CHAPTER 13

Statistical Models for Multiple-Scaled Analysis

Donald E. Myers

INTRODUCTION

Geographic information systems might be viewed as spatial database managers. Both for what a GIS can do and also for the analysis of such data it is important to consider the connotation "spatial." In a spatial data set each data entry is associated with a location and there is a relationship between the data value and its location. This "location" may be a point in space or it may be a geometrical object, for example, a rectangle as in a pixel for remotely sensed data. The size and shape of this location is also important, especially as it relates to the scale of the data set. The relationship between a data value and its "location" especially must be considered when data are to be analyzed at multiple scales or when the scale is to be changed. This relationship is also important when considering spatial variability. Since this relationship is seldom explicitly describable, statistical methods are used to treat the data "ensemble" rather than as individual data entries.

There are at least three areas in which statistical modeling is important for multiple-scaled analyses in GIS, particularly remotely sensed data. First, there is the quality of the data. In general all GIS packages treat the data values as being without error and do not provide for the incorporation of any measure of the reliability of the data. Secondly, there is what might be called the completeness of a data layer. Goodchild et al. (1992) have described six different data models. One corresponds to a form of compositing (raster of cells), two of the others correspond to sparse data sets and differ only in that the grid is regular for one and not for the other (digital elevation model and weather map). A fourth (contour map) might be viewed as having been derived from one of the other three, but the resulting data set does not uniquely determine one of the other models as the source. Multiple-scaled analyses will require a common data model. Thirdly, the spatial variability of a data

1-56670-104-X/97/$0.00+$.50

set is a crucial characteristic and is a function of scale; it is important to be able to relate the variability at one scale to that at another. Because the characterization and quantification of variability is important for understanding and characterizing the first two, it is more convenient to consider it first. Spatial variability often changes with respect to time and the characterization of this change is an important aspect of spatial-temporal modeling; see for example Biondi et al. (1994). Even with the soil or vegetation cover, data model variability plays a role. While each polygon may be labeled with a single soil or vegetation type, this labeling may be based on information collected on a much smaller sized polygon. For example, in EPA's EMAP program, data will be collected in a multiple tiered scheme and each tier will have its own degree of variability.

Spatial variation is also important for classification analysis of remotely sensed data since pixels are grouped into classes because of similarities, and classes are distinguished because of dissimilarities (Fabbri et al., 1993). That is, it is necessary to identify and quantify *within-band* correlations but it is also necessary to quantify and characterize *between-band* correlation. This may involve only correlations between bands for the same pixel or a combination of spatial and interband correlations. When data layers are of different scales, it is important to quantify the change in spatial and interband correlations when scales are changed in order to use the multiple layers for classification analysis.

For some measurements it is difficult to quantify the scale of the measurement; for example, hydraulic conductivity is commonly associated with a "representative volume" but the value is not an average over the volume. In many instances measurements can only be considered as point values in a limiting sense: hence, although they are quantitative they are not comparable to remotely sensed data with a fixed pixel size nor to averaged values.

Heterogenity, Homogeneity, and Scale

Since at least four of the data models relate at least indirectly to point values, it is useful to construct a model for point values that reflects different scales. Let

$$Z(x) = Y(x) + m(x) + e(x) \tag{1}$$

This can be interpreted as a decomposition of the heterogenities and homogeneities. Y(x) represents the local variation, m(x) the regional variation, and e(x) the extreme local variation or simply noise. The importance of each of these as well as the extent to which they contribute to the representation is related to scale. When measurements are made on a sufficiently large scale, local variation may be indistinguishable from noise. An image that is too homogeneous contains little or no useful information and features will not be recognizable. An image that is too heterogeneous may contain useful information but will be difficult to discern. Important features will be distinguished by local regions of homogeneity separated by heterogeneities. The distinction between the two will be maximally discernable at the right scale. This requires quantifying the degrees of homogeneity and heterogeneity.

