Geostatistics in physical geography. Part I: theory

MARGARET OLIVER*, RICHARD WEBSTER† and JOHN GERRARD‡

**Honorary Research Fellow, ‡Senior Lecturer in Geography, School of Geography, University of Birmingham, P. O. Box 363, Birmingham B15 2TT*

†Senior Principal Scientific Officer, Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ

Revised MS received 19 April, 1989

ABSTRACT

Geostatistics grew from meteorologists' desire to interpolate weather characteristics from sparse data, and from mining engineers who wanted to estimate quantities of minerals in bodies of rock from drill cores. In both cases the properties of interest behaved as spatially correlated random variables. Practice in mining is now underpinned by sound theory, the theory of regionalized variables. It is widely recognized that the theory is properly applicable in many other branches of earth, atmospheric and marine science.

Part I of this paper reviews the more elementary aspects of the theory and its assumptions. It describes the method of local estimation embodied in regionalized variable theory known as kriging. The central tool of geostatistics is the variogram, which expresses quantitatively and succinctly spatially correlated variation, and its estimation is described. The more common forms of variogram are illustrated, and authorized models for them are listed. There are sound procedures for choosing and fitting models to variograms now using well-tried computer programs such as MLP and Genstat.

Regionalized variable theory provides geographers with a concise and coherent methodology for describing and analysing spatially distributed data.

KEY WORDS: Autocorrelation, Estimation, Geostatistics, Kriging, Modelling, Regionalized variable theory, Variogram

INTRODUCTION

For many years geographers had to be content with the qualitative description of the variation of continuous variables over the earth's surface and to display it on maps. On these maps regions of interest were divided into areas with a characteristic type of rock, soil, vegetation, climate, or other attribute of concern. An exception in geography was topographic mapping: this represents a contrasting quantitative tradition that was followed later in meteorology.

Geographers realized that they could not measure all the attributes that interested them with the same resolution as topographic height to produce accurate descriptions and maps. They usually had to work from fragmentary information and to use it sensibly to estimate and predict the properties at intermediate positions with confidence.

Many techniques have been devised for interpolation and mapping. Most are quite empirical and many are intuitively reasonable, but they lack intellectual rigour. One approach designed to put interpolation on a sounder theoretical basis has been to treat geographic attributes as mathematical variables that depend on their positions on or above the earth's surface; i.e. to treat their values as functions of their spatial coordinates. Thus geographers envisaged a model of the kind:

$$z(x,y) = f(x,y) + \varepsilon, \quad (1)$$

where $z(x,y)$ is the value of a variable z at a place whose coordinates in two dimensions are x and y, and f denotes some deterministic function. The quantity ε is an error term embracing random fluctuation plus other residual information not described by the function. Polynomial functions such as

$$z(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy + \ldots + \varepsilon, \quad (2)$$

have been popular. They are linear in their parameters, the spatial coordinates can usually be determined without serious error, and the equations can be fitted by standard regression techniques to produce 'trend surfaces'. This has the disadvantage that the deterministic element of the trend is not always present and if so it might not be very large. As it happens the random component is usually the larger of the two, and is often so large as to mask any deterministic variation. The natural properties of the earth's surface seem to behave as essentially random variables, albeit spatially dependent, rather than as mathematical ones. There appeared to be no general theory and methodology that were applicable to them.

Similar problems of estimating and mapping properties arose in meteorology (Gandin, 1965) and in mining (Matheron, 1965) where the concentrations of minerals and the thickness of ore bodies vary in space. Gandin (1965) describes the application of optimum interpolation, developed by A. N. Kolmogorov as early as 1941, for estimating the values of atmospheric pressure and rainfall at sites between the recording stations. The need for solutions was more pressing in mining because of the enormous costs incurred, and it was in mining that the advance in spatial analysis was made. Matheron (1965, 1971) brought together a number of isolated results in spatial statistics (Kolmogorov, 1941; Krige, 1951; Matérn, 1960; Yaglom, 1962) into a coherent body of theory, the theory of regionalized variables. This theory describes comprehensively and quantitatively the kind of variation that is characteristic of geological deposits and many other properties of the earth's surface. All can be treated as spatially dependent random variables. Geostatistics is largely the application of this theory to practical problems.

In Part I of this paper we summarize those aspects of regionalized variable theory that are most likely to be useful to physical geographers and environmental scientists. Part II (Oliver *et al.*, 1989) brings together examples derived from several fields of study to illustrate their application.

Applications

Regionalized variable theory was developed largely by Matheron at the Paris School of Mines, and it is now applied widely in mining (e.g., Guarascio *et al.*, 1976; David, 1977; Journel and Huijbregts, 1978; Verly *et al.*, 1984) for estimating the concentrations of minerals in ore bodies and recoverable reserves, and in planning operations. Geostatistical methods are applicable throughout the earth sciences, especially where information is fragmentary and there is a need to maximize its use. Examples of such applications include the mapping and modelling of ground water (Gambolati and Volpi, 1979; Kitanidis and Vomvoris, 1983), rainfall monitoring (McCullagh, 1975), and the distribution of atmospheric pollutants (Lajaunie, 1984). In soil science the methods have been used to estimate nutrients and other soil constituents at unvisited sites and over larger areas (Burgess and Webster, 1980; McBratney *et al.*, 1982; Yost *et al.*, 1982a and b; Webster and McBratney, 1987), to improve the efficiency of sampling (Burgess *et al.*, 1981; Webster and Burgess, 1984; Oliver and Webster, 1986a), and to rationalize spatial classification (Wackernagel *et al.*, 1988; Oliver and Webster, 1989). Geostatistical methods can be used to explore the processes responsible for variation. Moffat *et al.* (1986) used geostatistics to determine the structure of the Chalk and Tertiary surfaces in the Chiltern Hills, and Yost *et al.* (1982a) and Oliver and Webster (1986b) to identify the causes of spatial variation in soil properties.

