variogram is derived using the relation defined in equation [11].

Such a function would seem fairly obviously applicable to three-

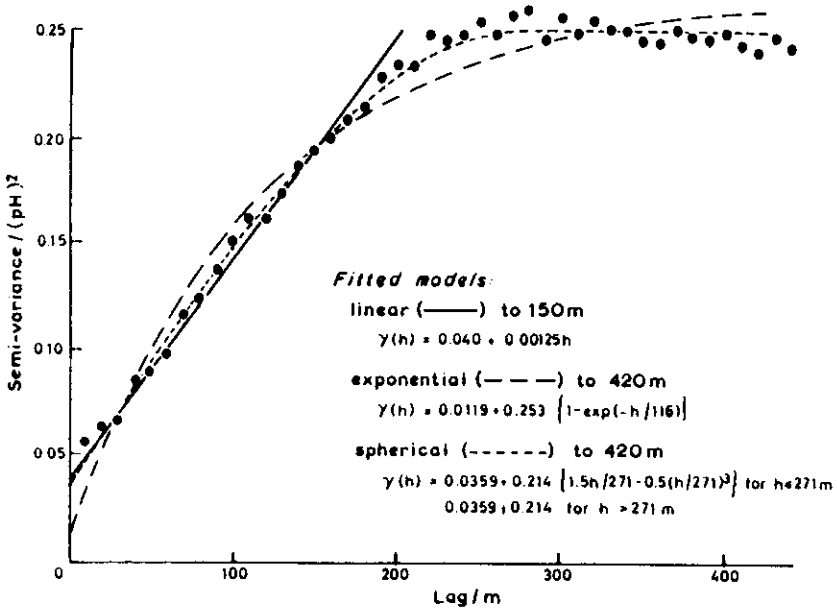

Figure 13. Sample semi-variogram of pH in the topsoil at Sandford St. Martin with linear, exponential, and spherical models fitted by least-squares approximation.

dimensional rock bodies, and it has indeed proved so. It is less obviously relevant for distributions in one and two dimensions, and yet it nearly always fits experimental results from soil sampling better than the one- and two-dimensional analogs, e.g., Figure 14, which are described later. The reason is presumably that there are additional sources of variation in soil at other spatial scales.

Similar interpretations of variograms near Witney (Webster, 1973) and Sandford (Webster and Cuanalo, 1975), both in Oxfordshire, and of semi-variograms at Tillycorthie in northeast Scotland (McBratney and Webster, 1981a) led the authors to seek distinct parcels of different soil type with linear dimensions equal to the range. In all instances they were able to find boundaries at approximately this average spacing using statistical search procedures devised by Webster (1978) and Hawkins and Merriam (1974), and so confirm their interpretations.

3. Exponential Models

The formula of the exponential model is

$$\gamma(h) = c_0 + c\,[1 - \exp(-h/r)] \qquad \text{for } h > 0$$
$$\gamma(0) = 0. \qquad [28]$$

The spatially dependent variance and nugget are c and c_0 as before, and r is a distance parameter controlling the spatial extent of the function. Here $\gamma(h)$ approaches the sill asymptotically, and so there is no strict finite range. Nevertheless, it is clear that for practical purposes the semi-variance ceases to increase beyond some point, and a commonly used rule of thumb is to take this as the effective range, $a' = 3r$ at which $\gamma(a')$ is then equal to approximately $c_0 + 0.95c$. Figure 13 shows the form of the model. The curve is the best fitting exponential to the sample semi-variogram of pH but clearly does not represent the sample values well.

The exponential model has an important place in statistical theory. It represents the essence of randomness in space. It is the semi-variogram of first-order auto-regressive and Markov processes. As autocorrelation function it has been the basis of several theoretical studies of the efficiency of sampling designs, by for example, Cochran (1946), Yates (1948), Quenouille (1949), and Matérn (1960). Semi-variograms of exponential form are also to be expected where differences in soil type are the main contributors to soil variation and where soil boundaries occur as a Poisson process. Burgess and Webster (1984) and Webster and Burgess (1984a) found the latter to be so along transects in many instances. If the intensity of the process is α, then the mean distance between the boundaries is $\bar{d} = 1/\alpha$, and the semi-variogram is

$$\gamma(h) = c\,[1 - \exp(-h/\bar{d})]$$
$$= c\,[1 - \exp(-\alpha h)]. \qquad [29]$$

Oliver (1984), working in the Wyre Forest of England, obtained exponential semi-variograms for a number of soil properties all with approximately the same distance parameter. She attributed their form largely to this process.

Sisson and Wierenga (1981) found that the rate at which water filtered into soil varied exponentially with distance and postulated it as a first-order auto-regressive process. On a very different scale Yost *et al.* (1982) estimated the semi-variograms of several chemical properties to 60 km over the Island of Hawaii and were able to fit exponential models in all cases. Similarly Xu and Webster (1984) found the pH of the soil in Zhangwu, a region covering 3500 km² in northeast China, to be distributed exponentially.

4. The Hyperbola

A novel function in this group is that used by Vieira *et al.* (1981) to describe variation in infiltration across irrigated land in the Central Valley of California, and shows the strong influence of physical chemistry in soil science even in this field. The function is a hyperbola,

$$\gamma(h) = \frac{h}{\alpha + \beta h}, \qquad [30]$$

which is the equation of the Langmuir adsorption of isotherm, with $\gamma(h)$ replacing the mass adsorbed and h the pressure. The model is transitive: It approaches the sill value $1/\beta$ asymptotically and has a limiting gradient of $1/\alpha$ at the origin. Nielsen[1] has stated that it is positive definite in two dimensions. For infiltration rate in mm hr^{-1} and lag in m, Vieira *et al.* found values for α and β as 1.207 and 0.105, respectively. There was no intercept, though it would be perfectly in order to add a nugget term.

C. Risky Models

In most instances soil properties appear transitive: The semi-variogram appears monotonic increasing to a sill, and this is to be expected in a finite region. The main difference among them is the degree of curvature. The exponential function curves gradually. The spherical model curves more tightly. But there are instances where the semi-variogram appears to curve more tightly still, even abruptly, and the investigator may be tempted to fit more tightly curving models. We consider two below.

1. The Circular Model

Just as we obtained the spherical model for three dimensions we can derive a circular model in two (Zubrzycki, 1957; Dalenius *et al.*, 1961). The area of intersection of two equal circles is given by

$$A = \frac{a^2}{2}\cos^{-1}\left(\frac{h}{a}\right) - \frac{h}{2\pi}\sqrt{a^2 - h^2} \qquad \text{for } h \leq a, \qquad [31]$$

where a is the diameter of the circles and h the distance between their centers. Expressing this as a fraction of the area of the circle gives

$$\rho(h) = \frac{2}{\pi}\left[\cos^{-1}\left(\frac{h}{a}\right) - \frac{h}{a}\sqrt{1 - \frac{h^2}{a^2}}\right] \qquad [32]$$

for the autocorrelation function, and the following semi-variogram:

$$\gamma(h) = c_0 + c\left[1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a}\right) + \frac{2h}{\pi a}\sqrt{1 - h^2/a^2}\right] \quad \text{for } 0 < h \leq a$$
$$\gamma(h) = c_0 + c \quad \text{for } h > a$$
$$\gamma(0) = 0. \qquad [33]$$

[1]Personal communication.

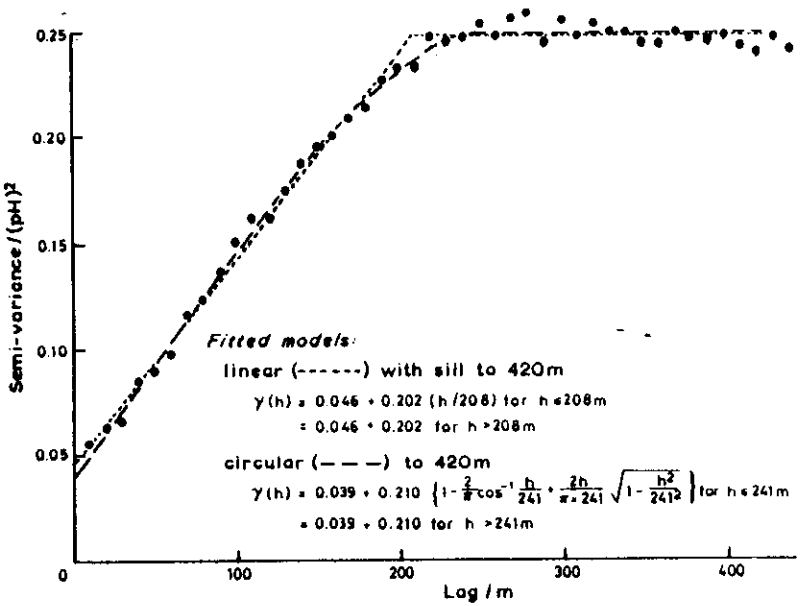

Figure 14. Sample semi-variogram of pH in the topsoil at Sandford St. Martin. Linear-with-sill and circular models are fitted by least-squares approximation.

This model and its consequences for plane sampling were explored in detail by Dalenius *et al.* (1961). Figure 14 shows it fitted to the sample semi-variogram of sand content at Sandford. It clearly describes the results well.

2. Linear Model with Sill

The extreme form of transitive model is linear with a sill:

$$\gamma(h) = c_0 + c\left(\frac{h}{a}\right) \qquad \text{for } 0 < h \leq a$$
$$\gamma(h) = c_0 + c \qquad \text{for } h > a$$
$$\gamma(0) = 0. \qquad [34]$$

I have expressed it in this way to emphasize its similarity to that of the spherical and circular models. Again, by analogy with the spherical scheme this can arise from linear zones of influence of equal length but with varying distance between their centers, h. For a given length, a, the overlap is simply $a-h$, or as a proportion $1-h/a$, provided $h \leq a$.

There are several examples in the literature where investigators have used this form of model for semi-variograms of soil. Figure 14 shows it fitted to

the same Sandford semi-variances as above, and again the fit is reasonable.

What, then, are the risks of using these models? They concern the positive definite condition. The circular model is positive definite in one and two dimensions, but not in three. This will usually not matter in soil survey, which is essentially concerned with variation in the plane. The linear model with sill is positive definite in one dimension, but not in more. This model should not be used in soil survey, however well it might appear to fit, unless variation is strictly limited to just the one dimension. Those investigators who have used the model in two-dimensional contexts have either not examined the variances of combinations of values or have had the fortune not to encounter negative variances in practice.

D. Nested Models

This paper began with a study of nested sampling and an analysis based on a nested model with variation deriving from several sources with different spatial scales. The components of variance measured the amount of variance contributed by each scale, and by accumulating them we were able to show how variance increased with increasing distance. Starting at the bottom of the hierarchy, the variance at the spacing of level 3 consists of the sum of components σ_4^2 and σ_3^2, that at level 2 consists of $\sigma_4^2 + \sigma_3^2 + \sigma_2^2$, and so on. Miesch (1975) showed that this is precisely the same as in the semi-variogram: The variance at a lag of 2, say $\gamma(2)$, comprises that at lag 1, $\gamma(1)$ or σ_1^2, plus a component deriving from distances between 1 and 2, say, σ_2^2. We have, therefore,

$$\gamma(2) = \gamma(1) + \sigma_2^2. \qquad [35]$$

Thus, the cumulative variances from the nested analysis constitute a semi-variogram.

Most of the semi-variograms that we have seen so far are already nested models in the sense that they have nugget variances. They can be represented by

$$\tfrac{1}{2} \text{ var } [z(x) - z(x + h)] = \gamma(h) = \gamma_0(h) + \gamma_1(h). \qquad [36]$$

In other words, the semi-variogram is the sum of two functions, one representing pure nugget, $\gamma_0(h)$, and the other a spatially dependent one, $\gamma_1(h)$. This can be extended by adding further functions.

In studies of mineralization it is often found that one more term is needed to describe the variation adequately; thus,

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) + \gamma_2(h). \qquad [37]$$

Mine surveyors and geochemists have found the double spherical nested model especially valuable, and my colleagues and I have fitted it to several sample semi-variograms of soil properties (McBratney *et al.*, 1982; Webster and Nortcliff, 1984). Figure 1.18 is an example. In these $\gamma_1(h)$ and $\gamma_2(h)$ are separate semi-variograms:

$$\begin{aligned}
\gamma_1(h) &= c_1\left[\frac{3}{2}\frac{h}{a_1} - \frac{1}{2}\left(\frac{h}{a_1}\right)^3\right] && \text{for } 0 < h \le a_1 \\
\gamma_1(h) &= c_1 && \text{for } h > a \\
\gamma_2(h) &= c_2\left[\frac{3}{2}\frac{h}{a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right] && \text{for } 0 < h \le a_2 \\
\gamma_2(h) &= c_2 && \text{for } h > a_2. \qquad [38]
\end{aligned}$$

The two sills c_1 and c_2 are in general different. The ranges will usually be very different; otherwise, they are unlikely to be distinguished.

The quantity $\gamma_0(h)$ is the nugget variance: a semi-variogram with its effective range much less than the smallest sample interval measured.