Time and Space

Time might be considered simply as an additional dimension, but there are several difficulties with this. One is that time is ordered; that is, there is a clear notion of a past, a present and a future. Time is irreversible whereas the Euclidean dimensions are ordered only by convention. The two or three Euclidean dimensions are distinguishable from each other by convention but not uniquely by order; the same metric is applicable to all three and a natural notion of distance in 2- or 3-D space is available but no such metric is natural for space-time.

One solution is to consider a discrete set of times and to view the same variable at different times as different variables. This approach is discussed in Rouhani and Myers (1990). It corresponds to the treatment of multiple data layers, one for each time point. Interband correlation would then correspond to temporal correlation. As will be noted in the later discussion on intervariable spatial correlation, there are difficulties in modeling space-time correlation. These are similar but not identical to the difficulties encountered in modeling multiple spatial correlations.

SPATIAL VARIABILITY

An Empirical Form

Consider a region V decomposed into a union of disjoint, congruent subregions $v_1,\ldots,v_m$. These might be thought of as pixels in a remotely sensed data set where V represents the entire image or some significant portion thereof. Suppose that $y_1,\ldots,y_m$ are the centers of the respective subregions. Let Z(x) represent a variable of interest defined at each point x in V. Let v be a congruent subregion centered at zero; then each v_i can be written as

$$\{y_i + x \mid x \in v\}$$

That is, each subregion is a translate of the subregion centered at zero, the translation being determined by its center. Further let

$$Z(V) = (1/V)\int_V Z(x)\,dx \tag{2}$$

be the average value of Z over V, i.e., the regional average value. Similarly let

$$Z(v_i) = (1/v)\int_v Z(x + y_i)\,dx \tag{3}$$

where v is the congruent subregion centered at the origin. $Z(v_i)$ is then the average of point values within v_i. A word about notation is necessary; v and V have been used in two different senses; in Eq. 2 and 3, they have been used to denote a region

of integration and also to denote the "size" (length, area, volume) of the same regions. The context should make it clear which sense is intended. Similarly although only a single integral sign appears in each of these expressions, depending on the dimension of the space, each of the above integrals will in fact be a single, double, or triple integral. This usage will appear elsewhere in the chapter.

Spatial variability can now be compared at three different scales. The variation of point values with respect to the regional average value is given by

$$S^2(O,V) = (1/V)\int_V [Z(x) - Z(V)]^2 \, dx \tag{4}$$

The variation of subregional average values with respect to the regional average value is given by

$$S^2(v,V) = (1/m)\sum_{[i=1,\ldots,m]} [Z(v_i) - Z(V)]^2 \tag{5}$$

Finally there is the average variation of point values within subregions

$$S^2(O,v) = (1/m)\sum_{[i=1,\ldots,m]} (1/v)\int_v [Z(v_i) - Z(x+y_i)]^2 \, dx \tag{6}$$

It remains to consider the relationship between these three. Note that Eq. 4 can also be written as

$$S^2(O,V) = (1/m)\sum_{[i=1,\ldots,m]} (1/v)\int_v [Z(x+y_i) - Z(V)]^2 \, dx \tag{7}$$

$$= (1/m)\sum_{[i=1,\ldots,m]} (1/v)\int_v [Z(x+y_i) - Z(v_i) + Z(v_i) - Z(V)]^2 \, dx \tag{8}$$

By squaring the integrand as the sum of two differences it is easy to see that

$$S^2(O,V) = S^2(O,v) + S^2(v,V) \tag{9}$$

For a given function Z(x) one would expect the variation of point values within an area or volume to increase with the size of the area or volume. For fixed V, however, the value of $S^2(O,V)$ is fixed and it will be seen that in general as the size of the subregions increase, $S^2(O,v)$ increases whereas $S^2(v,V)$ decreases. Considering the subregions to be the pixels in a remotely sensed image then $S^2(v,V)$ is the variation in the image whereas $S^2(O,v)$ is the unobserved (average) variation within pixels. If the data is noisy then one way to smooth out the noise is to increase the pixel size, but discrimination of the image may be lost as a consequence. This decomposition

is analogous to the partitioning of the variance used in ANOVA. While in general it is not possible to directly compute either $S^2(O,V)$ or $S^2(O,v)$, it is useful to have the identity given in Eq. 9. Note that

$$Z(V) = (1/m)\sum_{[i=1,\ldots,m]} Z(v_i)$$

hence relationships between first order moments are of less interest.