Geostatistical procedures are also applicable where there is a complete cover of information. For instance, stereo plotters and satellite sensors can produce unlimited digital data that need to be sampled for storage, analysis and comparison with data from other sources (Atkinson *et al.*, forthcoming). Sampling should be efficient whether the cover is complete or fragmentary, and this can be determined by a preliminary spatial analysis.

Physical geographers can encounter both situations described above, and so geostatistics is potentially very valuable to them. They can use it, for example, to estimate the values of properties at unsampled locations and over larger areas, determine the spatial scale of variation, plan efficient sampling, and determine the structure or pattern in particular variables to suggest likely causes of the variation.

THEORY

The general statistical approach to prediction embodied in regionalized variable theory combines a deterministic component, such as that of trend surface analysis, with a stochastic one, so that the spatial variation in an attribute is expressed by

$$z(\mathbf{x}) = \Sigma\, a_k f_k(\mathbf{x}) + \varepsilon(\mathbf{x}), \qquad (3)$$

where $\mathbf{x}$ denotes the spatial coordinates in one, two or three dimensions, the f_k, $k = 0, 1, \ldots$, are functions

of the spatial position, the a_k are unknown coefficients, and $\varepsilon(\mathbf{x})$ is a random component that is itself spatially dependent. Thus, the first term on the right-hand side of equation (3) represents the deterministic element of the variation, and the stochastic element is embodied in the second. As mentioned earlier, earth scientists have discovered empirically that the stochastic component is by far the larger in most instances. So for practical purposes all the variation can be represented by the second term in equation (3), and the first can be replaced by a constant to give

$$z(\mathbf{x}) = \mu_v + \varepsilon(\mathbf{x}), \tag{4}$$

where μ_v is the mean, and the quantity $\varepsilon(\mathbf{x})$ is the spatially dependent random component defined as follows. It has a mean of zero,

$$\mathrm{E}[\varepsilon(\mathbf{x})] = 0, \tag{5}$$

and a variance defined by

$$\mathrm{var}[\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}+\mathbf{h})] = \mathrm{E}[\{\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}+\mathbf{h})\}^2] = 2\gamma(\mathbf{h}), \tag{6}$$

where $\mathbf{h}$ is a vector, the *lag*, that separates the two places $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ in both distance and direction. Thus the variance of $\varepsilon(\mathbf{x})$ depends on the separation $\mathbf{h}$ and not on the actual position of $\mathbf{x}$. Matheron realized that with a constant mean equation (6) was equivalent to

$$\mathrm{var}[z(\mathbf{x}) - z(\mathbf{x}+\mathbf{h})] = \mathrm{E}[\{z(\mathbf{x}) - z(\mathbf{x}+\mathbf{h})\}^2] = 2\gamma(\mathbf{h}). \tag{7}$$

These assumptions that the mean and the variance of the differences are both stationary constitute Matheron's *Intrinsic Hypothesis*. The quantity $\gamma(\mathbf{h})$ is known as the *semi-variance*: it is half the expected squared difference between two values. As above, it depends on $\mathbf{h}$, and the function that relates γ to $\mathbf{h}$ is the *semi-variogram* or increasingly just the *variogram*. Where the intrinsic hypothesis holds, the variogram contains all the information about the spatial variation of the attribute of interest. Furthermore it enables the semi-variances to be estimated from a sample of a single realization of the underlying process.

Where a variable is second-order stationary, i.e. where both the mean and variance are constant, the semi-variance is equivalent to the auto-covariance of time-series analysis. The covariance at lag $\mathbf{h}$ is

$$C(\mathbf{h}) = \mathrm{E}[\{z(\mathbf{x}) - \mu\}\{z(\mathbf{x}+\mathbf{h}) - \mu\}], \tag{8}$$

where μ is the mean of the attribute. The semi-variance is then

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}), \tag{9}$$

where $C(0)$ is the covariance at zero lag, or the *a priori* variance of the process. The autocorrelation coefficient, used in earth sciences by, for example, Nieuwenhuis and van den Berg (1971), Thornes (1973), and Webster and Cuanalo (1975) is closely related:

$$\begin{aligned}\rho(\mathbf{h}) &= C(\mathbf{h})/C(0) \\ &= 1 - \{\gamma(\mathbf{h})/C(0)\}.\end{aligned} \tag{10}$$

In many instances the variance appears to increase without limit as the lag distance increases. There is no finite covariance then, and equations (8), (9) and (10) do not apply. The covariance and autocorrelation functions cannot be used. The variogram, however, still exists, and because of its weaker underlying assumptions it is useful in a wider range of situations. In particular, it can be used with greater confidence for reconnaissance when little or nothing is known beforehand.

In large regions the mean values of variables will vary from one part to another. A variable will usually be locally stationary within some neighbourhood v, however. The variable is then said to be *quasi-intrinsic*, and it is for this reason that the subscript V is inserted in equation (4). In practice it means that the intrinsic hypothesis can be assumed and the variogram used to describe the variation within limited neighbourhoods. This is frequently all that is required to estimate or interpolate the variable at unvisited sites satisfactorily.

KRIGING

One of the most important uses of regionalized variable theory is for local estimation by the method known as kriging. D. G. Krige (1951, 1966) developed the method empirically for estimating amounts of gold in bodies of rock from fragmentary information in the mines of South Africa. Kolmogorov's (1941) method of optimum interpolation is, however, the first recognizable formulation of kriging. Kriging is a general term that embraces several estimation procedures (Krige *et al.*, 1989). What makes kriging unique and highly commendable compared with other methods of estimation is that its estimates are unbiased and have minimum variances. In this sense it is optimal. Furthermore the estimation variances themselves can be estimated, and so the technique can be used with known confidence. Kriging is also an exact interpolator, i.e. the kriged value at a sampling point is the measured value there and the variance is zero. Laslett *et al.* (1987) compared kriging with other

techniques of interpolation and showed that kriging was the only one that performed reliably in all circumstances.