These three semi-variograms comprise the nested model fitted to the semi-variogram of readily extractable copper content in the topsoil of southeast Scotland and are shown separately in Figure 15, from McBratney *et al.* (1982). The coefficients are given below. The variances are in units of $(\log_{10} \mu g \text{ Cu/g soil})^2$ and the ranges in km.

c_0 (nugget)	0.0213
c_1	0.0257
c_2	0.0196
a_1	2.26
a_2	15.5

Recently Burrough (1983b) has postulated a nested structure for soil variation that arises from independent soil-forming factors that have operated over distinctly different spatial scales. He mentions the effects of geology, relief, and earthworms, to which one can add those of tree-throw and manmade divisions into fields and farms as examples of other scales. The effects can be ranked according to their ranges, and Burrough proposed a geometric scale in much the same way as Youden and Mehlich (1937) did for their nested sampling. As an example Burrough chose ranges, a, on a scale such that $a_1 = \frac{1}{3}a_2 = \frac{1}{9}a_3 = \frac{1}{27}a_4 \ldots$. Thus in one dimension the semi-variogram has the general form of equation [37] comprised as

$$\gamma(h) = c_1\frac{h}{a_1} + c_2\frac{h}{a_2} + c_3\frac{h}{a_3} + c_4\frac{h}{a_4} + \ldots . \qquad [39]$$

In other words, it consists of a series of linear semi-variograms with sills that together form a linear spline. Burrough estimates the sill values, c_i, by solving a system of linear equations and shows how such models fit actual one-dimensional semi-variograms. The fits are undoubtedly good. His

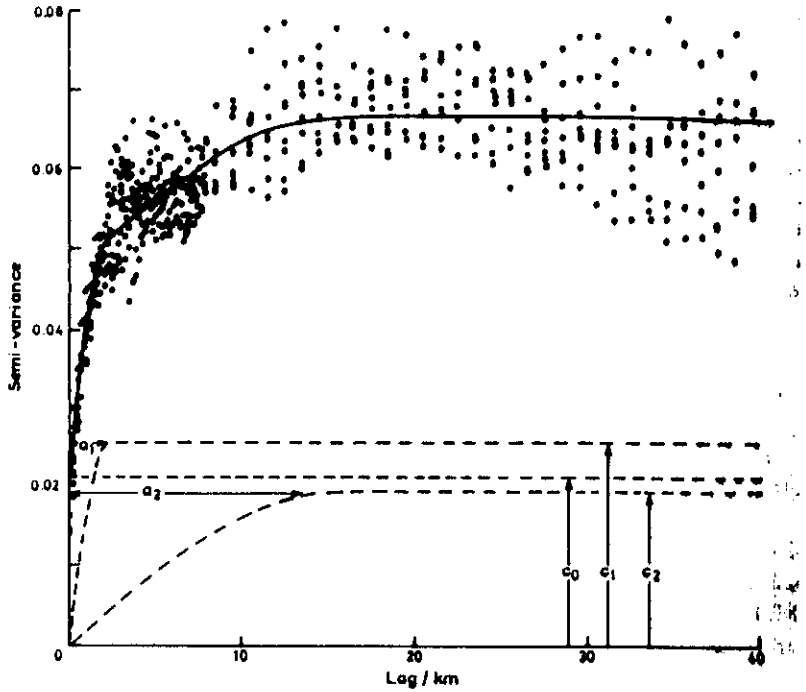

Figure 15. Semi-variogram of available copper in the topsoil of southeast Scotland. The three component semi-variograms of the nested spherical model are shown by the dashed lines (see McBratney *et al.*, 1982).

expression for the semi-variogram of pH in the topsoil at Sandford, Figure 14, converted into the terms of equation [34] is

$$\begin{aligned} \gamma(h) &= 0.048 + 0.200\frac{h}{209} && \text{for } 0 < h < 209 \\ \gamma(h) &= 0.048 + 0.200 && \text{for } h > 209 \\ \gamma(0) &= 0, \end{aligned} \qquad [40]$$

which agrees very closely with the model I fitted in Figure 14.

The model is postulated specifically for one dimension. In more dimensions linear splines are not generally positive definite, and the component semi-variograms should be replaced by their two- or three-dimensional analogs, the circular or spherical models.

E. Models for Anisotropy

As discussed earlier, soil does not necessarily vary equally in all lateral directions. Figures 6 to 9 showed an example in which the stone content varied anisotropically. Another example, Figure 21, appears later, and McBratney and Webster (1983) analyzed a further case, Figure 17.

The semi-variogram for the individual directions in Figure 6 are very close to linear, though with different slopes. The contours of the semi-variogram, Figure 7, are approximately elliptical, and this suggests that the semi-variogram can be represented by a simple sinusoidal function. Burgess and Webster (1980a) initially represented this semi-variogram using polar coordinates by

$$\gamma(h, \theta) = c_0 + u(\theta)|\mathbf{h}|, \qquad [41]$$

where

$$u(\theta) = A\cos^2(\theta - \phi) + B\sin^2(\theta - \phi). \qquad [42]$$

The parameters are ϕ, the direction of maximum variation; A, the gradient of the semi-variogram in that direction; and B, the gradient in the direction $\phi + \pi/2$.

Burgess *et al.* (1981) later modified this to

$$u(\theta) = [A^2\cos^2(\theta - \phi) + B^2\sin^2(\theta - \phi)]^{1/2}. \qquad [43]$$

This last expression now defines strict geometric anisotropy (David, 1977), which can be made isotropic by a linear transformation of the coordinate system. It might help to appreciate this by imagining the land on a rubber sheet. By stretching the sheet in direction ϕ in the proportion A to B the semi-variogram in that direction will have the same gradient as that in the perpendicular direction. The proportion A/B can be regarded as the anisotropy ratio.

On fitting this model to the sample values, Burgess *et al.* (1981) obtained the following:

$$\begin{aligned} \gamma(h, \theta) = 9.06 &+ [7.18^2\cos^2(\theta - 1.00) \\ &+ 2.40^2\sin^2(\theta - 1.00)]^{1/2}|\mathbf{h}|. \end{aligned} \qquad [44]$$

As above, this would appear simply as a series of concentric elliptical isarithms in plan. On the cylindrical projection it can be represented as a defining envelope, shown by the lines of maximum and minimum gradient in Figure 9. These are

$$\gamma_1(h) = 9.06 + 7.18\,h$$

and

$$\gamma_2(h) = 9.06 + 2.40\,h, \qquad [45]$$

respectively. Where the anistropy is geometric, this form of display contains all the information in the semi-variogram except the orientation, and that can be stated.

In the linear case the quantity $u(\theta)$ is the gradient in direction θ. In transitive models the gradient $u(\theta)$ in equation [42] is replaced by a range

parameter, say $r(\theta)$, and A and B then represent the maximum and minimum ranges of the model. No one seems to have encountered semi-variograms of this kind so far for soil.

A further type of anisotropy of which soil scientists must be aware is where the property of interest has different sills in the different directions. This is known as zonal anisotropy (Journel and Huijbregts, 1978). Again there are no reports of its occurrence in the horizontal plane in the soil literature. Almost certainly it occurs when the vertical dimension is added, but no one has analyzed three-dimensional soil bodies or even vertical planes in this way yet to my knowledge.

V. Fitting Models

Choosing models to describe observed semi-variances and the procedures for fitting them are matters of some controversy. Indeed, they are more controversial than any other aspect of spatial analysis covered in this review.

The choice of model will obviously be governed by the general graphic appearance of the sample semi-variogram. For example, one will choose a function for pH at Sandford that turns more tightly than the exponential curve, Figure 13. But not any function that appears to fit will do: As above, functions must be positive definite.

An investigator will also be influenced by what is already known about the soil of a region—by the form that the semi-variogram is expected to take. The investigator may wish to interpret the semi-variogram in terms of some physical process or geological origin, and a model that makes this possible is likely to be preferred to one that does not.

The choice of model will also depend on the purpose for which it is wanted. The criteria for choosing models for estimation or interpolation might be quite different from those used for illustration or explanation. An investigator may also fit different models for different-sized regions depending on the maximum lag that is of interest (Webster and Burgess, 1984b).

Having decided the form of the model and the lag over which it is to be fitted, there is then a choice of fitting procedure. The linear model is, of course, easy to fit by simple regression. Usually the number of pairs of comparisons in the estimated semi-variances differ from one another, and so the regression should be weighted accordingly. Some models can be made linear by transformation and the parameters estimated again by regression. The Langmuir isotherm, for example, in the form given in equation [30] can be transformed to

$$\frac{h}{\gamma} = \alpha + \beta h, \qquad [46]$$

from which α and β are readily found.

Most of the useful models are non-linear, and so a less straightforward approach must be adopted. Journel and Huijbregts (1978) recommend procedures for finding the nugget variance and sill of transitive models. Many sample semi-variograms are approximately linear over the first few lags, and so by fitting straight lines and extrapolating intercepts on the ordinate can be found. As Figure 13 shows, this value is much the same as that obtained by non-linear least-squares approximation of the whole curve.

Their recommendation for finding a sill is less sound, mainly because of misunderstanding among readers. It will pay, therefore, to make a digression at this point to clarify matters.

The basic notion of transitive variation is that there is a distinct range within which variation increases with increasing lag distance, h. Beyond this range $\gamma(h)$ remains constant, however much larger the lag. In principle the semi-variance remains constant to infinity at what is known as the *a priori* variance of the property of interest. One aim in sampling and model fitting is to estimate this *a priori* variance by the sill of the semi-variogram.

If we were able to sample in infinite space then the spatial dependence that occurs within limited lags would be negligible. We could disregard it. The *a priori* variance would be equivalent to the population variance, σ^2, of classical statistics, which we could estimate in the usual way by

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} [z(\mathbf{x}_i) - \bar{z}]^2, \qquad [47]$$

where $\bar{z}$ is the sample mean. Of course, we never have an infinite space in which to sample, but if the region, B, is very large in relation to the range of the semi-variogram, the regional variance, say σ_B^2, and our estimate of it, $\hat{\sigma}_B^2$, might be so close to the *a priori* variance that it will serve.

In practice, however, we are more often concerned with situations in which the range of the semi-variogram is a considerable proportion of the distance across the region. Now the population variance of a finite region, or *dispersion variance* as it is known to geostatisticians, is the average semi-variance within that region. Formally it is

$$\sigma_B^2 = \frac{1}{B^2}\int_B \int_B (\mathbf{x} - \mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}', \qquad [48]$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two points that describe the region independently. It is represented in Figure 16 by the area under the curve. Clearly, this is less than the area beneath the sill. In fact the dispersion variance, which one would estimate by equation [47] from a sample is that value which makes the two hachured areas equal. It must be less than the sill, unless the semi-variogram is pure nugget at the working scale. Thus, the practice of

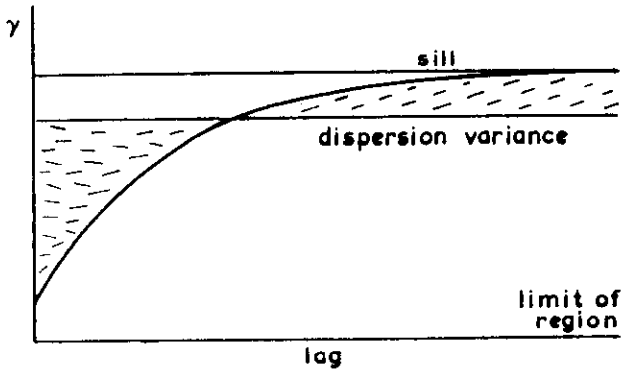

Figure 16. Transitive semi-variogram and dispersion variance of a finite region.

estimating the sill by the sample variance within a region is not to be recommended.

A. Recommended Practice

Despite the criticisms of automatic least-squares fitting made by Journel and Huijbregts (1978) and others, my colleagues and I have found that it works well where there are sufficient data. For non-linear models we first obtain values for the non-linear parameters, notably the range, iteratively and then fit the remainder by solving a matrix equation. We weight the semi-variances by the number of pairs of comparisons used to estimate them. We use the Maximum Likelihood Program, MLP, written by Ross (1980) at Rothamsted for this and program our own models into it. Other sound suggestions are made by Armstrong (1984) in a recent discussion of the problem. She clearly favors a weighted area method involving Laplace transforms that is especially well suited for exponential semi-variograms. Kitanidis (1983) discusses the problem further and proposes a true maximum likelihood solution for fitting parameters to the whole set of covariances among the data. It is likely to be restricted to fairly small sets of observations, however. Readers wishing to analyze spatially distributed soil data should read these papers and also that by Delfiner (1976).

VI. Fractal Representation

We have already remarked on the nested nature of soil variation. We noted that soil can apparently vary increasingly without limit as the area increases. At the other end of the scale most semi-variograms recorded on point supports have nugget variances. There always seems to be some

variance within the smallest sampling interval, and more to encounter if we go further afield. It is clear, therefore, that the semi-variogram obtained in any study depends very much on the scale of the study. There is no absolute semi-variogram for a soil property.

Gajem *et al.* (1981) illustrate this dependence on scale well. They examined a region of 85 ha at three scales by sampling on transects at 20 cm, 2 m, and 20 m intervals and increasing the lengths of the transects in proportion. As the length of transect increased so did the variance and the range of dependence.

Such facts led Burrough (1983a) to consider whether soil exhibited true similarity at different scales and to explore its representation by fractals (Mandelbrot, 1982). He took as starting point the stochastic fractal of linear Brownian motion,

$$z(x) = z(x + h) + \varepsilon, \qquad [49]$$

where ε is a Gaussian random deviate. Its semi-variogram has the one-dimensional form,

$$E\{[z(x) - z(x + h)]^2\} = 2\gamma(h) = h^{2H}, \qquad [50]$$

where $H = 0.5$. If the sampling interval is divided by any arbitrary positive value r and the result rescaled in the ratio r^H then the new semi-variogram will be identical with the old one. In this sense Brownian motion is a self-similar or fractal process.

In ordinary Brownian motion the successive values of ε are totally independent. Mandelbrot, however, showed that by increasing H in the range 0.5 to 1 he could generate a family of trails smoother than Brownian motion in which successive values of ε were positively correlated. Conversely, if he diminished H between 0.5 and 0 so that the ε were negatively correlated, then the tracks were noisier than those of Brownian motion.

At the limit $H = 1$ the semi-variogram is a parabola and variation is smooth and differentiable. At the other extreme $H = 0$ indicates pure noise, which seems impossible in continuous space. Indeed, Journel and Huijbregts (1978) preclude these limiting values: The function $\gamma(h) = -h^2$ is not conditional positive definite. Following Orey (1970) and Berry and Lewis (1980) Burrough points out that H is related to the Hausdorff-Besicovitch or fractal dimension D by $H = 2-D$, and by plotting the semi-variogram on double logarithmic graph paper he estimates D from the slope, $m = 4 - 2D$.