Frequency Distributions

More information would be obtained if one could compare the frequency distributions of point values within pixels against the frequency distribution of pixel values within the image as well as against the frequency distribution of point values in the entire image. If the latter distribution were Normal then the other two distributions would also be Normal, or nearly so, with common means but with variances related as in Eq. 9. It is also important to recognize that while the distributions of point values within a pixel or point values within V are both continuous, the distribution of pixel values is discrete since the number of pixels is finite. Unfortunately, in general one cannot describe the interrelationship between these distributions. Because of a Central Limit Theorem effect, the distribution of the $Z(v_i)$ values will tend to be approximately Normal as the size of v_i's increase even without the assumption of Normality for point values in V. In many applications an assumption of Normality may not be unreasonable and has the advantage that the joint Normal distribution is completely characterized by the first and second moments.

The above discussion of spatial variability does not make use of any assumptions about spatial dependence or correlation, but an image that contains useful information is not just a set of random values. Whether pixels, data on a regular grid or data at irregularly spaced locations, there is some degree or form of interrelationships between the values at different locations and the relationships between the values are in turn affected by the proximity or distances between data locations. To adequately address these issues it is necessary to proceed beyond empirical descriptions or measures and consider models, i.e., theoretical quantification.

Theoretical Models

Rather than considering only one image, one might consider the image to be one of multiple possibilities that could be generated. This is analogous to considering Z(x) to be a random function. The covariance function, $Cov\{Z(x),Z(y)\}$, is a common measure of spatial variation and spatial dependence which exists if Z(x) has finite variance for each x. In general this covariance will be a function of both x,y. If Z(x) is second order stationary then

$$C(h) = Cov\{Z(x+h), Z(x)\} \tag{10}$$

exists and is only a function of the separation vector h. Under somewhat weaker assumptions

$$\gamma(h) = 0.5\ \mathrm{Var}\left[Z(x+h) - Z(x)\right] \tag{11}$$

exists and is only a function of the separation vector h. If Z(x) has constant mean then

$$\gamma(h) = 0.5\ \mathrm{E}\left\{\left[Z(x+h) - Z(x)\right]^2\right\} \tag{11a}$$

$\gamma(h)$ is called the variogram and if Z(x) is second order stationary then

$$\gamma(h) = C(0) - C(h) \tag{12}$$

Var denotes variance, Cov denotes covariance and E denotes expected value. While stationarity, in its various forms, is a statistical property of the random function it is also possible to rationalize such a property on non-statistical grounds. Each data point has coordinates and these coordinates relate to an "origin," yet clearly there is no uniquely determined origin for all spatial measurements. It is then desirable that the quantification of spatial variability and spatial correlation be given in a manner that is not dependent on the choice of the origin for the coordinate system. This is exactly what is required when the covariance function and the variogram are to be functions of h only and not functions of the actual positions. The vector h represents the relative relationship of two locations but not their absolute locations.

Covariance functions must satisfy a condition known as positive definiteness whereas the variogram need only be conditionally negative definite. Covariances must be bounded whereas variograms need not be. When the variogram (or the covariance function) is a function of only the length of h instead of both the length and the direction then Z(x) (and its variogram or covariance) is said to be isotropic. The power model is an example of an isotropic variogram that does not correspond to a covariance. The power model is given by

$$\gamma(r) = r^a; \quad 0 < a < 2 \tag{13}$$

In the case of a fractal image, the exponent a is related to the fractal dimension; see Burrough (1983a, 1983b), Jaggi et al. (1993). The variogram (or the covariance function) can be used to quantify the spatial correlation between point values. In addition, the effect of the change of scale on the variogram is easily computed. As is implied by Eq. 11, variograms that correspond to covariances must be bounded and the asymptotic value is the variance; hence they are characterized by a range of correlation and the variance. In practice the variogram is modeled as a positive linear combination of one or more standard models; in particular this allows for modeling spatial correlation at multiple scales.