At its simplest kriging is a method of weighted averaging of the observed values of a property Z within a neighbourhood, V, from the measured values $z(\mathbf{x}_i)$ of the property at n sites, $\mathbf{x}_i$, $i = 1,2 \ldots, n$. Estimates can be made over a block of land or in a body of rock B by

$$\hat{z}(B) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i), \tag{11}$$

where λ_i are weights associated with the sampling points. To ensure that the estimates are unbiased the weights λ_i sum to 1:

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{12}$$

The estimation variance for $\hat{z}(B)$ is given by

$$\sigma^2(B) = \mathrm{E}[\{\hat{z}(B) - z(B)\}^2] = 2\sum_{i=1}^{n} \lambda_i \bar{\gamma}(\mathbf{x}_i, B) - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j) - \bar{\gamma}(B,B), \tag{13}$$

where $\gamma(\mathbf{x}_i, \mathbf{x}_j)$ is the semi-variance between the ith and jth sampling points, $\bar{\gamma}(\mathbf{x}_i, B)$ is the average semi-variance between the block B and the ith sampling point, and $\bar{\gamma}(B,B)$ is the average semi-variance within the block B, i.e. the within-block variance. The estimation variance is minimized subject to condition (12) when

$$\sum_{i=1}^{n} \lambda_i \gamma(\mathbf{x}_i, \mathbf{x}_j) + \psi = \bar{\gamma}(\mathbf{x}_j, B) \text{ for all } j, \tag{14}$$

which introduces a Lagrange multiplier, ψ, to achieve minimization. The weights are found by solving these kriging equations, and the estimate is then obtained by inserting the weights into equation (11). The estimation variance or kriging variance is estimated from the solution by

$$\sigma^2(B) = \sum_{i=1}^{n} \bar{\gamma}(\mathbf{x}_i, B) + \psi - \bar{\gamma}(B,B). \tag{15}$$

Equations (13) to (15) show that the weights and the kriging variances depend on the variogram and on the configuration of the sampling points in relation to the block to be estimated and not on the observations themselves. To obtain the semi-variances for equation (13) involves finding a function for the variogram, and we deal with this later. In general the sampling points within or near the block carry large weights while more distant ones have small weights. Points that are clustered individually have smaller weights than isolated points, and near points can be screened by others lying between them and the block. Thus the estimate is local. The variogram must be estimated as accurately as possible over the first few lags, and the model should fit well there.

The block B may be of any reasonable size and shape. At its smallest it can be a 'point', $\mathbf{x}_0$, of the same size and shape as that on which the original measurements were made, i.e. the *support* of the sample. In these circumstances $\mathbf{x}_0$ replaces B in equations (11) to (15) and we have *punctual kriging*. The quantity $\bar{\gamma}(B,B)$ is zero and so disappears from equations (13) and (15). Punctual kriging can be used for predicting values at unvisited or unrecorded sites from data in the neighbourhood, and with the same support. Both punctual and block estimates can be used for interpolation in mapping. Values of the property can be estimated at points and blocks spaced as closely as desired to produce a statistical surface that can be 'contoured' by any of the standard computer programs.

A map drawn from punctual estimates is often regarded as the best that can be made because the interpolated surface passes through the data. In the presence of a nugget variance, however, there will be local discontinuities at the sampling points. These can obscure the spatially dependent variation. By computing estimates over larger blocks this nugget effect, which may be due either to measurement error or very short range variation that is conservatively represented in the variogram or both, can be avoided. The interpolated surface from block kriging is smoother, and the longer range variation can be detected more easily. We illustrate this in Part II of the paper. There is often little difference between the estimates themselves for points and blocks, but the estimation variances decrease as the block size over which estimates are made increases: it is another facet of the spatial smoothing. Consequently block estimates appear more reliable than those for points.

Simple kriging described above is, however, just the simplest in a family of techniques of spatial estimation. Co-kriging is the most obvious extension in which additional variables are incorporated into the linear kriging model to improve the estimation in a way analogous to their use in multiple regression (McBratney and Webster, 1983; Vauclin *et al.*, 1983). A more complex extension is universal kriging. Here the simple model of equation (4) no longer applies. Variation is assumed to comprise both a drift and a random component, equation (3), and universal kriging takes account of both. However, the method is by no means universally applicable (Webster and Burgess, 1980), and investigators need to be quite

sure that it is appropriate to their circumstances when using it. Lastly we mention disjunctive kriging (Matheron, 1976). An investigator may wish to make decisions based on the probability that the estimates exceed or are less than some critical threshold. If the variable has a known probability distribution, ideally normal, then such probabilities can be estimated from that distribution. In many instances this is not so, and in these circumstances disjunctive kriging solves the problem. It transforms the data non-linearly to a normal distribution and then combines them, also non-linearly, to arrive at its estimates. Its estimation variances are often less than those of simple kriging. Webster and Oliver (1989) describe the technique and give examples elsewhere.

OPTIMIZING SAMPLING

As described above the estimation variances for simple kriging, equation (13), depend on knowing only the variogram and the configuration of the observations in relation to the point or block to be estimated; they do not depend on the observed values themselves. This fact can be exploited in designing sampling schemes for mapping spatial variables. Burgess *et al.* (1981) and McBratney *et al.* (1981) showed how to do this. They computed the estimation variances for estimates at points and over blocks on regular grids for a range of sampling intensities. They plotted the variances against the grid spacing and then determined the optimal spacing for a given precision from the graph. Oliver and Webster (1987) used this technique and then followed it by sampling to map the particle size distribution of the soil. Webster and Burgess (1984) showed that the approach can also be used to optimize the location of sites from which to bulk samples. If the variogram is known and used in this way the necessary sampling effort is usually found to be less than that suggested by classical statistics; in many instances much less.