Burrough estimated D for a number of soil properties in several different regions and obtained values ranging from 1.5 to 2. His values of H therefore lay between 0 and 0.5, suggesting that soil varied in a way more random than Brownian motion. While this may be so, Burrough concluded that the different independent factors that affect soil variation combine in a way that overshadow any self similarity. The fractal Brownian model seems less appropriate for soil than the nested models described above

(Burrough, 1983b). Nevertheless, the fractal representation should be borne in mind by soil scientists, and D values of soil properties computed for reference against those of Brownian fractals (Burrough, in press).

VII. Cross Correlation

So far we have considered the spatial correlation of a soil property with itself; i.e., its autocorrelation. The same concepts extend to two properties, so that the values of the one depend in the statistical sense on those of the other at other places nearby. Where this is so the variables are said to be co-regionalized or cross correlated.

If we adhere to the intrinsic hypothesis as the strongest assumption then we can express the spatial interdependence of two variables, say u and v, by their cross semi-variogram, $\gamma_{uv}(\mathbf{h})$. The cross semi-variance is defined by

$$\gamma_{uv}(\mathbf{h}) = \mathrm{E}\{[z_u(\mathbf{x}) - z_u(\mathbf{x} + \mathbf{h})]\,[z_v(\mathbf{x}) - z_v(\mathbf{x} + \mathbf{h})]\}, \quad [51]$$

and is estimated like the auto semi-variance by

$$\hat{\gamma}_{uv}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})} [z_u(\mathbf{x}_i) - z_u(\mathbf{x}_i + \mathbf{h})]\,[z_v(\mathbf{x}_i) - z_v(\mathbf{x}_i + \mathbf{h})], \quad [52]$$

where $N(\mathbf{h})$ is the number of pairs of comparisons that can be made at lag $\mathbf{h}$. Sample cross semi-variograms can be computed in one dimension from transect data and in two dimensions from grids, and the same devices can be adopted for dealing with missing data and less regular sampling as are used for single variables.

There is, however, one feature of cross correlation that does not occur with single variables. It concerns the symmetry of the relation. If we assume that the two properties are second-order stationary and consider their relations in one dimension, then we can define the cross covariance $C_{uv}(h)$ at lag h as

$$C_{uv}(h)\ \mathrm{E}\{[z_u(x) - \bar{z}_u]\,[z_v(x + h) - \bar{z}_v]\}. \quad [53]$$

The cross correlation is then

$$\rho_{uv}(h) = C_{uv}(h)/\,[C_{uu}(0) \cdot C_{vv}(0)]^2, \quad [54]$$

and the cross semi-variance is related to it by

$$\begin{aligned} 2\gamma_{uv}(h) &= 2C_{uv}(0) - C_{uv}(h) - C_{vu}(h) \\ &= 2C_{uv}(0) - C_{uv}(-h) - C_{vu}(-h). \end{aligned} \quad [55]$$

If we interchange u and v in either equation [51] or equation [55] for $\gamma_{uv}(h)$ it makes no difference. If, however, we interchange them in the equation for the cross covariance, equation [53], we obtain a different result in general.

The value of $C_{uv}(h)$ is only incidentally the same as $C_{vu}(h)$. Small differences in their estimates can be regarded as sampling effects, but any large difference means that one variable lags behind the other. In these circumstances the coregionalization can be represented fully only by the cross covariogram.

The same rules for fitting models apply to cross semi-variograms as to auto semi-variograms. The models must be such that any linear combination of values,

$$Y_k = \sum_{i=1}^{n} \lambda_{ki}\, z_k(\mathbf{x}_i), \quad [56]$$

has a non-negative variance for $k = u$ and $k = v$, and indeed for any other variables in the model. This implies that at any lag $\mathbf{h}$ the cross covariance matrix of the variables is positive definite. The matter is somewhat complex, and readers who wish to delve more deeply should consult Journel and Huijbregts (1978).

Vauclin *et al.* (1983) and McBratney and Webster (1983) have analyzed co-regionalizations of soil properties, and the following example from the latter paper illustrates the techniques.

McBratney and Webster examined the cross correlation of particle size fractions in the topsoil and subsoil of a 10.6 ha field on the Woburn Experimental Station in central England. The field had been sampled at the nodes of a 10 m square grid, and cores 6 cm in diameter were taken from 0 to 20 cm and 40 to 60 cm depth. The percentages of sand (2 mm to 63 μm), silt (63 μm to 2 μm), and clay (< 2 μm) were determined on every subsoil core but on only every fourth sample of topsoil. The percentages were transformed to logit ($\ln[p/(100 - p)]$ where p is a percentage) to stabilize variances.

As might be expected there was some strong simple correlation among the fractions. This was, however, restricted largely to the sand and silt, for which correlation coefficients were all larger than 0.8 in absolute value. There was also a marked anisotropy, and a part of the analysis was concerned to establish the extent to which this was common to all the particle size fractions.

All the semi-variograms were linear, at least to 100 m, and so had the form of equation [44]. Table 3 lists the coefficients: ϕ the angle of maximum gradient, A the gradient on that direction, B the gradient in direction $\phi + \pi/2$, and c_0 the nugget variance. The table introduces a feature of cross correlation that does not occur with autocorrelation: The semi-variances can be negative. Figures 17, 18, and 19 show three examples from the possible 15 cross semi-variograms, the three that the authors later used for co-kriging. In two of them the semi-variances are negative.

Table 3 shows that the anisotropy ratio A/B was very similar for sand and silt at both depths, and that the directions of anisotropy were much the

Table 3. Coefficients of the Geometric Anisotropic Semi-Variogram for Sand and Silt, Equation [41], Found by Fitting the Model Independently to Each Variable

	Topsoil		Subsoil	
	Sand	Silt	Sand	Silt
Nugget variances, c_0				
Topsoil sand	0	0	0	0
silt		0	0	0
Subsoil sand			0.0675	−0.0665
silt				0.0968
Maximum gradients, A				
Topsoil sand	0.0128	−0.0145	0.0198	−0.0239
silt		0.0172	−0.0230	0.0285
Subsoil sand			0.0341	−0.0418
silt				0.0547
Minimum gradients, B				
Topsoil sand	0.0080	−0.0087	0.0118	−0.0138
silt		0.0107	−0.0132	0.0159
Subsoil sand			0.0230	−0.0272
silt				0.0342
Anisotropy ratios, A/B				
Topsoil sand	1.590	1.671	1.680	1.731
silt		1.709	1.739	1.788
Subsoil sand			1.483	1.538
silt				1.570
Weighted average 1.590				
Directions of maximum variation,				
Topsoil sand	1.148	1.147	1.166	1.188
silt		1.151	1.176	1.208
Subsoil sand			1.165	1.163
silt				1.192
Weighted average 1.172				

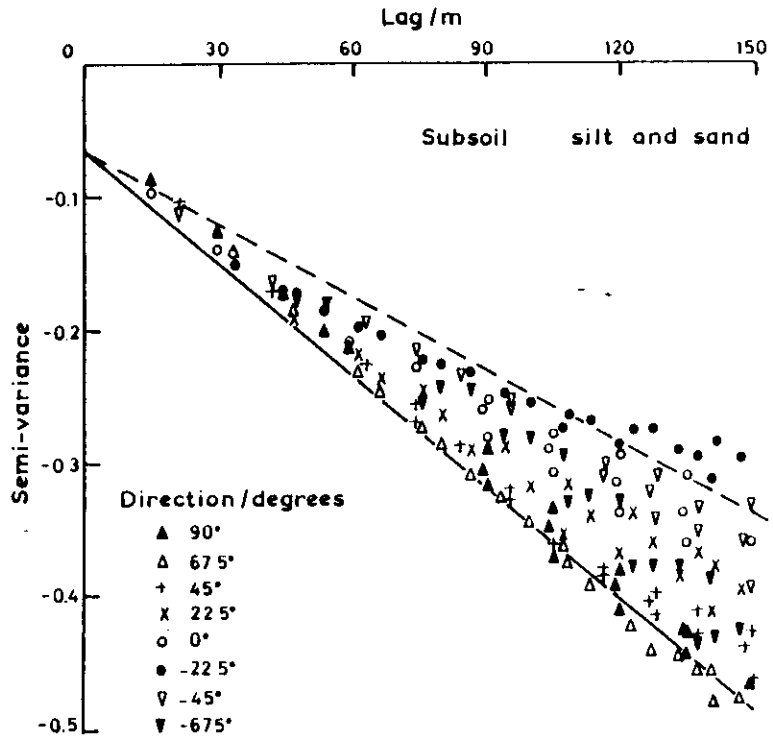

Figure 17. Semi-variogram of subsoil silt and sand at Woburn. Here and in Figures 18 and 19 the symbols for the semi-variances show the different directions and the oblique lines the enclosing envelope of the fitted model (see McBratney and Webster, 1983).

same. It seemed that the values obtained were estimates of common values for the two parameters. Weighted averages of them were therefore computed, giving $A/B = 1.59$ and $\phi = 1.17$ rad. With these held constant the model of equation [44] was fitted afresh to give the coefficients listed in Table 4. The authors also presented model coefficients for clay auto semi-variograms. The anisotropy ratios were fairly similar, but the angles of orientation were different, especially that for the subsoil.

VIII. Changing Drift

Most of the semi-variograms that we have seen so far have had nugget variances, and all have been either linear or convex upward. Occasionally, however, we may encounter semi-variograms of soil properties that

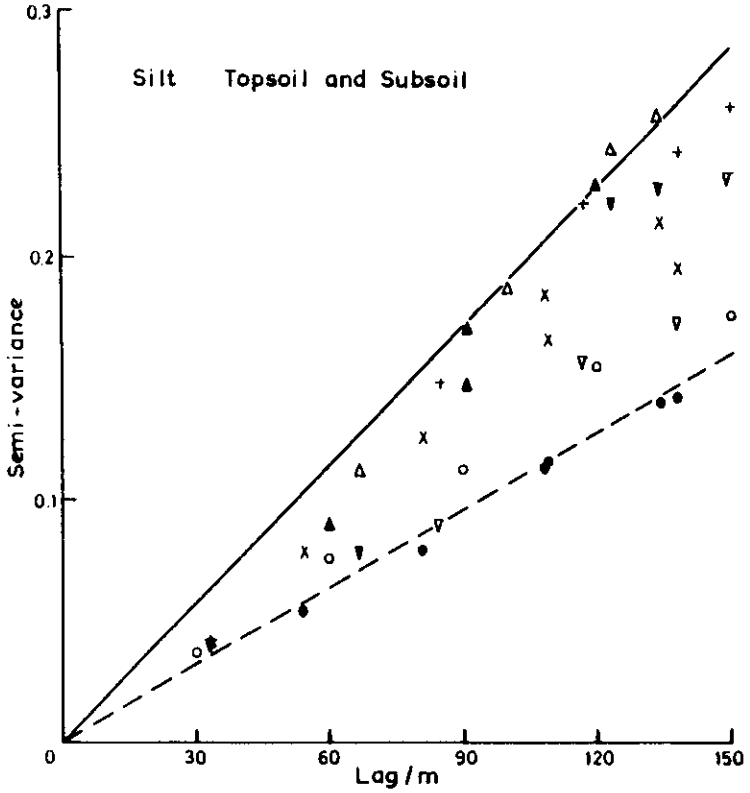

Figure 18. Semi-variogram of silt in the topsoil and that in the subsoil at Woburn.

approach the origin and are concave upward, approximately parabolic there. This shape should immediately signal a departure from the assumptions of the intrinsic hypothesis. It is a sign of smooth continuity in the measurements of the soil, so that there are local trends.

Figure 20 shows an example in which the electrical resistivity of the soil was measured at 1-m intervals (Webster and Burgess, 1980), and Figure 21 presents the semi-variograms for the four principal directions of the grid. There are a few places where the resistivity appears fairly constant apart from point-to-point fluctuation; e.g., around 52 m and from 94 m onwards in the upper graph. Elsewhere, however, there are local trends extending over 4 m to about 10 m. The expected value of resistivity is not constant, even within small neighborhoods, but changes with position. It is a function of position, thus

$$\mathrm{E}[z(\mathbf{x})] = m(\mathbf{x}). \qquad [57]$$

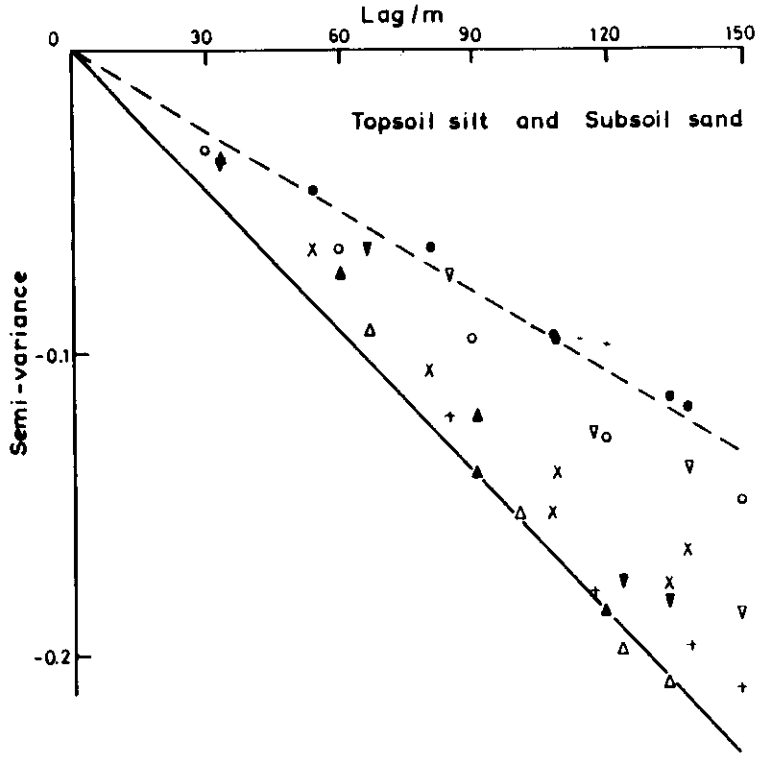

Figure 19. Semi-variogram of topsoil silt and subsoil sand at Woburn.