If we consider the theoretical analogue of Eq. 9 by applying the expectation operator and the definition of the variogram we obtain

$$\gamma(0, V) = \gamma(0, v) + \gamma(v, V) \tag{14}$$

where

$$\gamma(0, V) = \left(1/V^2\right)\int_V \int_V \gamma(u - w)\, du\, dw \tag{15a}$$

$$\gamma(0, v) = \left(1/v^2\right)\int_v \int_v \gamma(u - w)\, du\, dw \tag{15b}$$

$$\gamma(v, V) = \left(1/vV\right)\int_v \int_V \gamma(u - w)\, du\, dw \tag{15c}$$

Each of these three terms represents an average value of the variogram for pairs of points; in Eq. 15a the pairs of points are in V, in Eq. 15b both points are in v, in Eq. 15c one point is in v and one in V. As a theoretical relationship it can be useful in determining an optimal pixel size. This problem of optimal support size occurs in many areas of application. In mining one must balance the size of an ore block with the problem of adequate discrimination of blocks of ore from blocks of waste; see for example, Parker (1979). In environmental applications it is related to the question of optimal-sized regions to remediate. This kind of relationship has occurred in many areas of application. For example, Smith developed an empirical version of Eq. 10 in the context of applications in agronomy. For extensions of this result see Modjeska and Rawlings (1983), Zhang et al. (1990, 1994).

Non-Point Data

Remotely sensed data generated by sensors on satellites do not associate values with points. Rather there is a value representing an average for each pixel. When the spatial variation and spatial dependence of such data are considered then adjustments must be made in the use of the variogram or covariance function.

Consider the decomposition of the region V into congruent subregions as used above in the discussion of the empirical model. Then the analog of the point value variogram would be

$$\gamma_v(h) = 0.5\ \mathrm{Var}\left[Z\left(v_{x+h}\right) - Z\left(v_x\right)\right] \tag{16}$$

where v_{x+h}, v_x are the translations of the subvolume v by the vectors x + h, x, respectively. Using Eq. 3 above

$$\left[Z(v_{x+h}) - Z(v_x)\right]^2 = \left[(1/v)\int_v \left[Z(u+x+h) - Z(u+x)\right] du\right]^2$$

$$= (1/v^2)\int_v\int_v \left[Z(u+x+h) - Z(u+x)\right]\left[Z(w+x+h) - Z(w+x)\right] du\, dw \quad (17)$$

$$\gamma_v(h) = (1/v^2)\int_v\int_v \gamma(w-u+h)\, du\, dw - (1/v^2)\int_v\int_v \gamma(u-w)\, du\, dw$$

Note especially the second term on the right (with a negative sign). This is the average value of the variogram for all possible pairs of points within the subvolume v. There is also an averaging effect in the first term. A similar result is obtained for the covariance function. The second term on the right in Eq. 17 is the same as that in Eq. 15b.

Modeling the Variability

Except for some instances in hydrology, one generally does not have state equations from which the covariance function or variogram may be derived. Hence, the spatial correlation must be estimated and modeled from data. Since direct estimation of the covariance function requires separate estimation of the (constant) mean, estimation and modeling of the variogram has advantages. The simplest and most obvious estimator for the variogram is given by

$$\gamma^*(r,\theta) = \left[(1/2)/N(h)\right]\sum\left\{Z(x+h) - Z(x)\right\}^2 \quad (18)$$