ESTIMATING THE VARIOGRAM

The variogram is central to geostatistics. We have already shown that it is essential for optimal estimation and interpolation by kriging. In addition the variogram summarizes the spatial variation in the region of interest provided that the intrinsic hypothesis holds. The semi-variance for any given lag **h** in one, two or three dimensions is readily estimated from sample data. The usual formula for computing it is

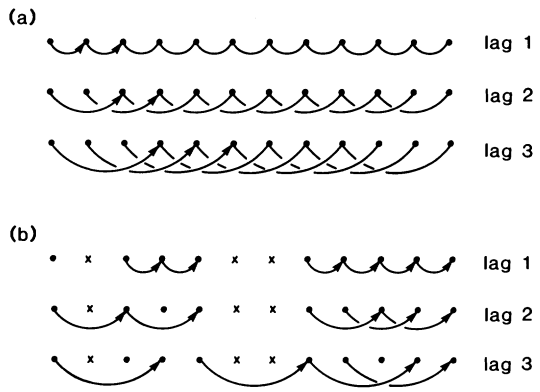

FIGURE 1. Comparisons for estimating semi-variances on linear transects at lags of 1, 2, and 3 sampling intervals, (a) for complete data, and (b) where some observations are missing, marked by crosses

$$\gamma(\mathbf{h}) = \frac{1}{2\,M(\mathbf{h})} \sum_{i=1}^{M(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2, \quad (16)$$

where $M(\mathbf{h})$ is the number of pairs of observations separated by the lag **h**. Figure 1a shows how the comparisons between points are made along a regular transect for $|\mathbf{h}| = 1$, 2 and 3 lag intervals. Thus by increasing h an ordered set of values is obtained, and this constitutes the sample or experimental variogram. Missing values are allowed for by including only the actual number of comparisons as in Figure 1b. For two-dimensional data the lag interval **h** can be grouped by both distance and direction (David, 1977; Webster, 1985). To detect directional differences or anisotropy the variogram should be estimated in at least three directions.

The variogram describes the magnitude, spatial scale and general form of the variation. It can indicate whether the data are second-order stationary or just intrinsic. A variogram that appears to rise with a concave upward form from the origin may indicate local drift or global trend. A full structural analysis, in Olea's (1975) sense, should then be performed so that the deterministic and stochastic components, equation (3), can be distinguished. The variogram then describes the residuals from the drift or trend.

Several points must be considered when estimating and interpreting the variogram. The sample variogram of any property in a given region is not unique. It is a function of the scale of the investigation, i.e. the size of the region, and the support of the sample, that is the size, shape and orientation of the areas of land or bodies of material on which individual measurements are made. The larger the support the more

variation each measurement encompasses, and so the less there is in the intervening space. This has a smoothing effect on the variogram. The support must remain constant throughout an investigation and should be reported. When the support is small the variogram can be regarded as a punctual variogram. Estimated semi-variances are subject to error, and their precision depends on the number of comparisons at each lag and therefore on the sample size. We recommend a minimum of a hundred comparisons at the first lag, and following practice in the analysis of time series, suggest that estimates should be made for lag distances no more than a fifth of the entire transect for one dimension. There is no simple way of determining confidence limits on variograms analytically. Simulation studies by Omre (1984) and McBratney and Webster (1986) show that confidence limits are very wide for the longer lags. The larger the sample and the shorter the lag the better the semi-variance is estimated.

In some instances the data contain outliers, in others they are strongly skewed. Geochemical data in particular have long upper tails in their distributions. Both outliers and long tails can have a disproportionate effect on the value of the semi-variance. Cressie and Hawkins (1980) proposed a robust method for estimating the variogram. They discovered that the fourth root of the squared differences has a distribution close to normal, and they used this to compute a mean, $\overline{y}$, given by

$$\overline{y}=\frac{1}{M}\sum_{i=1}^{M}[\{z(\mathbf{x})-z(\mathbf{x}+\mathbf{h})\}^2]^{\frac{1}{4}}. \qquad (17)$$

The quantity $\overline{y}$ must be transformed back to a semi-variance, and Cressie and Hawkins showed that the required transformation is

$$\hat{\gamma}(\mathbf{h})=\overline{y}^4/2(0{\cdot}457+0{\cdot}494M^{-1}+0{\cdot}045M^{-2}). \qquad (18)$$

McBratney and Webster (1986) examined the method and concluded that it conferred little benefit. Positive skewness can usually be removed by transformation. Taking logarithms is often effective. A more general normalizing transformation is that embodied in disjunctive kriging which uses Hermite polynomials (Matheron, 1976). This will convert almost any distribution to normal, though the later use of the variogram does assume second-order stationarity.

There is an alternative and much older procedure for estimating the variogram, albeit more crudely. Miesch (1975) has shown that a first approximation to the variogram can be obtained by a nested analysis. The procedure has an important role in an overall spatial investigation, as we shall demonstrate in Part II. It is an adaptation of classical multi-stage sampling and analysis originally devised by Youden and Mehlich (1937) to determine the range of distances over which most of the spatial variation in the soil of a region occurred. The principle of multi-stage sampling is that an individual observation embodies variation from each stage in a hierarchy, including the unresolved variation from the lowest stage, and the contributions from each stage are estimated by the analysis. These contributions are known as components of variance, and provided they can be regarded as independent their confidence limits can be estimated. If the sampling is suitably randomized the estimates are unbiased. The sum of the individual components of variance is the total variance of the sample. Snedecor and Cochran (1980) and Webster (1977) describe the method fully.

In Youden and Mehlich's (1937) adaptation the stages in the heirarchy represent specific spatial scales, and the components of variance estimate the variation attributable to them. When the components of variance are accumulated, starting with the smallest spacing, they are equivalent to the semi-variances of regionalized variable theory (Miesch, 1975). We describe this link in Part II of the paper.

The ordered set of accumulated components of variance form a crude variogram, and they can be plotted against sample spacing (see Part II for an example). The crudeness is a consequence of the sampling design because it is feasible to include only a few lag distances. The great merit of the method is that the variation over several orders of magnitude of distance can be covered in a single sampling. It is especially valuable when little or nothing is known about the spatial scale of variation (Oliver and Webster, 1986a), and so can be used as the first stage in a more comprehensive survey (McBratney *et al.*, 1981).