Our model of the soil for the stationary case, equation [7], must be changed by replacing μ, the mean, by the more general $m(\mathbf{x})$, giving

$$z(\mathbf{x}) = m(\mathbf{x}) + \varepsilon(\mathbf{x}). \qquad [58]$$

The quantity $m(\mathbf{x})$ representing the trend is known in regionalized variable theory as the *drift*. The term is well entrenched, and despite possible confusion with material transported by ice we shall adhere to it.

A. Structural Analysis

Where the drift changes, rearranging equation [58] gives $\varepsilon(\mathbf{x})$ as the deviation from the expectation

$$\varepsilon(\mathbf{x}) = z(\mathbf{x}) - m(\mathbf{x}). \qquad [59]$$

In these circumstances equations [6] and [8] are not equivalent. The raw semi-variance no longer estimates the expected squared difference between

Table 4. Coefficients of the Geometric Anisotropic Semi-Variogram for Sand and Silt, Equation [41], with Common Anisotropy Ratio, 1.59, and Direction, 1.172 rad

	Topsoil		Subsoil	
	Sand	Silt	Sand	Silt
Nugget variances, c_0				
Topsoil sand	0	0	0	0
silt		0	0	0
Subsoil sand			0.0715	−0.0668
silt				0.0998
Maximum gradients, A				
Topsoil sand	0.0128	−0.0143	0.0195	−0.0233
silt		0.0169	−0.0224	0.0275
Minimum gradients, B				
Topsoil sand	0.0080	−0.0090	0.0122	−0.0146
silt		0.0106	−0.0141	0.0173
Subsoil sand			0.0215	−0.0263
silt				0.0338

the residuals at two places, and we need an analysis that will reveal both the drift and the deviations from it. This is by no means straightforward. Nevertheless, it can be important since it forms the basis of universal kriging.

The aim of the analysis, then, is to estimate the semi-variogram of the residuals, equation [8]. The problem arises because, as equation [59] shows, these residuals depend not only on the measured values, $z(\mathbf{x}_i)$, but also on the drift values, $m(\mathbf{x}_i)$, which we do not know. The problem is compounded because the semi-variogram must be known in order to estimate the drift. Thus, for a given set of data we could compute a semi-variogram if we knew the drifts, which we could calculate if we knew the semi-variogram. This apparent impasse is resolved by a structural analysis that combines trial and error with good judgment. Its solution owes much to Olea (1975) for an English text that leads step by step from the theory to its practical use. Olea (1977) has also provided a computer program for the analysis.

We proceed as follows. We first assume an expression for the drift:

$$m(\mathbf{x}) = \sum_{j=0}^{k} a_j f_j(\mathbf{x}) \qquad [60]$$

where $a_j, j = 0, 1, 2 \ldots k$ are unknown coefficients, and $f_j(\mathbf{x})$ are known

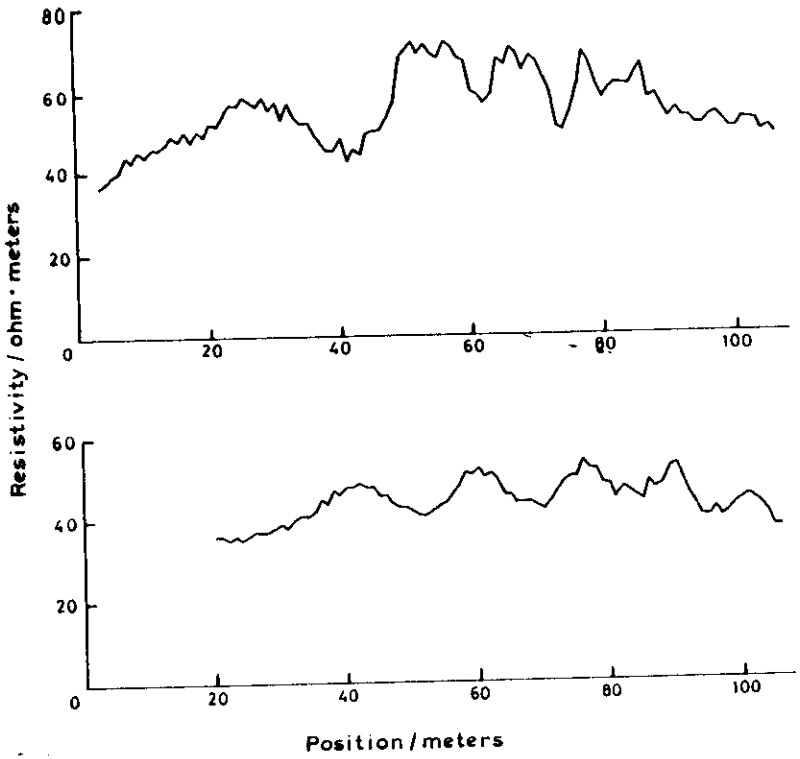

Figure 20. Electrical resistivity along two transects at Bekesbourne, England (see Webster and Burgess, 1980).

functions of $\mathbf{x}$. We also assume a theoretical semi-variogram and use it to evaluate the coefficients a_j in equation [60]. This gives us value for $m(\mathbf{x})$. Using these estimates of the drift we obtain residuals between them and the data and compute the semi-variogram of the residuals. This should match the theoretical semi-variogram.

Deriving the coefficients in equation [60] is lengthy and need not concern us here: Olea (1975) gives the full details. In practice we can usually keep the task fairly simple by sampling initially on a regular grid, by analyzing one direction at a time, and by restricting the neighborhood over which we consider the semi-variogram as in the stationary case. Given these restraints, the expression for the drift need never be more complex than a second-order polynominal, and the semi-variogram can be well approximated by a straight line,

$$\gamma(h) = wh, \qquad [61]$$

usually without a nugget variance.

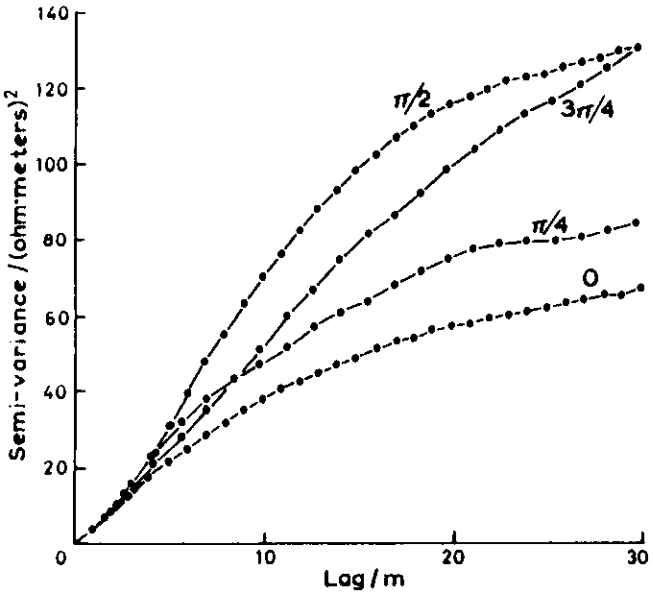

Figure 21. Semi-variograms of electrical resistivity in the soil at Bekesbourne, England, for four directions.

Equation [60] for the drift then takes one of the following forms:

$$m(x) = a_0$$
$$m(x) = a_0 + a_1 x$$
or $$m(x) = a_0 + a_1 x + a_2 x^2, \qquad [62]$$

where x relates to an origin within the neighborhood. These expressions can be combined with a linear semi-variogram that can be changed by lengthening or shortening the neighborhood over which it applies.

The constant a_0 cannot be calculated from the semi-variogram, but in any case is not needed since the semi-variogram of the estimated residuals does not depend on it. Olea shows that for regular sampling the other coefficients are given by

$$a_1 = \frac{z(x_n) - z(x_1)}{c(n-1)}, \qquad [63]$$

for linear drift, and for quadratic drift:

$$a_1 = \frac{z(x_n) - z(x_1)}{c(n-1)} - (n-1)\,a_2 c \qquad [64]$$

$$a_2 = -\frac{3[2_{i=1}(x_n) - z(x_1)]}{(n-2)(u-1)\,c^2}. \qquad [65]$$

Here x_1 and x_n are the extreme sampling points in the neighborhood, c is the sampling interval, and $\bar{z}$ is the mean of the sample in the neighborhood; i.e.,

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z(x_i). \qquad [66]$$

When residuals are calculated from either linear or quadratic drift their semi-variances are biased estimates of the semi-variances of the true residuals (Olea, 1975). The bias can be calculated by assuming that the semi-variogram is linear and that its slope is well estimated by the slope of the semi-variogram of the estimated residuals at the origin—effectively the slope over the first lag interval.

For linear drift the slope is

$$w = \frac{n-1}{n-2}\gamma_E(1), \qquad [67]$$

where $\gamma_E(1)$ is the estimated semi-variance at lag 1. The bias at lag h is then given by

$$\frac{wh^2}{c(n-1)}, \qquad [68]$$

and the semi-variances of the true residuals are

$$\gamma(h) = \gamma_E(h) + \frac{wh^2}{c(n-1)}. \qquad [69]$$

For quadratic drift the slope of the true semi-variogram is estimated from $\gamma_E(1)$ by

$$w = \frac{n-1}{n-3}\gamma_E(1), \qquad [70]$$

and the full semi-variances are then

$$\gamma(h) = \gamma_E(h) + \frac{wh^2[c^2(2n^2 - 2n - 1) - 2cnh + h^2]}{(n-2)(n-1)\,nc^3}. \qquad [71]$$

A choice is first made of combination of drift power and neighborhood size, and the analysis is performed. If the experimental semi-variogram matches the theoretical one well then the combination is judged an appropriate model to describe the variation of the soil property in the region. If, however, the match is poor then either the expression for the drift or the size of neighborhood or both are changed and the analysis is repeated. It is not always easy to decide whether a particular match is good, and even if the match seems good there might be better ones. So the analysis is repeated with successively larger neighborhoods, first for linear drift and then for quadratic drifts, and the results compared. The best fit is

chosen by inspecting graphs of the results, and there may be more than one equally good combination.

Webster and Burgess (1980) performed a full structural analysis on the resistivity data whose raw semi-variogram is shown in Figure 21. Figures 22 to 25 show examples for neighborhoods of just five and nine terms, equal to 4 and 8 lag intervals. Note, however, that an interval across the diagonals is $\sqrt{2}$ times that along the rows or columns of the grid. In the figures the circles represent the assumed linear semi-variogram and the stars the actual semi-variograms of the residuals.

As might be expected the agreement between assumed and actual semi-variograms is good for both linear and quadratic drifts when the

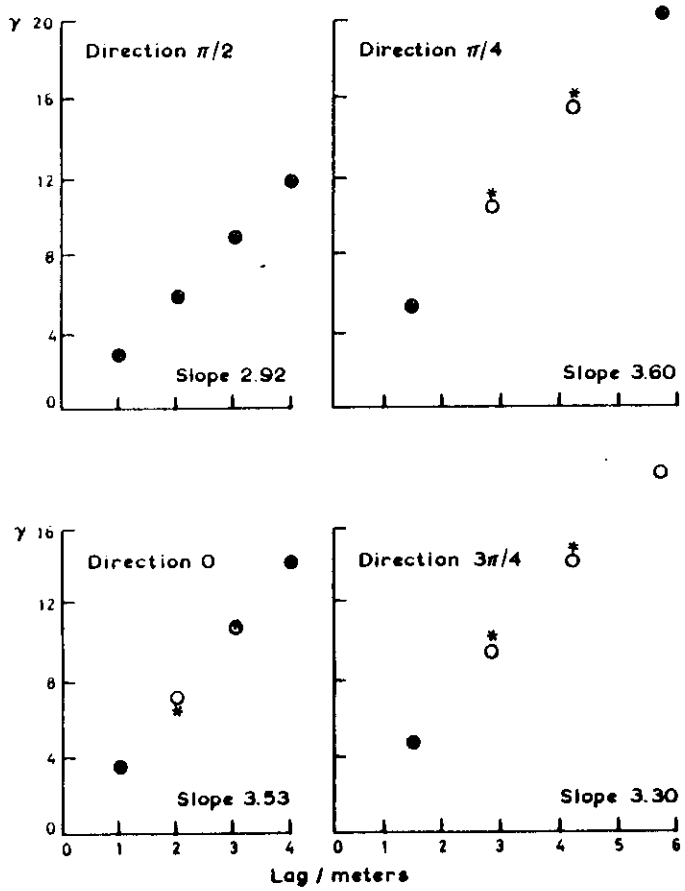

Figure 22. Semi-variograms of residuals from linear drift over five terms (four lag intervals) at Bekesbourne. Here and in Figures 22 to 24 the ordinate is in (ohm. meters)^2 and the lag in meters.