where the sum is taken over all pairs of data locations (x + h, x) such that (i) $r - d/2 < |h| < r + d/2$ and (ii) $\theta - \varepsilon/2 < \arg(h) < \theta + \varepsilon/2$ where |h| is the length of the separation vector h and arg(h) is the angle of the vector, N(h) is the number of pairs satisfying these conditions. The use of distance classes and angle windows is necessary because of the use of irregularly spaced data locations, but is useful for regularly spaced locations as well. In practice θ is fixed and Eq. 18 is computed for various choices of r (commonly called lags). Common choices for θ are 0, 45, 90, 135 with $\varepsilon = 45$. It is necessary to consider different directions in order to determine whether the spatial correlation is isotropic or anisotropic. If only the range is dependent on the direction the anisotropy is called geometric and is easily incorporated. But more general forms of anisotropy are not so easily modeled. Note that Eq. 18 only produces estimated values of the unknown variogram whereas a functional form is needed. It is not sufficient to fit a curve to the plot since not every function will satisfy the required positive definiteness. Modeling the spatial correlation function then requires first fitting a functional type and then the values of the parameters. The second step lends itself to the use of weighted least squares. Because of the dependence, the sample variance is not an estimate of the value of C(0), i.e., the maximal value of the variogram in the case of second order stationarity. The general problem of variogram estimation and modeling is discussed in Myers (1991a,b).

The question of the number of data locations adequate for variogram estimation has been considered by Warrick and Myers (1987), Webster and Oliver (1993), as well as others. Note that it is not sufficient to merely consider the number, since the pattern of data locations is equally important. When non-point data are used, for example pixel values, the estimation and modeling of the variogram is complicated because the sample variogram given in Eq. 18 will be distorted; the estimate of the variance of the random function will be too small (because of the second term in Eq. 17), and the apparent range of dependence will be increased. There is also the potential for not distinguishing any spatial dependence. Non-stationarity of the random function will also complicate estimation and modeling of the variogram. Unfortunately when data from only one realization is available stationarity is not testable. For a discussion of this problem, see Myers (1989a).

INTERPOLATION OF SPATIAL DATA

If Z(x) is known at all points in a volume v or V then computing the average is straightforward. But if Z(x) is only known at discrete points, as on a regular grid or irregularly spaced, then some form of estimation is required. The problem is even more complicated if some of the data locations are outside the volume of interest. To utilize a common data model it is necessary to have data on a regular grid; hence if data are only available on an irregular grid then some form of interpolation is required. Similarly, most contouring algorithms require data on a regular grid. While some software packages will accept irregularly spaced data there are two distinct steps in the analysis. The first step is interpolation to a regular grid and then contouring from the regular grid. In order to minimize the effect of the contouring algorithm it is desirable to have data available on a regular grid with very small mesh; hence it may even be necessary to interpolate data from a regular grid. Both the interpolation and the contouring steps involve some uncertainty and a contour map produced from irregularly spaced data is not uniformly reliable although most contouring packages do not indicate this. Hence some contour maps may be very misleading or misrepresentative.

There are two general approaches to interpolation: for each grid point an estimator is constructed which is a function of the data; or alternatively an interpolating function is constructed which then can be evaluated at the desired locations. The former approach implicitly determines an interpolating function but may not provide an explicit representation. For a large class of interpolation algorithms these two approaches are equivalent and can be transformed one into the other.

Deterministic Methods

Consider the model given in Eq. 1 but with e(x) = 0 and

$$m(x) = \sum_{[j=0,\ldots,p]} c_j f_j(x_k) \tag{19}$$

where the $f_j(x)$, $j = 0,\ldots,p$ are known linearly independent functions but with unknown coefficients. These are commonly taken as monomials in the coordinates of x. The following is a fairly general form of an interpolating function

$$Z^*(x) = \sum_{[i=1,\ldots,n]} b_i g(x - x_i) + \sum_{[j=0,\ldots,p]} a_j f_j(x) \tag{20}$$

where the kernel g is a known conditionally positive definite function. If the interpolator is required to be exact, i.e., $Z^*(x_i) = Z(x_i)$ for $i = 1,\ldots,n$ then the following equations are obtained.