FORMS AND MODELS OF VARIOGRAMS

An ordered set of values, $\hat{\gamma}(\mathbf{h})$, a sample variogram, when plotted displays the average change of a property with changing lag. Semi-variances are estimated at discrete values of $\mathbf{h}$, whereas the true variogram is continuous. Furthermore the estimates are subject to error, and unless a large sample is taken (several hundred points) the experimental variogram

will appear erratic. An investigator will usually want to fit some kind of model to the sample values to represent the true variogram for a region. Suitable models should be able to incorporate the main features of variograms that we describe below. The models must also be conditional negative semi-definite, CNSD, (Journel and Huijbregts, 1978). This means that the variance of any linear combination of the values of a regionalized variable provided by the model must be positive or zero: variances cannot be negative.

As it happens there are just a few simple models that satisfy all these constraints. They fall into two broad groups which for convenience we may call unbounded (Fig. 2a, b and c) and bounded (Fig. 2d, e, f, g and h) models. Unbounded models have no finite *a priori* variance and the intrinsic hypothesis only holds. Bounded or transitive models reach an upper bound, known as the *sill*. The bound is the *a priori* variance of the random process, and its presence means that the variable is second-order stationary. Such models may indicate the occurrence of transition structures, e.g. blocks of land that are independent of each other but within which the values are highly correlated. The transition structures might represent discrete units with similar properties, such as types of strata, or they can overlap in all degrees to give rise to continuous variation. Their precise behaviour can be determined only by further investigation.

Figure 2 shows a few of the characteristic variograms and models for one dimension. The simplest models for unbounded variation are power functions. Their general form in one dimension or for isotropic variation in more is

$$\gamma(h) = wh^a \text{ for } 0 < a < 2. \tag{19}$$

where w is a linear parameter describing the intensity of the spatial variation, and $h = |\mathbf{h}|$ is the lag. The parameter a determines the shape of the variogram as Figure 2a to c shows. If $a = 1$ then we have the linear form. If $a < 1$ then the curve is convex upward, and conversely if $a > 1$ then the curve is concave upwards. The limits $a = 0$ and $a = 2$ are strict and excluded: a value of $a = 0$ would indicate white noise and therefore discontinuous variation, while a value of 2 would imply smooth differentiable variation and therefore not random. This type of model can be linked with the theory of fractals (Mandelbrot, 1982; Burrough, 1981, 1983).

Figure 2d, e, f shows examples of bounded variograms. These describe second-order stationary variation. The simplest is the bounded linear model, Fig. 2d which is given by

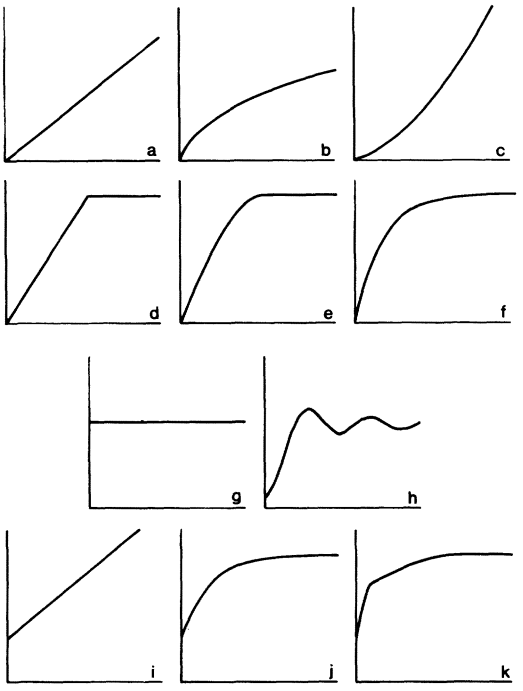

FIGURE 2. Examples of some characteristic forms of variogram

$$\begin{cases} \gamma(h) = c\,(h/a) \text{ for } h \leqslant a \\ \gamma(h) = c. \qquad \text{for } h > a \end{cases} \tag{20}$$

The variogram rises linearly to its sill, c, at $h = a$, which is the *range* of the model. The range defines the limit of spatial dependence. The model is valid for one dimension only: it is not authorized for more because it is not CNSD in two and three dimensions.

Figure 2e shows the type of transitive variogram that is very common in the earth sciences. The rising part curves before flattening to its sill. It is often fitted well by a spherical model, which for isotropic variation is given by

$$\begin{cases} \gamma(h) = c\left\{\dfrac{3h}{2a} - \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3\right\} \text{ for } h \leqslant a \\ \gamma(h) = c, \qquad\qquad\qquad \text{for } h > a \end{cases} \tag{21}$$

where a is the range, and c is the sill variance as before. This model is also CNSD in two and three dimensions. This type of variogram has been interpreted as evidence of the variation consisting of transition structures in three dimensions, i.e. different types of soil or rock types that have similar extent.

Figure 2f is also a transitive variogram. It differs from the previous two in that it approaches its sill asymptotically and is fitted by an exponential model. The formula is

$$\gamma(h) = c\{1 - \exp(-h/r)\}. \quad (22)$$

There is no finite range, though in practice an effective range a' is determined from the distance parameter r of the model, as $a' = 3r$. The model is CNSD for one, two or three dimensions, and it has been found to fit the experimental variogram well in many studies. It can arise from transition structures that vary in size in a random manner and also from autoregressive and Markov processes.

We have already remarked that the power function for unbounded variation, equation (17), may not have an exponent of zero since this would imply total independence in the variable and discontinuity. In equation (17) the semi-variance at lag $|\mathbf{h}| = 0$ would be w. Nevertheless there are many instances where the experimental variogram appears flat and to cut the ordinate at some such positive value. We can formalize this combination of a flat variogram and spatial variance of 0 at lag zero by defining

$$\gamma(h) = c_0\{1 - \delta(h)\}, \quad (23)$$

where $\delta(h)$ is the Dirac function taking the value 1 when $h = 0$ and zero otherwise.