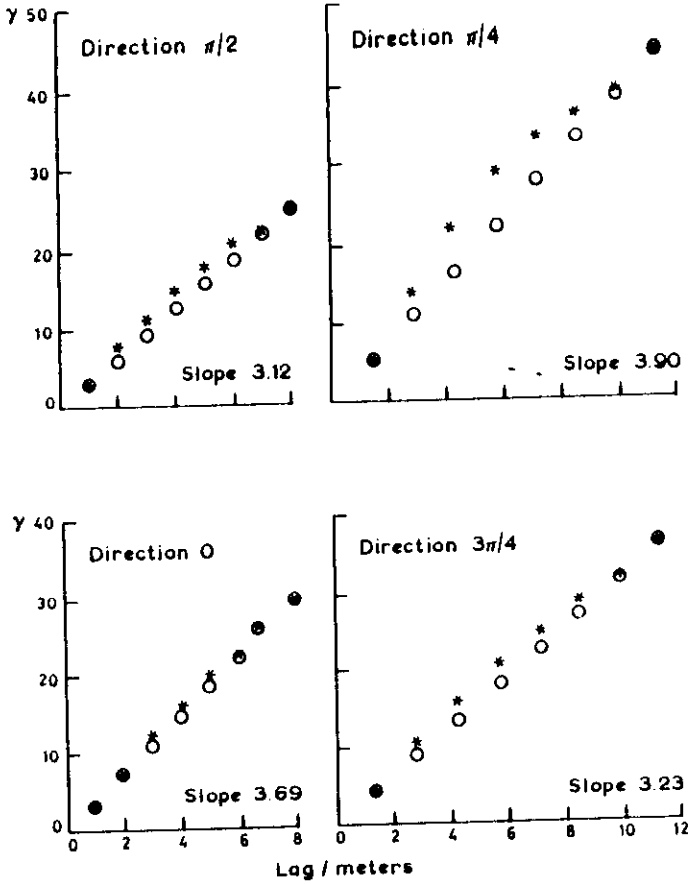

Figure 23. Semi-variograms of residuals from linear drift over nine terms.

neighborhood is restricted to five terms. When enlarged to nine terms the agreement is still good for the quadratic drift, but somewhat less so for the linear drift, especially in the north-south direction. Webster and Burgess found that on increasing the drift from these the agreement became steadily worse. Further, because they wished to use the semi-variogram for kriging (see later) across large gaps in their grid of data they chose the linear model in preference to the quadratic because the latter was inadequately constrained there.

On fitting equation [43] for geometric anisotropy to the semi-variograms from the linear drift over nine terms the following is obtained:

$$\gamma(h, \theta) = [3.913^2 \cos^2(\theta - 0.416) + 3.029^2 \sin^2(\theta - 0.416)]^{1/2} |\mathbf{h}|, \quad [72]$$

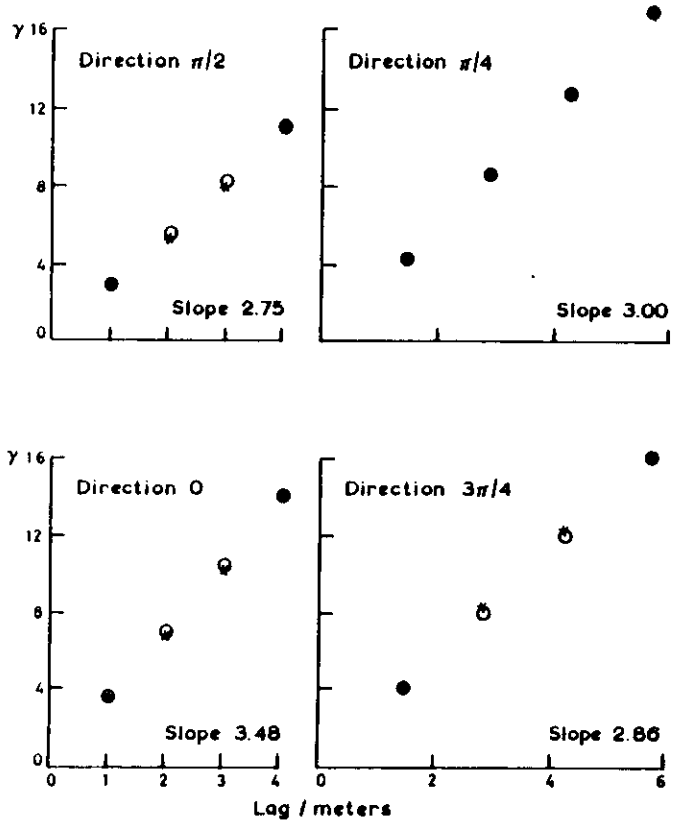

Figure 24. Semi-variograms of residuals from quadratic drift over five terms.

Table 5. Slopes in (ohm m)^2 m^{-1} of Semi-Variograms Assuming Constant Drift and of Residuals from Linear and Quadratic Drift

	Constant drift average slope		Linear drift slope		Quadratic drift slope	
Direction	5 terms	9 terms	5 terms	9 terms	5 terms	9 terms
NW to SE	5.82	6.88	2.92	3.12	2.75	2.73
SW to NE	4.42	4.03	3.53	3.69	3.48	3.49
N to S	5.59	4.60	3.60	3.90	3.00	3.45
W to E	4.91	5.26	3.30	3.23	2.86	3.05

where $h = |\mathbf{h}|$ is in meters and θ is in radians. Table 5 compares the calculated slopes, w, and values given by the fitted model, and shows good agreement between the two.

IX. Extension to the Power Spectrum

The semi-variogram and the co-variogram both express the spatial dependence of a property in the spatial domain. In time series analysis they express variation in the domain of time. Many temporal phenomena are periodic, and the semi-variogram will itself then show periodicity. In these cases it can be helpful to transform the semi-variogram, or better the co-variogram or correlogram into the frequency domain by Fourier analysis.

The transformed function is the *power spectrum*, or simply the *spectrum*. It expresses the distribution of variance with frequency. By definition the spectrum is given by

$$\Gamma(f) = \lim_{n\to\infty} \mathrm{E}[G(f)] = \int_{-\infty}^{\infty} C(h)e^{-j2\pi fh}dh, \qquad [73]$$

where f denotes the frequency, G is the spectral density for a finite series of length n, and j is $\sqrt{-1}$. If $C(h)$ is replaced by the autocorrelation then the transform gives the spectral density, which is

$$\mathrm{P} = \frac{\Gamma(f)}{\sigma^2} = \int_{-\infty}^{\infty} \rho(h)e^{-j2\pi fh}dh, \qquad [74]$$

and this, like the autocorrelation, no longer depends on the scale of measurement.

Estimating the spectrum is a large subject in its own right and still a matter of some controversy. Readers who wish to use the technique should consult one or more of the standard works by, for example, Chatfield (1980), Jenkins and Watts (1968), and Kendall and Stuart (1976). Basically, the required transform of the co-variogram is

$$\hat{\Gamma}(f) = \frac{1}{A}\left[\hat{C}(0) + 2\sum_{k=1}^{m}\hat{C}(i)\cos\pi kf\right], \qquad [75]$$

with the generally agreed value for A of 2π. The value of m is the maximum lag over which the summation is performed. Estimates so obtained, however, are subject to substantial error, and therefore they are usually smoothed. There are several methods for this. They involve viewing the

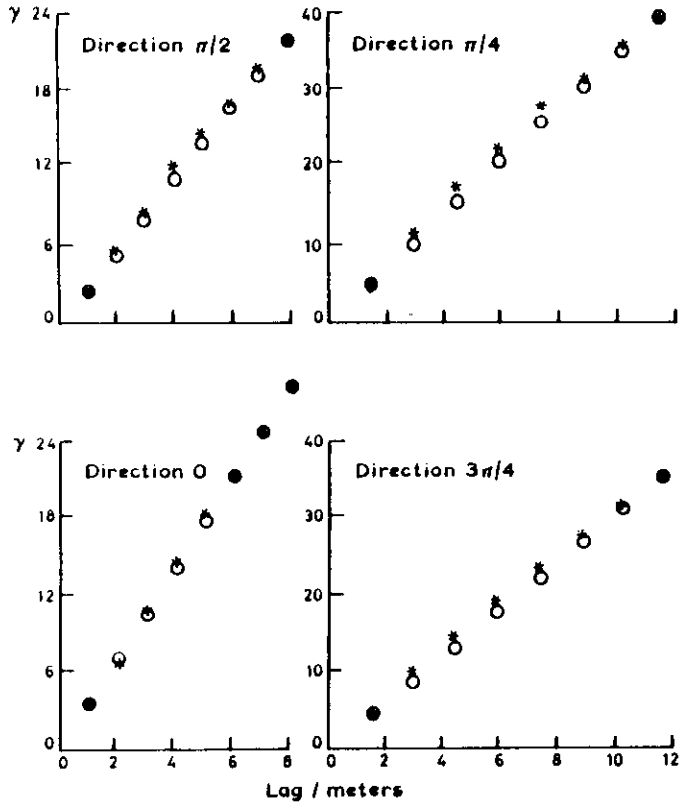

Figure 25. Semi-variogram of residuals from quadratic drift over nine terms.

spectrum through a *window*, the width of which can be chosen as desired: The wider the window the greater is the smoothing, though the larger becomes the risk of bias in the estimates. The ideal width is the one that achieves the best compromise between precision and accuracy in the estimates. The shape of the window can also be varied.

An example to illustrate this extension of spatial analysis is taken from Webster (1977). A transect had been sampled at 4-m intervals across gilgai terrain on the Bland Plain in eastern Australia, and at each sampling point several soil properties had been recorded. Among these was the electrical conductivity at 30- to 40-cm depth, measured in a 1:5 soil:water extract. There was a total of 355 observations over 1420 m. The measured values were strongly skewed, and they were therefore transformed to logarithms to stabilize the variances, and on this scale the series seemed fully second-order stationary.

The correlogram was computed. It is shown as its complement, the semi-variogram, in Figure 27. It was then transformed to its spectrum using the Bartlett window as smoothing function (Bartlett, 1966). This function is defined for the frequency domain as

$$W(f) = m\left(\frac{\sin fm}{fm}\right)^2, \qquad [76]$$

where m is the maximum lag of the transformation. In computing, however, it is usually applied to the correlogram in the spatial or time domain as a *lag window*, $w(f)$, which is simply

$$\begin{aligned} w(f) &= 1 - k/m \qquad \text{for } k \le m \\ w(f) &= 0 \qquad\qquad \text{for } k > m. \end{aligned} \qquad [77]$$

This then gives the computing formula for spectral density as

$$R(f) = \frac{1}{2\pi}\left[1 + 2\sum_{k=1}^{m}\hat{\rho}(k)\,w(k)\cos\frac{\pi fk}{L}\right], \qquad [78]$$

where f is the frequency ranging from 0 to ½ cycle and L is the number of steps by which f is incremented. A half cycle corresponds to two sampling intervals and is the limit, often known as the *Nyquist* frequency. If the spectrum is needed for shorter wavelengths, then sampling must be more intensive. In this study the frequency therefore ranged from 0 to 1/8 cycles per meter (wavelength 8 m) and L was set to 100.

The spectrum was computed with m set to four different sampling intervals, 10, 25, 40, and 60. Figure 26 shows the results. With a lag window of 10 the spectrum declines smoothly over the frequency range. As m is increased, however, more detail becomes apparent, and in particular a strong peak emerges at 0.12 cycles. This frequency, 0.12 cycles per 4 m, corresponds to a wavelength of 33.3 m, or a little more than eight sampling intervals. It signifies periodic fluctuation in the conductivity with this wavelength, a result that matched the morphological evidence.

In the original study no attempt was made to model the correlogram, and the semi-variogram was not computed. In the light of the foregoing sections it is worth seeing how the modern geostatistical approach treats such periodicity. The conductivities in their logarithmic form are stationary, and so the sample semi-variogram is a mirror image of the co-variogram. It too is periodic, therefore, and some kind of angular function is needed as a model of it. Journel and Huijbregts (1978) suggest some appropriate models and their limiting values to ensure that the semi-variograms are positive definite. In this instance an angular function alone is not sufficient because the semi-variogram clearly increases from a small value at lag 1 to much larger values, reminiscent of a sill about which it thereafter fluctuates.

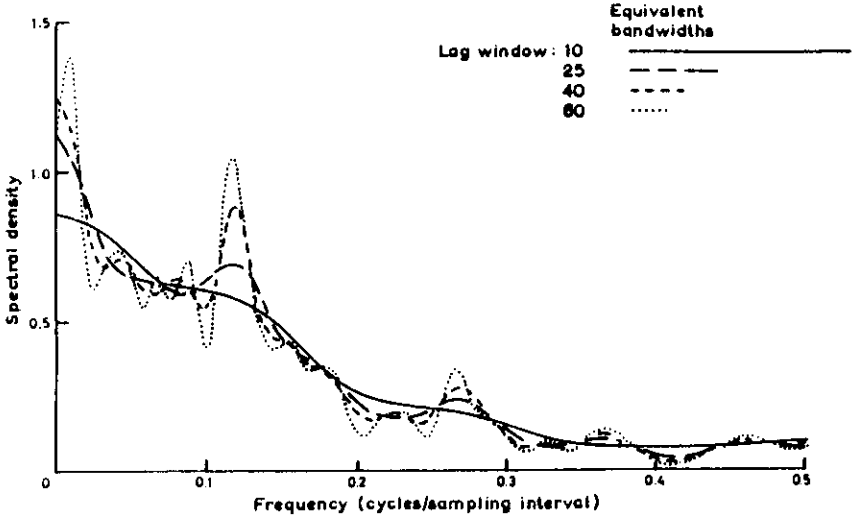

Figure 26. The spectrum of electrical conductivity of the soil at 30 to 40 cm on the Bland Plain computed with Bartlett lag windows of four widths (see Webster, 1977).

This additional component seemed best modeled by an exponential term. The best fitting model finally chosen was

$$\gamma(h) = c + b \exp(-h/a) + u \sin\left(\frac{2\pi h}{l}\right) + v \cos\left(\frac{2\pi h}{l}\right) \text{ for } h > 0$$
$$\gamma(0) = 0, \qquad [79]$$

with its six parameters estimated by

$$\hat{a} = 0.723 \qquad \hat{u} = 0.00937$$
$$b = 0.220 \qquad \hat{v} = 0.00833$$
$$\hat{c} = 0.188 \qquad l = 8.89.$$

Figure 27 shows the sample semi-variances and the fitted function. Note especially the estimated wavelength, l, of 8.89 sampling intervals, which is sensibly the same as that found by spectral analysis.