$$\sum_{[i=1,\ldots,n]} b_i g(x_k - x_i) + \sum_{[j=0,\ldots,p]} a_j f_j(x_k) = Z(x_k); \quad k = 1,\ldots n \tag{21a}$$

Even for fixed g and functions $f_j(x)$, the coefficients are not uniquely determined because the coefficient matrix is not invertible. If the kernel g is conditionally positive definite with respect to the $f_j(x)$ and the following conditions are adjoined to the system in Eq. 21a

$$\sum_{[i=1,\ldots,n]} b_j f_j(x_i) = 0; \quad j = 0,\ldots,p \tag{21b}$$

then the coefficient matrix is invertible and hence the coefficients in Eq. 20 are uniquely determined. The choice of the $f_j(x)$ affects how the interpolating function "extrapolates" and the choice of g determines the smoothness of the interpolating function.

The Thin-Plate Spline is a special case of Eq. 20 as is the Radial Basis Function interpolator. Using this interpolating function, spatial averages can then be estimated using irregularly spaced data locations by integration, i.e.,

$$(1/V)\int_V Z(x)\,dx \approx (1/V)\int_V Z^*(x)\,dx \tag{22}$$

To quantify the error in the estimation of the spatial integral in this way requires quantifying the error in the interpolation step for all the points in V. One way to do this is to use a statistical approach. Particularly in the case where only data are known, statistical methods have certain advantages.

Regression Methods

The minimum variance unbiased estimator Z(x) is given by the conditional expectation of Z(x) given $Z(x_1),\ldots,Z(x_n)$. Unfortunately, to compute the conditional expectation requires knowing the joint distribution of Z(x), $Z(x_1),\ldots,Z(x_n)$. In the special case of joint Normality the conditional expectation is easily computed and is given by the regression of Z(x) on $Z(x_1),\ldots,Z(x_n)$. That is,

$$Z^*(X) = \sum_{[i=1,\ldots,n]} \lambda_i Z(x_i) \tag{23}$$

This suggests the use of such an estimator even when the joint distribution is unknown. Since the estimator is linear in the coefficients they can be determined by requiring that $Z^*(x)$ be unbiased and have minimal error variance. Eq. 23 is known as the kriging estimator. If Z(x) is regressed on the $f_j(x)$ one obtains the well-known Trend Surface estimator. Using the form of the estimator given in Eq. 23 and using the variogram to quantify the spatial correlation the following linear system of equations is obtained

$$\sum_{[i=1,\ldots,n]} \lambda_i \gamma(x_k - x_i) + \sum_{[j=0,\ldots,p]} \mu_j f_j(x_k) = \gamma(x_k - x); \quad k = 1,\ldots n \tag{24a}$$

$$\sum_{[i=1,\ldots,n]} \lambda_j f_j(x_i) = f_j(x); \quad j = 0,\ldots,p \tag{24b}$$

The coefficient matrix is invertible and hence the system has a unique solution. The μ_j's are Lagrange multipliers introduced in the optimization step because of the unbiasedness condition. The part of the system given in Eq. 24b is obtained by imposing the unbiasedness condition. Once again for a given kernel function $\gamma(h)$ and functions $f_j(x)$ the estimator is uniquely determined, but in this case $\gamma(h)$ and functions $f_j(x)$ are modeled and chosen using data. Note, however, that the system will have a unique solution for any choice of a valid variogram and linearly independent functions. By the use of elementary linear algebra, Eq. 23 can be shown to be equivalent to Eq. 20 and the system Eq. 24a and 24b is equivalent to the system given by Eq. 21a and 21b. In the case that there is no spatial correlation, i.e., the variogram model represents a lack of correlation (known as a "pure nugget" model), then the values of the estimator given by Eq. 23 will coincide with those of the Trend Surface interpolator except at the data locations. In this case there is no local variation, only regional variation and a noise term.