This kind of behaviour was recognized early in the history of geostatistics in gold mining. It was known as the 'nugget effect', and was attributed largely to the chance occurrence of gold nuggets in drill cores. In most other fields it arises from a combination of measurement errors and spatially dependent variation on scales much shorter than the smallest sampling interval. Horizontal variograms, such as the one in Fig. 2g, are 'pure nugget': they indicate that there is no spatial dependence among the observations at the scale of sampling. Soil variograms from an initial survey of the Wyre Forest (Oliver and Webster, 1987) were like this because the sampling sites were too far apart and thus spatially independent. More intensive sampling is needed in such situations to reveal the spatially dependent variation.

Figure 2h is a 'hole effect' variogram. It suggests repetition in the variation that is neither wholly random nor yet periodic. The more marked the reversal of slope the more regular the repetition. One formula for a hole effect model that is CNSD in three dimensions is given by

$$\gamma(h) = c\left\{1 - \sin\left(\frac{2\pi h}{r}\right)/h\right\}, \quad (24)$$

where r is the wavelength of the hole effect and c its amplitude.

In many instances the observed variation is too complex to be described by one of these simple models. More complex mathematical functions can be sought, but the usual solution is to combine two or more of the simple functions listed above. Any combination of CNSD functions is itself CNSD and so there is no need to prove the function to be CNSD, which is not easy. The most common requirement is for a model with both nugget and spatially dependent variance. For example, an unbounded variogram might be represented by a nugget variance c_0 plus a power function (Fig. 2i):

$$\gamma(h) = c_0\{1 - \delta(h)\} + w\,h^a, \quad (25)$$

and a bounded one that approaches its sill asymptotically by a nugget variance plus an exponential component (Fig. 2j):

$$\gamma(h) = c_0\{1 - \delta(h)\} + c\{1 - \exp(-h/a)\}. \quad (26)$$

The Dirac function is usually omitted from the formulae, since it is understood that $\gamma(0) = 0$.

Where spatial dependence can be detected at two distinct scales the model can combine two simple models from the above. One combination that has been much used in mining is the double spherical model which in one dimension and for isotropic variation is

$$\begin{cases} \gamma(h) = c_1\left\{\dfrac{3h}{2a_1} - \dfrac{1}{2}\left(\dfrac{h}{a_1}\right)^3\right\} + \\ \qquad\qquad c_2\left\{\dfrac{3h}{2a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3\right\} \text{ for } 0 < h \leqslant a_1 \\ \gamma(h) = c_1 + c_2\left\{\dfrac{3h}{a_2} - \dfrac{1}{2}\left(\dfrac{h}{a_2}\right)^3\right\} \text{ for } a_1 < h \leqslant a_2 \\ \gamma_2(h) = c_1 + c_2, \text{ for } h > a_2 \end{cases} \quad (27)$$

where c_1 and a_1 are the sill and range of the shorter range component and c_2 and a_2 are those of the longer range component. This can be combined further with a nugget component. McBratney *et al.* (1982) used such a model Fig. 2k to describe the variation of copper and cobalt in the soil of south-east Scotland. The long-range component, with $a_2 \simeq 15$ km, seemed clearly attributable to the major geological formations. The shorter range one, with $a_1 \simeq 3$ km, they attributed to variation from farm to farm.

The examples we have given here are for one-dimensional or isotropic variation. Natural features, however, do not always vary at the same rate in all directions. For instance river alluvium varies much more intensely at right angles to the river's course than it does parallel to it. The variogram will express this anisotropy. If the anisotropy can be accounted for by a simple linear transformation of the coordinates then it is said to be geometric or affine (Journel and Huijbregts, 1978). Such anisotropy can then be represented by

$$\Omega(\theta) = \{A^2\cos^2(\theta\text{-}\varphi) + B^2\sin^2(\theta\text{-}\varphi)\}^{\frac{1}{2}}. \quad (28)$$

Here A is the gradient of the variogram in the direction of maximum variation or distance parameter, B is the gradient in the direction of minimum variance or distance parameter in the perpendicular direction, and φ is the angle of the maximum gradient or distance parameter. The ratio $A{:}B$ measures the anisotropy. The anisotropic function, $\Omega(\theta)$, can be applied to either the gradient of an unbounded model, such as a power function:

$$\gamma(h,\theta) = \Omega(\theta)|\mathbf{h}|^a, \quad (29)$$

or to the distance parameter of a bounded model, such as an exponential function:

$$\gamma(h,\theta) = c\,[1 - \exp\{-|\mathbf{h}|/\Omega(\theta)\}]. \quad (30)$$

We give an example of anisotropy in Part II.

CHOOSING AND FITTING MODELS

We have shown above some of the common forms of variogram and the simple mathematical functions that may be used to describe them. There remain the tasks of choosing a function from among the plausible ones and fitting it to the experimental values. These are important especially where the variogram is to be used later for kriging.

Some of the early geostatistical practitioners fitted models by eye, and there are many who still do it qualitatively by trial and error. McBratney and Webster (1986) found it unreliable, as might be expected, and we do not recommend it. The most generally satisfactory method is to fit models by weighted least squares approximation (Cressie, 1985; McBratney and Webster, 1986).

The weights can be the number of pairs of comparisons, M, that contribute to the estimates, and these are the ones most widely used. More elaborate schemes weight the estimates, either instead or in addition, by some inverse function of the estimates themselves or the expected semi-variances: the larger the semi-variance the less confidence one has in it. McBratney and Webster (1986) discuss these and their merits at some length. Cressie proposed a weight

$$w(\mathbf{h}) = M(\mathbf{h})/\mathrm{E}[\gamma(\mathbf{h})]^2, \quad (31)$$

where $M(\mathbf{h})$ is the number of pairs of comparisons at lag $\mathbf{h}$ and $\mathrm{E}[\gamma(\mathbf{h})]^2$ is the semi-variance expected from the model, while G. M. Laslett, quoted in McBratney and Webster (1986), suggested the improvement:

$$w(\mathbf{h}) = M(\mathbf{h})\,\hat{\gamma}(\mathbf{h})/\mathrm{E}[\gamma(\mathbf{h})]^3. \quad (32)$$

The differences in the fitted models produced by these schemes are usually small, but they do produce models that fit better at the shorter lags.