In geostatistical parlance any semi-variogram whose increase is not monotonic is said to exhibit a "hole effect." This signifies some kind of repetition in the regional distribution of the variable. It was evident in Webster and Cuanalo's (1975) original correlograms of the soil across the Jurassic limestones, sandstones, and clays of north Oxfordshire when extended to lags of 600 m. McBratney and Webster (1981a) found similar hole effects in some semi-variograms, again to 600 m, of soil on the drift-covered Buchan Plateau in northeast Scotland. Where the periodicity is as strong as it is at Caragabal on the Bland Plain, however, a more

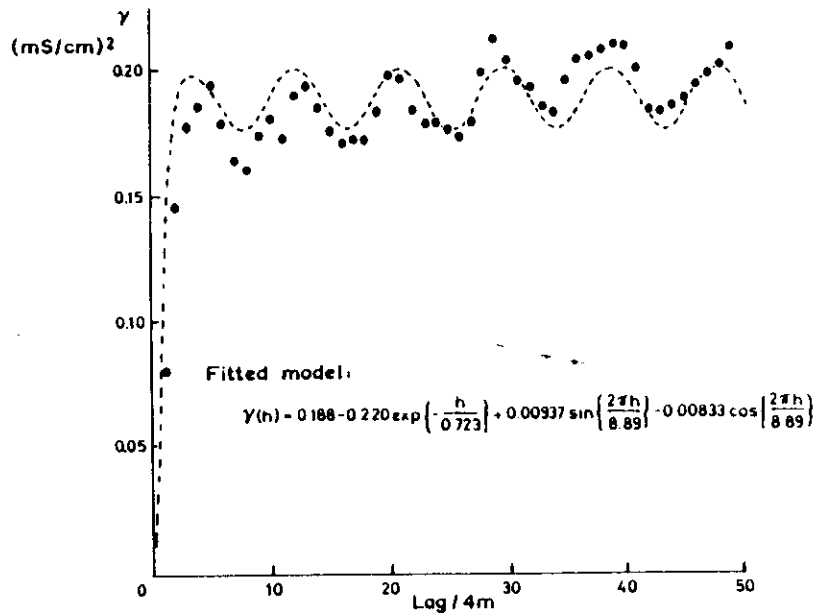

Figure 27. The semi-variogram of electrical conductivity of the soil at 30 to 40 cm on the Bland Plain with a periodic model fitted.

connotative term is desirable, and I suggest that such semi-variograms be described as periodic.

X. Optimal Estimation—Kriging

One of the prime reasons for obtaining a semi-variogram is to use it for estimation, and no review of modern spatial analysis would be complete without some account of how this is done. In soil survey we can recognize two main kinds of estimate. In the one we may wish to estimate the average value of a soil property within some defined region. The other is interpolation: We may wish to predict the values of a property at places that we have not visited. If such places are few we could go there and measure them, of course. In principle, however, they are infinite, and we may wish to predict the values at any or all of them and to make an isarithmic ("contour") map of the property.

If the region of interest is sufficiently uniform then the mean of an unbiased sample estimates both the population mean and the value at any one place. Usually it is not, and in these circumstances surveyors classify the soil so that they can use class means as predictors for individual sites.

Their model is that the value of a property z at a place i in class j is the sum of three terms:

$$z_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \qquad [80]$$

where μ is the general mean of the property in the whole region, α_j is the difference between the general mean and the mean of class j, and ε_{ij} is a random component with zero mean and variance σ_w^2, usually assumed to be normally distributed. This is the model underlying the sampling studies by Thornburn and co-workers in Illinois (Thornburn and Larsen, 1959; Morse and Thornburn, 1961) and the research on spatial prediction by Beckett's group at Oxford (e.g., Webster and Beckett, 1968; Beckett and Burrough, 1971).

Following classical procedure class means are estimated by sampling each class and computing

$$\bar{z}_j = \hat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} z_{ij}. \qquad [81]$$

Associated with $\hat{\mu}_j$ is its standard error or the square, the estimation variance. The latter is defined as

$$\sigma_{E_j}^2 = E[(\mu_j - \hat{\mu}_j)^2] = \sigma_{w_j}^2/n, \qquad [82]$$

and estimated by

$$\hat{\sigma}_{E_j}^2 = s_{E_j}^2 = \frac{1}{n_j(n_j - 1)} \sum_{i=1}^{n_j} (z_{ij} - \bar{z}_j)^2. \qquad [83]$$

If the soil has been classified at the same categoric level throughout the region, then the variances within all the classes should be the same, $\sigma_{w_j}^2 = \sigma_w^2$.

The pooled within-class variance should therefore be well estimated from a large sample.

The estimation variance for predicting a point, assuming the class is known, is the within-class variance plus the estimation variance for the mean of that class, thus:

$$\sigma_{E_p}^2 = \sigma_w^2 + \sigma_w^2/n_j. \qquad [84]$$

We note that in classical procedure the errors associated with prediction or interpolation are determined very largely by the size of the within-class variance: $\sigma_{E_p}^2$ cannot be less than σ_w^2, no matter how intensive the sampling. The quality of the classification is paramount, and this is the statistical reason for devoting so much effort to creating a good classification in traditional practice.

The dependence on classification for improving prediction and the inability to take into account either the intensity of sampling or its configuration are serious weaknesses of traditional practice. Regionalized variable theory provides an alternative, and we devote the remainder of this review to it. It is a large subject, and we can examine only the broad principles and some of its simpler forms.

A. Kriging Defined

The method of estimation embodied in regionalized variable theory is known in earth sciences as *kriging*, after D.G. Krige, who devised it empirically for use in the South African goldfields (see Krige, 1966). It is essentially a means of *weighted local averaging* in which the weights are chosen so as to give unbiased estimates while at the same time minimizing the estimation variance. Kriging is in this sense optimal.

Consider a typical situation in which we have measured a property at a number of places, $\mathbf{x}_i$, within a region to give values $z(\mathbf{x}_i)$, $i = 1, 2 \ldots n$. From these we wish to estimate the value of the property at some place B. The place might be a "point," that is, an area of the same size and shape as those on which the measurements were made. Alternatively it might be a larger area or block. The procedures for estimation in the two instances are known as *punctual kriging* and *block kriging*, respectively. Punctual kriging can be regarded as a special case of block kriging, and we shall treat it so here. In either event our estimate at B is the linear sum

$$\hat{z}(B) = \lambda_1 z(\mathbf{x}_1) + \lambda_2 z(\mathbf{x}_2) + \ldots + \lambda_n z(\mathbf{x}_n), \qquad [85]$$

where the λ_i are the weights associated with the sampling points.

We want our estimate to be unbiased; i.e.,

$$E[(z(B) - \hat{z}(B)] = 0, \qquad [86]$$

and this is assured if the weights sum to 1:

$$\sum_{i=1}^{n} \lambda_i = 1. \qquad [87]$$

The estimation variance at B is the expected square difference between our estimate and the true value, and is

$$\sigma_E^2(B) = E[[z(B) - \hat{z}(B)]^2] = 2 \sum_{i=1}^{n} \lambda_i \bar{\gamma}(\mathbf{x}_i, B) - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j) - \bar{\gamma}(B, B). \qquad [88]$$

Here $\gamma(\mathbf{x}_i, \mathbf{x}_j)$ is the semi-variance of the property between $\mathbf{x}_i$ and $\mathbf{x}_j$, taking account of both the distance, $\mathbf{x}_i - \mathbf{x}_j$, separating them and the angle, $\bar{\gamma}(\mathbf{x}_i, B)$ is the average semi-variance between $\mathbf{x}_i$ and all points within the block, and $\bar{\gamma}(B, B)$ is the average semi-variance within the block; i.e., the within-block variance defined as in equation [48] earlier. In punctual kriging this last

term is zero—there is no variance at a point—and the quantity $\bar{\gamma}(x_i, B)$ is just the semi-variance between x_i and the point to be estimated.

The estimation variance is minimized, subject to the constraint that $\lambda_i = 1$, by finding the partial derivatives with respect to each λ_i, and introduces a Lagrange multiplier, ψ. The minimum is obtained when

$$\sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) + \psi = \bar{\gamma}(x_j, B) \qquad \text{for } j = 1, 2, \ldots n. \qquad [89]$$

There are thus n equations in n unknowns plus a set of unbias conditions from which to determine ψ. These are best represented in matrix form by

$$\mathbf{A}\begin{bmatrix} \lambda \\ \psi \end{bmatrix} = \mathbf{b}, \qquad [90]$$

where

$$\mathbf{A} = \begin{bmatrix} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \ldots & \gamma(x_1, x_n) & 1 \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \ldots & \gamma(x_2, x_n) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(x_n, x_1) & \gamma(x_n, x_2) & \ldots & \gamma(x_n, x_n) & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{bmatrix},$$

$$\begin{bmatrix} \lambda \\ \psi \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \psi \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} \gamma(x_1, B) \\ \gamma(x_2, B) \\ \vdots \\ \gamma(x_n, B) \\ 1 \end{bmatrix}. \qquad [91]$$

Having solved equation [90] to find the weights, the estimation variance is obtained from

$$\sigma_E^2 = \mathbf{b}^T \begin{bmatrix} \lambda \\ \psi \end{bmatrix} - \bar{\gamma}(B, B). \qquad [92]$$

Kriging thus provides not only unbiased estimates of minimum variance, but also a measure of the estimation variance. In this respect it is superior to other methods of interpolation.

B. Example

Applying these equations to the stone content on Plas Gogerddan will reveal several interesting features of kriging. The semi-variances are taken from equation [44].

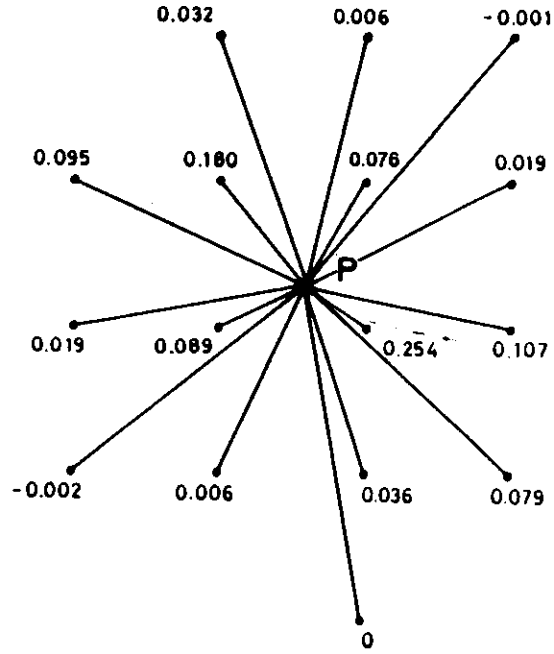

Figure 28. Weights for kriging stone content at a point, *P*, at Plas Gogerddan.

Suppose we wish to estimate the stone content at a point *P*. Figure 28 shows *P* lying one-third of a sampling interval from the bottom and right-hand side of a grid cell. The sampling points are at the nodes of the grid, and the nearest 16 are shown. The values at the nodes are the weights. The nearest is much the largest, and the ones in the outer shell are much smaller in absolute value. Notice especially that those in the bottom left and top right corners are so small as to be negligible. The figure also shows the effect of anisotropy. Variation is greatest approximately in the direction of these points and least in the perpendicular direction. As a result points lying in the latter direction carry more weight than those in the former for the same distance.

The particular weights depend of course on the semi-variogram, especially the size of the nugget variance, and on the configuration of sampling points and place to be predicted. Nevertheless, the fact that near data carry much weight and far data none of consequence is quite general and very important. It means that kriging is essentially local, and conforms to our intuitive notion of a sensible estimation procedure. It also means that for kriging the semi-variogram needs to be well estimated only to the lag of the furthest point to carry effective weight. In many instances the semi-variogram will be linear to this lag, and so fitting a model is easy (see Figure 13). Of course, there must be ample data within the range of the semi-

variogram if it is transitive, since any beyond it are spatially independent of the point being estimated. Local estimation, or interpolation, is not worth attempting with spatially independent data.

The local nature of kriging is computationally important, for it means that we rarely need more than 16 observations from which to krige one estimate, and matrix **A** need not be larger than 17 × 17. This does not matter when kriging just one point or block, but to make a map might require thousands of matrix inversions if the data are irregularly scattered. The size of the matrix is then very serious.

Applying the procedure to the whole of the field at Plas Gogerddan gives the results shown in Figures 29 to 33. Figure 29 is an isarithmic map drawn by interpolating a 3 × 3 square grid of points within each original grid cell and then threading contour lines through the grid. Notice in particular the discontinuity around many of the original grid nodes. These are due to the nugget effect and the minimum variance criterion. The latter ensures that the interpolated value at a sampling point is the observed value there. The presence of a nugget variance means that the semi-variogram is effectively composed of two distinct functions, one, $\gamma_1(h, \theta) = w(\theta)\, h$, describing the spatial dependence, and the other the purely random function, $\gamma_0 = c_0$, that represents the discontinuity. Its effect appears even more dramatically in Figure 30, a block diagram of the interpolated surface.

A user of a map will often want to know not about points but about small blocks, of, say, 1 to 10 ha. In these circumstances a block kriged map should be made. Figure 31 shows such a map made from the same data. Each block was 30.5 × 30.5 m, and again there were nine estimates for each original grid cell on a 5.1-m mesh. The result is smoother than the punctually kriged map. Note, however, that the aim is not primarily cosmetic: it is to provide the best estimates for 30.5 × 30.5-m blocks.

As mentioned above, kriging provides estimates of the estimation variances. These themselves can be displayed in the form of a reliability map. Where both sampling and interpolation are on regular grids, as here, the estimation variances have a regular pattern that is repeated. Figure 32 gives the estimation variances for grid cells in the central part of the map, (a) for punctual kriging and (b) for block kriging. A block diagram of the block kriging variance for the whole field, Figure 33, is interesting in emphasizing the way in which the error increases sharply near the edge of the field, beyond which there are no data. The small bump near the bottom is caused by a missing value.

The above is just one example to illustrate the power and reliability of kriging for local estimation and mapping. There are now a number of examples in the literature of soil science (Burgess and Webster, 1980a, 1980b; McBratney *et al.*, 1982; Yost *et al.*, 1982; Vieira *et al.*, 1981; McBratney and Webster, 1983).