Most of the models are non-linear in one or more of their parameters. They must be fitted iteratively, and an efficient and numerically sound computer program is essential. We use MLP, the Maximum Likelihood Program, written by Ross (1987). The same algorithms are embodied in the more widely available Genstat 5 (Genstat 5 Committee, 1987).

One can apply the same criteria of fitting, namely the minimum residual sum of squared deviations from the model, to choose among several plausible models. This works well if all the models have the same number of parameters. However, the goodness of fit can always be improved by adding parameters, and some kind of compromise must usually be struck between simple parsimony and elaborate close fit. Many criteria have been proposed for selecting models in regression analysis, all embracing a penalty for increased complexity. One that seems to work well is Akaike's (1973) Information Criterion, used by McBratney and Webster (1986) and described by them at greater length later (Webster and McBratney, 1989).

The Akaike Information Criterion (AIC) is defined as

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

and it is estimated by

$$A = \left\{n\ln\left(\frac{2\pi}{n}\right) + n + 2\right\} + n\ln RSS + 2p, \quad (33)$$

where n is the number of individuals, i.e. the number of lags at which the semi-variance is estimated, p is the number of parameters and RSS is the residual sum of squares. The quantity in curly brackets is constant for a given set of data, and so models can be compared by

computing $n \ln RSS + 2p$ only, and choosing that model for which this quantity is least. If all models have the same number of parameters then minimizing the AIC is equivalent to minimizing *RSS*. If, however, *RSS* is diminished only by increasing p then the AIC might actually increase: the AIC contains the penalty for adding to the complexity.

CONCLUSION

Regionalized variable theory provides a concise, coherent and useful body of theory physical geographers can use to describe spatial variation in phenomena over the earth's surface. The theory provides quantitative tools for estimation and interpolation, and for planning efficient sampling. In the second part of the paper we describe examples to demonstrate the effectiveness of the techniques in the earth sciences.

ACKNOWLEDGEMENTS

We thank Mrs Joyce Munden for preparing the figures.

REFERENCES

AKAIKE, H. (1973) 'Information theory and an extension of maximum likelihood principle', in PETROV, B. N. and CSAKI, F. (ed.) *Second international symposium on information theory* (Akadémia Kiadó, Budapest) pp. 267–81

ATKINSON, P. M., CURRAN, P. J. and WEBSTER, R. (1989) 'Sampling remote imagery for storage, retrieval and reconstruction ', *Prof. Geogr.* (forthcoming)

BURGESS, T. M. and WEBSTER, R. (1980) 'Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging', *J. Soil Sci.* 31: 315–31

BURGESS, T. M., WEBSTER, R. and McBRATNEY, A. B. (1981) 'Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy', *J. Soil Sci.* 32: 643–59

BURROUGH, P. A. (1981) 'Fractal dimensions of landscapes and other environmental data', *Nature (London)* 294: 240–42

BURROUGH, P. A. (1983) 'Multiscale sources of spatial variation in soil. I. The application of fractal concepts to nested levels of soil variation', *J. Soil Sci.* 34: 577–97

CRESSIE, N. (1985) 'Fitting variogram models by weighted least squares', *Math. Geol.* 17: 563–86

CRESSIE, N. and HAWKINS, D. M. (1980) 'Robust estimation of the variogram I', *Math. Geol.* 17: 115–25

DAVID, M. (1977) *Geostatistical ore reserve estimation* (Elsevier, Amsterdam)

GAMBOLATI, G. and VOLPI, G. (1979) 'Groundwater contour mapping in Venice by stochastic interpolators. I. Theory', *Water Res. Res.* 15: 281–90

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

GENSTAT 5 COMMITTEE (1987) *Genstat 5 reference manual* (Clarendon Press, Oxford)

GUARASCIO, M., DAVID, M. and HUIJBREGTS, C. (1976) (eds) *Advanced geostatistics in the mining industry* (Reidel, Dordrecht)

JOURNEL, A. G. and HUIJBREGTS, C. J. (1978) *Mining geostatistics* (Academic Press, London)

KITANIDIS, P. K. and VOMVORIS, E. G. (1983) 'A geostatistical approach to the inverse problem in groundwater modelling (steady state) and one-dimensional simulations', *Water Res. Res.* 19: 677–90

KOLMOGOROV, A. N. (1941) 'Interpolirovanie i ekstrapolirovanie statsionarnykh sluchainykh posledovatel'nostei (Interpolated and extrapolated stationary random sequences)', *Izvestiya AN SSSR, seriya matematicheskaya* 5: No. 1

KRIGE, D. G. (1951) 'A statistical approach to some basic mine evaluation problems on the Witswatersrand', *J. Chem. Metall. and Mining Soc. of S. Afr.* 52: 119–39

KRIGE, D. G. (1966) 'Two-dimensional weighted moving average trend surface for ore evaluation', *J. S. Afr. Inst. of Mining and Metall.* 66: 13–38

KRIGE, D. G., GUARASCIO, M. and CAMISANI-CALZOLARI, F. A. (1989) 'Early South African geostatistical techniques in today's perspective', in ARMSTRONG, M. (ed.) *Geostatistics, volume I* (Kluwer, Dordrecht) pp. 1–19

LAJAUNIE, G. (1984) 'A geostatistical approach to air pollution modelling', in VERLY, G., DAVID, M., JOURNEL, A. G. and MARECHAL, A. (eds) *Geostatistics for natural resources characterization* (Reidel, Dordrecht) pp. 877–91

LASLETT, G. M., McBRATNEY, A. B., PAHL, P. J. and HUTCHINSON, M. F. (1987) 'Comparison of several spatial prediction methods for soil pH', *J. Soil Sci.* 38: 325–41

MANDELBROT, B. B. (1982) *The fractal geometry of nature* (Freeman, San Francisco)

MATÉRN, B. (1960) 'Spatial variation', *Meddelanden från Statens Skogsforskninginstitut* 49 (5): 1–144

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

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

MATHERON, G. (1976) 'A simple substitute for conditional expectation: the disjunctive kriging', in GUARASCIO, M., DAVID, M. and HUIJBREGTS, C. (eds) *Advanced geostatistics in the mining industry* (Reidel, Dordrecht) pp. 221–36

McBRATNEY, A. B. and WEBSTER, R. (1983) 'Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy', *J. Soil Sci.* 34: 137–62

McBRATNEY, A. B. and WEBSTER, R. (1986) 'Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates', *J. Soil Sci.* 37: 617–39

McBRATNEY, A. B., WEBSTER, R. and BURGESS, T. M. (1981) 'The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and Method', *Computers and Geosciences* 7: 331–34

McBRATNEY, A. B., WEBSTER, R., McLAREN, R. G. and SPIERS, R. B. (1982) 'Regional variation of extractable copper and cobalt in the topsoil of south-east Scotland', *Agronomie* 2: 969–82

McCULLAGH, M. J. (1975) 'Estimating by kriging the reliability of the proposed Trent telemetry network', *Comp. Appl.* 2: 357–74

MIESCH, A. T. (1975) 'Variograms and variance components in geochemistry and ore evaluation', in WHITTEN, E. H. T. (ed.) *Quantitative studies in the geological sciences* (Geological Society of America Memoir, 142) pp. 333–40

MOFFAT, A. J., CATT, J. A., WEBSTER, R. and BROWN, E. H. (1986) 'A re-examination of the evidence for a Plio-Pleistocene marine transgression in the Chiltern Hills. I. Structure and surfaces', *Earth Surface Processes and Landforms* 11: 95–106

NIEUWENHUIS, J. D. and van den BERG, J. A. (1971) 'Slope investigations in the Morvan', *Revue de Géomorphologie Dynamique* 20: 161–76

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

OLIVER, M. A. and WEBSTER, R. (1986a) 'Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables', *Geogr. Anal.* 18: 227–42

OLIVER, M. A. and WEBSTER, R. (1986b) 'Semivariograms for modelling the spatial pattern of landform and soil properties', *Earth Surface Processes and Landforms* 11: 491–504

OLIVER, M. A. and WEBSTER, R. (1987) 'The elucidation of soil pattern in the Wyre Forest in the West Midlands, England. II. Spatial distribution', *J. Soil Sci.* 38: 293–307

OLIVER, M. A. and WEBSTER, R. (1989) 'A geostatistical basis for spatial weighting in multivariate classification', *Math. Geol.* 21: 15–35

OLIVER, M., WEBSTER, R. and GERRARD, J. (1989) 'Geostatistics in physical geography. Part II: applications'. *Trans. Inst. Br. Geogr.* N.S. 14: 270–86

OMRE, H. (1984) 'The variogram and its estimation', in VERLY, G., DAVID, M., JOURNEL, A. G. and MARECHAL, A. (eds) *Geostatistics for natural resources characterization* (Reidel, Dordrecht) pp. 107–25

ROSS, G. J. S. (1987) *MLP user manual* (Numerical Algorithms Group, Oxford)

SNEDECOR, G. W. and COCHRAN, W. G. (1980) *Statistical methods.* (Iowa State University Press, Ames). 7th edition

THORNES, J. B. (1973) 'Markov chains and slope series: the scale problem', *Geogr. Anal.* 5: 322–28

VAUCLIN, M., VIEIRA, S. R., VACHAUD, G. and NIELSEN, D. R. (1983) 'The use of cokriging with limited field soil observations', *Soil Sci. Soc. Am. J.* 47: 175–84

VERLY, G., DAVID, M., JOURNEL, A. G. and MARECHAL, A. (eds) (1984) *'Geostatistics for natural resources characterization'* (Reidel, Dordrecht)

WACKERNAGEL, H., WEBSTER, R. and OLIVER, M. A. (1988) 'A geostatistical method for segmenting multivariate sequences of soil data', in BOCK, H. H. (ed.) *Classification and related methods of data analysis* (North-Holland, Amsterdam) pp. 641–50

WEBSTER, R. (1977) *Quantitative and numerical methods in soil classification and survey* (Clarendon Press, Oxford)

WEBSTER, R. (1985) 'Quantitative spatial analysis of soil in the field', *Adv. in Soil Sci.* 3: 1–70

WEBSTER, R. and BURGESS, T. M. (1980) 'Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging', *J. Soil Sci.* 31: 505–24

WEBSTER, R. and BURGESS, T. M. (1984) 'Sampling and bulking strategies for estimating soil properties in small regions', *J. Soil Sci.* 35: 127–40

WEBSTER, R. and CUANALO de la C, H. E. (1975) 'Soil transect correlograms of North Oxfordshire and their interpretation', *J. Soil Sci.* 26: 176–94

WEBSTER, R. and McBRATNEY, A. B. (1987) 'Mapping soil fertility at Broom's Barn by simple kriging', *J. Sci. of Food and Agr.* 38: 97–115

WEBSTER, R. and McBRATNEY, A. B. (1989) 'On the Akaike Information Criterion for choosing models for variograms of soil properties', *J. Soil Sci.* 40: (forthcoming)

WEBSTER, R. and OLIVER, M. A. (1989) 'Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability', *J. Soil Sci.* 40: (forthcoming)

YAGLOM, A. M. (1962) *'An introduction to the theory of stationary random functions'* (Prentice-Hall, Englewood Cliffs, N.J.)

YOST, R. S., UEHARA, G. and FOX, R. L. (1982a) 'Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms', *Soil Sci. Soc. Am. J.* 46: 1028–32

YOST, R. S., UEHARA, G. and FOX, R. L. (1982b) 'Geostatistical analysis of soil chemical properties of large land areas. II. Kriging', *Soil Sci. Soc. Am. J.* 46: 1033–37

YOUDEN, W. J. and MEHLICH, A. (1937) 'Selection of efficient methods for soil sampling', *Contr. Boyce Thompson Inst. for Plant Res.* 9: 59–70