Before leaving the topic we should be clear that a kriged map of a soil property can be a misleading representation of reality in one respect. Even a

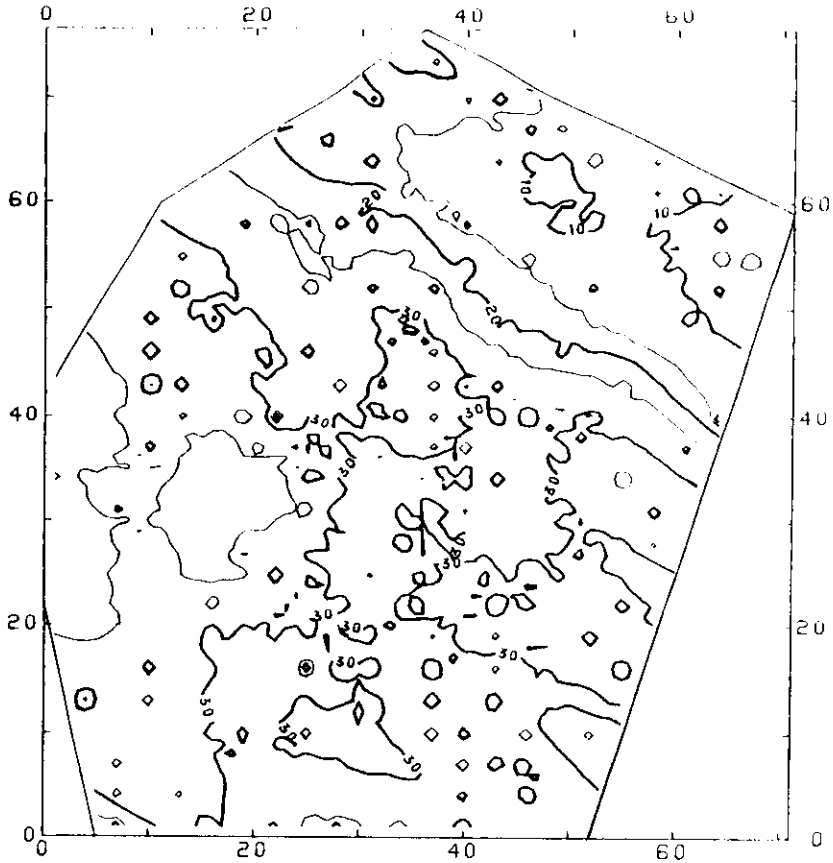

Figure 29. Isarithmic map of stone content made by punctual kriging at Plas Gogerddan.

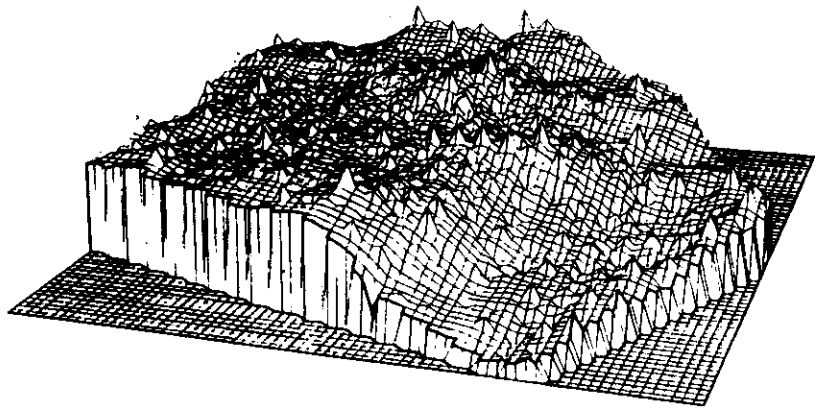

Figure 30. Block diagram of stone content estimated by punctual kriging.

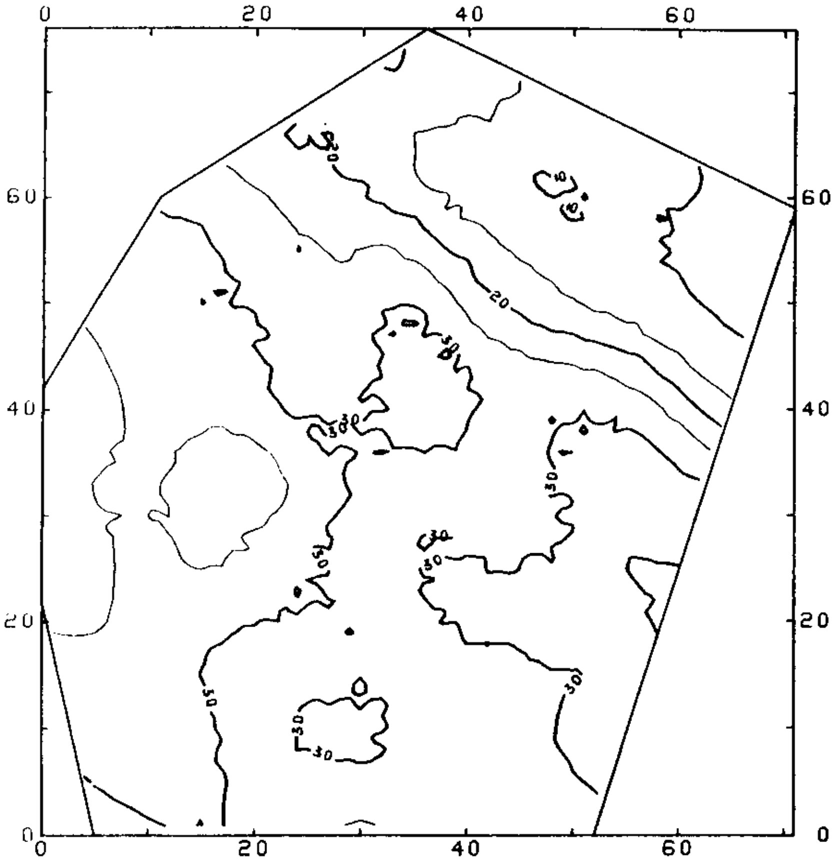

Figure 31. Isarithmic map of stone content made by block kriging.

punctually kriged map shows smooth variation between sampling points. Yet if there is a nugget variance, then we know that the soil does not vary smoothly at our working scale. If we wish to represent the soil with all its variation from sample data then we should simulate it in a way that retains all the characteristics of the semi-variogram. Kriging, by averaging, does not do this. Readers who wish to follow the matter further should consult Journel and Huijbregts (1978).

XI. Designing Sampling Schemes

A second way in which the results of spatial analysis can be used is in designing optimal-sampling schemes for surveys of soil properties. This potential application of regionalized variables theory to soil survey is due

Punctual kriging

0	12.89	12.89	0
12.89	12.87	12.87	12.89
12.89	12.87	12.87	12.89
0	12.89	12.89	0

Block kriging

0.929	0.935	0.935	0.929
0.935	0.940	0.940	0.935
0.935	0.940	0.940	0.953
0.929	0.935	0.935	0.929

Figure 32. Variances for punctual and block kriging of stone content. Values are in (percent)². For punctual kriging the block discs are sampling points and open circles the interpolated ones. In the lower figure they represent the centers of the estimated blocks.

largely to Burgess *et al.* (1981) and McBratney *et al.* (1981), and this account is taken largely from their work.

To understand the basis of this application consider equations [88] and [92] for the kriging variance. We can see from these that the errors of estimation depend on the semi-variogram and through it on the configuration of sampling points in relation to the block to be estimated. They do not depend on the observed values themselves. So, if the semi-variogram is known then the kriging variances can be estimated before performing the sampling.

The kriging variance is not constant from place to place as Figure 32 shows. Rather, it tends to increase the further the estimated point or block is from the data, at least where the semi-variogram is a monotonic increasing function, which is usually so for soil. The kriging variance can be diminished on average by sampling more intensively. For any given

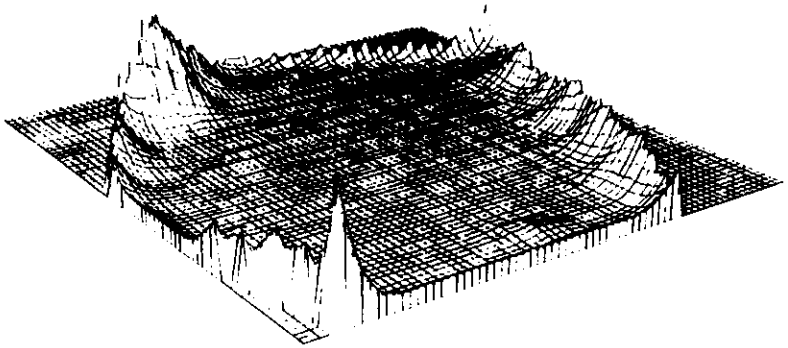

Figure 33. Diagram of the block kriging variance.

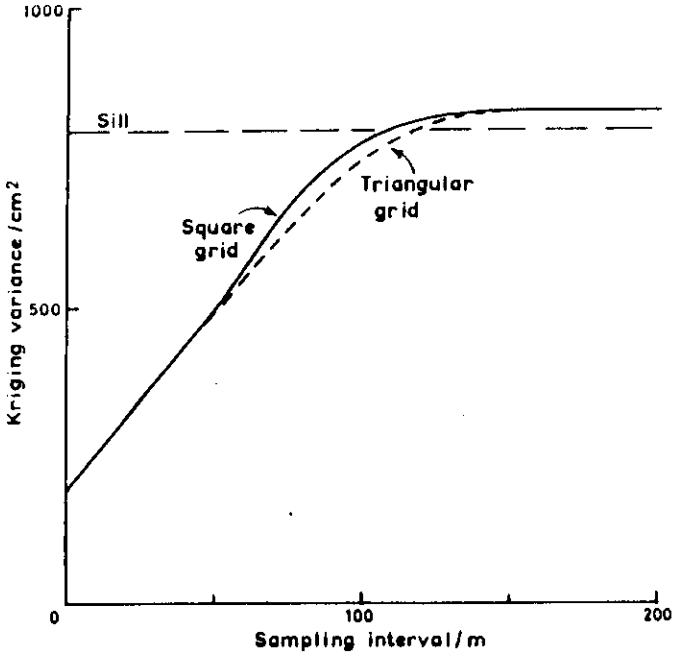

Figure 34. Graphs of maximum variance for punctual kriging of thickness of cover loam against sample spacing at Hole Farm, Norfolk, on square and triangular grids.

intensity, however, the maximum kriging variance will still occur where the data are furthest from the estimated point. One criterion of a good sampling scheme is that it minimizes the maximum kriging variance, say, $\sigma^2_{k_{max}}$. If variation is isotropic then it can be achieved by sampling on a regular equilateral triangular grid. For unit sampling density, the side of the triangle is 1.0746 and the maximum distance, d_{max}, between a sampling point and a point to be estimated, at the center of a triangle, is 0.6304. Any other sampling scheme, especially random and other irregular schemes, will have some larger values of d_{max} and therefore larger maximum kriging variances. Nevertheless, a square grid with $d_{max} = 1/\sqrt{2} = 0.7071$ for unit sampling density will generally give $\sigma^2_{k_{max}}$ only a little larger than than that for the triangular scheme, and for reasons of convenience in indexing computer management, site location and logistics will usually be preferred.

In planning a survey this information may be used in one of two ways. If the budget for the survey is fixed then so is the sampling intensity. The maximum estimation variance is determined by solving equation [92]. For punctual estimates it is solved just once for the centers of the grid cells. For block estimates, however, $\sigma^2_{k_{max}}$ can occur where the block is centered over a grid node: It depends on the size of the blocks in relation to the grid mesh. The equation must therefore be solved twice to find the maximum.

If, however, funding depends on the precision required, then a maximum tolerable error may be specified for the survey. In such a situation the aim is to determine the sampling intensity that provides just the required precision. The aim is achieved by solving equation [92] for a range of spacings. A graph of $\sigma^2_{k_{max}}$ against sample spacing is plotted and the required spacing read from the graph. The survey can then be performed most economically by sampling at this intensity.

An example from Burgess *et al.* (1981) illustrates the procedure where variation is isotropic. It derives from a study at Hole Farm in eastern England where crop performance had been found to depend on the thickness of cover loam. In one field of 18 ha the thickness had been measured by boring on a 20 m square grid to give some 450 observations. Semi-variances were calculated to 300 m and an isotropic spherical model, equation [25], with the following coefficients fitted:

$$\begin{aligned} \text{nugget variance } c &= 187.0 \text{ cm}^2 \\ \text{sill-nugget } c &= 603.8 \text{ cm}^2 \\ \text{range } a &= 101.2 \text{ m.} \end{aligned}$$

Figure 34 shows the graph of maximum kriging variance for punctual estimates against sample spacing for square grids. For comparison the graph for the equilateral triangular configuration is also plotted, and shows that there could be a small gain in efficiency using it. Note that the kriging variance cannot be less than the nugget variance, however dense the sampling, and that it reaches a maximum somewhat greater than the sill value. This maximum is in fact the sill value plus the Lagrange multiplier, ψ,

of equation [90]. The spacing at which it is reached is $\sqrt{2}$ times the range of the semi-variogram for the square grid. At this spacing the interpolated point at the center of a grid cell becomes independent of the data. The semi-variances in equation [90] equal the sill and all weights are equal to $1/n$. The Lagrange multiplier takes the value σ^2/n and represents the additional uncertainty in estimating the value at a place from local data rather than the whole population, and in these circumstances equation [92] is equivalent to the classical equation [84].

Examples of graphs for block kriging from the same semi-variogram are shown in Figure 35 for two sizes of block. As above, two graphs can be drawn, one for blocks whose centers lie in the centers of the grid cells and the other for blocks centered over grid nodes. At most spacings the kriging variances are larger for the former, but there is a small range of spacings where they are the smaller of the two, and this is evident for the 100-m square blocks.

A reasonable maximum for the kriging variance is perhaps 100 cm^2, equivalent to a standard error of 10 cm. For 40 × 40-m blocks this gives a grid spacing of approximately 40 m and about 75 m for 100 × 100-m squares.

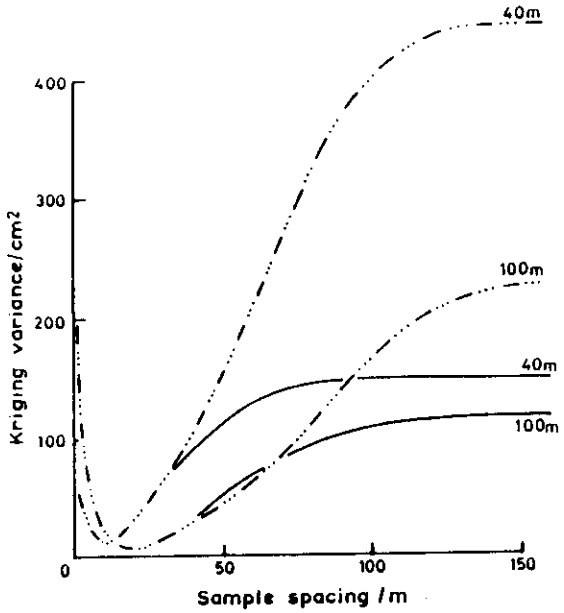

Figure 35. Graphs of maximum variance for kriging cover loam thickness over 40 × 40-m and 100 × 100-m blocks at Hole Farm. The solid lines are for square grids with the blocks centered over the grid nodes, the broken lines are for blocks centered in the centers of the grid squares.

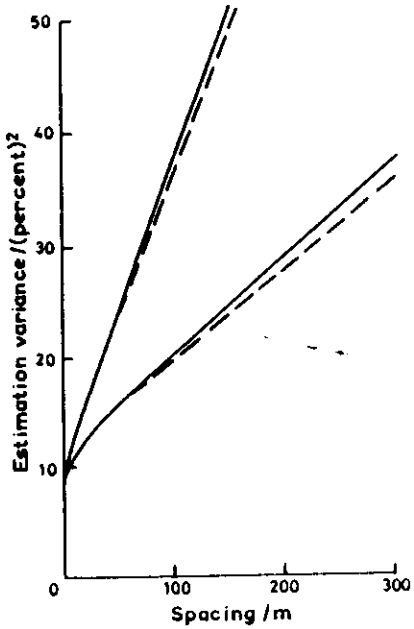

Figure 36. The relation between punctual kriging variance of stone content at Plas Gogerddan and sample spacing in the directions of maximum and minimum variation.

A. Effect of Anisotropy

Equilateral sampling grids are best where variation is isotropic, and there are many regions in which this is so. As above, however, the soil can vary anisotropically, and in these situations some modification is needed. If the anisotropy is geometric the modification is simple.

The problem of determining the optimal intensity is treated initially as though variation were isotropic with the semi-variogram for the direction, ϕ, in which the gradient or distance parameter is greatest. In the linear case this gives

$$\gamma(h) = c_0 + A|\mathbf{h}|, \qquad [93]$$

from equation [41]. The graph of $\sigma^2_{k_{min}}$ against intensity is drawn and the optimal intensity read from it. This gives the spacing, m, in that direction. This can be repeated for direction $\phi + \pi/2$. Figure 36 shows such graphs for punctual estimation for the stone content in Cae Ruel whose semi-variogram is given earlier. However, the second graph is not needed since the spacing for this direction is easily computed by multiplying m by the anisotropy ratio A/B.

Thus, the optimal strategy is to sample along parallel transects aligned in the direction of maximum variation. The spacing on the transects is m and that between them mA/B. It is the familiar pattern of soil survey.

If costs are fixed then the best strategy is still as above. The sampling intervals are now such that the product $m \times mA/B$ equals the reciprocal of the sampling intensity.

Acknowledgments

It is my pleasure to thank Dr T.M. Burgess and Dr A.B. McBratney whose collaboration produced so many of the results that I have quoted, and Mrs. Joyce Munden for much of the fresh computing and graphics. I also thank Dr P.A. Burrough for his advice on fractals and Dr H.E. Cuanalo for the Sandford data.

References

Armstrong, M. 1984. Improving the estimation and modelling of the variogram. In: G. Verly, M. David, A.G. Journel, and A. Marechal (eds.), *Geostatistics for natural resource characterization*. Reidel, Dordrecht. pp 1–19.

Armstrong, M., and R. Jabin. 1981. Variogram models must be positive-definite. *Math. Geol.* 13:455–459.

Bartlett, M.S. 1966. *An introduction to stochastic processes with special reference to methods and applications*. 2nd ed. Cambridge University Press.

Beckett, P.H.T., and P.A. Burrough. 1971. The relation between the cost and utility in soil survey. V. The cost-effectiveness of different soil survey procedures. *J. Soil Sci.* 22:481–489.

Beckett, P.H.T., and R. Webster. 1971. Soil variability: a review. *Soils Fert.* 34:1–15.

Berry, M., and Z.V. Lewis, 1980. On the Weierstrass-Mandelbrot fractal function. *Proceedings of the Royal Society* A 370:459–484.

Burgess, T.M., and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *J. Soil Sci.* 31:315–331.

Burgess, T.M., and R. Webster. 1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *J. Soil Sci.* 31:333–341.

Burgess, T.M., and R. Webster. 1984. Optimal sampling strategies for mapping soil types. I. Distribution of boundary spacings. *J. Soil Sci.* 35. 641–654.

Burgess, T.M., R. Webster, and A.B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *J. Soil Sci.* 32:643–659.

Burrough, P.A. 1983a. Multiscale sources of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation. *J. Soil Sci.* 34:577–597.

Burrough, P.A. 1983b. Multiscale sources of spatial variation in soil. II. A non-Brownian fractal model and its application in soil survey. *J. Soil Sci.* 34:599–560.

Burrough, P.A. In press. The application of fractal ideas to geophysical phenomena. *Bull. Inst. Math. Appl.*

Campbell, J.B. 1978. Spatial variation of sand content and pH within single contiguous delineations of two soil mapping units. *Soil Sci. Soc. Am. J.* 42:460–464.

Chatfield, C. 1980. *The analysis of time series*. Chapman and Hall, London.

Christakos, G. 1984. On the problem of permissible covariance and variogram models. *Water Resour. Res.* 20:251–265.

Cochran, W.G. 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Stat.* 17:164–177.

Cressie, N., and D.M. Hawkins. 1980. Robust estimation of the variogram. I. *Math. Geol.* 12:115–125.

Dalenius, T., J. Hajek, and S. Zubrzycki. 1961. On plane sampling and related geometrical problems. *Proc. 4th Berkeley Symp. Prob. Stat.* 1:125–150.

David, M. 1977. *Geostatistical ore reserve estimation*. Elsevier, Scientific Publishing Co., Amsterdam.

Delfiner, P. 1976. Linear estimation of non stationary spatial phenomena. In: M. Guarascio, M. David, and C. Huijbregts (eds.), *Advanced geostatistics in the mining industry*. Reidel, Dordrecht. pp. 49–68.

Dunn, M.R. 1983. A simple sufficient condition for a variogram model to yield positive definite variances under restrictions. *Math. Geol.* 15:553–564.

Gajem, Y.M., A.W. Warrick, and D.E. Myers. 1981. Spatial dependence of physical properties of a typic torrifluvent soil. *Soil Sci. Soc. Am. J.* 45:709–715.

Gandin, L.S. 1965. *Objective analysis of meteorological fields*. Israel Program for Scientific Translations, Jerusalem.

Harbaugh, J.W., and D.F. Merriam. 1968. *Computer applications in stratigraphic analysis*. John Wiley and Sons, New York.

Hawkins, D.M., and D.F. Merriam, 1974. Zonation of multivariate sequences of digitized geologic data. *Mathematical Geology* 6:263–269.

Jenkins, G.M., and D.G. Watts. 1968. *Spectral analysis and its applications*. Holden-Day, San Francisco.

Journel, A.G., and C.J. Huijbregts. 1978. *Mining geostatistics*. Academic Press, London.

Kendall, M.G., and A. Stuart. 1976. *The advanced theory of statistics. 3. Design and analysis, and time series*. 3rd ed. Griffin, London.

Kitanidis, P.K. 1983. Statistical estimation of polynominal generalized covariance functions and hydrological applications. *Water Resour. Res.* 19:909–921.

Kozlovskii, F.I., and N.P. Sorokina. 1976. The soil individual and elementary analysis of the soil-cover pattern. In: V.M. Fridland (ed.), *Soil combinations and their genesis*. Amerind Publishing Co., New Delhi. pp. 55–64. (Translated from Russian)

Krige, D.G. 1966. Two dimensional weighted moving average trend surfaces for ore-evaluation. *J. South African Inst. Min. Metal.* 66:13–38.

Mandelbrot, B.B. 1982. *The fractal geometry of nature*. W.H. Freeman, London.

Matérn, B. 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden* från Statens Skogsforskningsinstitut 49:1–144.

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, Fontainebleau*, No. 5.

McBratney, A.B., and R. Webster. 1981a. Spatial dependence and classification of the soil along a transect in northeast Scotland. *Geoderma* 26:63–82.

McBratney, A.B., and R. Webster. 1981b. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. II. Program and examples. *Comp. Geosci.* 7:335–365.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. *J. Soil Sci.* 34:137–162.

McBratney, A.B., R. Webster, and T.M. Burgess. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and method. *Comp. Geosci.* 7:331–334.

McBratney, A.B., R. Webster, R.G. McLaren, and R.B. Spiers. 1982. Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland. *Agronomie* 2:969–982.

Miesch, A.T. 1975. Variograms and variance components in geochemistry and ore evaluation. *Geol. Soc. Am. Mem.* 142:333–340.

Morse, R.K., and T.H. Thorburn. 1961. Reliability of soil maps. *Proc. 5th Int. Conf. Soil Mech. Found. Eng.* 1:259–262.

Nortcliff, S. 1978. Soil variability and reconnaissance soil mapping: a statistical study in Norfolk. *J. Soil Sci.* 29:403–418.

Oliver, M.A. 1984. Soil variation in the Wyre Forest: its elucidation and measurement. Ph.D. Thesis, Birmingham University.

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

Olea, R.A. 1977. *Measuring spatial dependence with semi-variograms*. Series on Spatial Analysis No. 3. Kansas Geological Survey, Lawrence.

Orey, S. 1970. Gaussian sample functions and the Hausdorff dimension of level crossings. *Zeitschrift für Wahrschein-lichkeitstheorie und Verwandte Gebiete* 15:249–256.

Quenouille, M.H. 1949. Problems in plane sampling. *Ann. Math. Stat.* 20:355–375.

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

Russo, D., and E. Bresler. 1982. Soil hydraulic properties as stochastic processes. II. Errors of estimates in a heterogeneous field. *Soil Sci. Soc. Am. J.* 46:20–26.

Sisson, J.B., and P.H. Wierenga. 1981. Spatial variability of steady-state infiltration rates as a stochastic process. *Soil Sci. Soc. Am. J.* 45:699–704.

Thornburn, T.H., and W.R. Larsen. 1959. A statistical study of soil sampling. *J. Soil Mech. Found. Div. Proc. Am. Soc. Civ. Eng.* 85, SMS, 1–13.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Neilsen. 1983. The use of cokriging with limited field soil observations. *Soil Sci. Soc. Am. J.* 47:175–184.

Vieira, S.R., D.R. Neilsen, and J.W. Biggar. 1981. Spatial variability of field-measured infiltration rate. *Soil Sci. Soc. Am. J.* 45:1040–1048.

Walker, P.H., G.F. Hall, and R. Protz. 1968. Soil trends and variability across selected landscapes in Iowa. *Soil Sci. Soc. Am. Proc.* 32:97–101.

Webster, R. 1973. Automatic soil-boundary location from transect data. *Math. Geol.* 5:27–37.

Webster, R. 1977. Spectral analysis of gilgai soil. *Austr. J. Soil Res.* 15:191–204.

Webster, R. 1978. Optimally partitioning soil transects. *J. Soil Sci.* 29:388–402.

Webster, R. 1981. Experience of kriging from field measurements of soil properties. In: M.-C. Girard (ed.), *Traitement informatique des données de sol*. Tome 1. Institut National Agronomique, Paris-Grignon. pp. 101–109.

Webster, R., and P.H.T. Beckett. 1968. Quality and usefulness of soil maps. *Nature* (London) 219:680–682.

Webster, R., and T.M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. *J. Soil Sci.* 31:505–524.

Webster, R., and T.M. Burgess. 1984a. Une approche probabiliste de la cartographie du sol. *Sciences de la Terre, Série Informatique Géologique* 18:175–185.

Webster, R., and T.M. Burgess. 1984b. Sampling and bulking strategies for estimating soil properties of small regions. *J. Soil Sci.* 35:127–140.

Webster, R., and B.E. Butler. 1976. Soil classification and survey studies at Ginninderra. *Austr. J. Soil Res.* 14:1–24.

Webster, R., and H.E. Cuanalo de la C. 1975. Soil transect correlograms of north Oxfordshire and their interpretation. *J. Soil Sci.* 26:176–194.

Webster, R., and S. Nortcliff. 1984. Improved estimation of micro nutrients in hectare plots of the Sonning series. *J. Soil Sci.* 35:667–672.

Xu Jiyan, and R. Webster. 1984. A geostatistical study of topsoil properties in Zhangwu County, China. *Catena* 11:13–26.

Yates, F. 1948. Systematic sampling. *Phil. Trans. Roy. Soc. Lond.* A 241:345–377.

Yost, R.S., G. Uehara, and R.L. Fox. 1982. Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. *Soil Sci. Soc. Am. J.* 46:1028–1032.

Youden, W.J., and A. Mehlich. 1937. Selection of efficient methods for soil sampling. *Contr. Boyce Thompson Inst. Plant Res.* 9:59–70.

Zubrzycki, S. 1957. On estimating gangue parameters. *Zastosowania Matematakyi* 3:105–153. (In Polish)