To estimate spatial integrals, the form of the estimator remains the same, but the coefficients are obtained by a different set of equations, i.e., several simple changes are made in Eq. 24a and 24b. On the right hand side of Eq. 24a replace $\gamma(x_k - x)$ by

$$\gamma(x_k, V) = (1/V)\int_V \gamma(x_k - x)\, dx \tag{25}$$

and replace $f_j(x)$ in Eq. 24b by

$$f_j(x, V) = (1/V)\int_V f_j(x)\, dx \tag{26}$$

For additional details on both the deterministic and regression methods see Myers (1991b, 1994a, 1994b). The estimator given by Eq. 23 and the system given by Eq. 24a and 24b are both easily adapted to the use of non-point data as well. The same result is obtained if Eq. 23 is used to obtain a value at each point in V and then these are averaged, although this would obviously take more computer time than direct estimation.

Estimation Variances

The error of estimation corresponding to Eq. 23 is given by $Z^*(x) - Z(x)$; the variance of this error is a quadratic form in the λ_i's. Substituting the solution from Eq. 24a and 24b back into this quadratic form gives

$$\sigma^2 = \sum_{[i=1,\ldots,n]} \lambda_i \gamma(x - x_i) + \sum_{[j=0,\ldots,p]} \mu_j f_j(x) \tag{27}$$

It is tempting to want to use this variance to construct confidence intervals for the errors, but some caution should be exercised since this variance does not depend on the data values (except as the data are used to model the variogram). Even if the errors are assumed to be Normally distributed this does not result in true confidence intervals. However, it is useful to construct a contour plot of these minimized error variances since this plot indicates the (spatial) relative reliability of the interpolated values. It is much more difficult to associate a measure of relative reliability for the interpolating function given by Eq. 20.

When Eq. 23 is used to directly estimate spatial integrals then an additional term will appear in the minimized error variance, which then becomes

$$\begin{aligned}\sigma_V^2 = &\sum_{[i=1,\ldots,n]} \lambda_i \gamma(x_i, V) \\ &+ \sum_{[j=0,\ldots,p]} \mu_j f_j(x, V) - (1/V^2)\int_V \int_V \gamma(u - w)\, du\, dw\end{aligned} \tag{28}$$

where the additional term is the same as Eq. 15a. See also Eq. 25 and 26. It is not surprising that the minimized error variance (usually called the kriging variance) is less for estimating spatial integrals than for interpolating at one point.

Software

Most standard statistical packages include trend surface analysis and some include the Thin Plate Spline but do not allow for a choice of the kernel function in Eq. 20. ARC/INFO incorporates some options for geostatistical analysis but does not provide the full flexibility described above. The EPA has released a public domain package, Geo-Eas, that allows for estimating and modeling variograms as well as kriging and the construction of contour plots. Geo-Eas only includes the case of m(x) a constant. A tutorial on the use of Geo-Eas in the analysis of lead contaminanted soil is given in Myers (1991b). In general the GIS packages that are readily available do not incorporate spatial statistical tools.

Non-Linear Transforms

In statistics it is common to use a non-linear transformation to equalize variances or obtain Normality. When treating spatial data such transformations may solve some problems but may also create others particularly when considering spatial variation and non-point data. The effect of such transformations on non-point data is considered in Myers (1993a). There are two transformations that are of special interest, however. In the case of logNormality, a logarithmic transformation is useful. Let

$$W(x) = \mathrm{Ln}\{Z(x)\} \tag{29}$$

Using the transformed data $W(x_1),\ldots, W(x_n)$ the variogram of W could be modeled and the counterpart of Eq. 23 used to estimate W(x); then Z(x) could be estimated by

$$Z^{**}(x) = \exp\{W^*(x)\} \tag{30}$$

However $Z^{**}(x)$ is biased, i.e., $E\{Z^{**}(x) - Z(x)\} \neq 0$. In the case of multivariate logNormality, the bias adjustment can be computed (which is multiplicative) but more generally it cannot. As an ad-hoc procedure, one might use $M \exp\{W^*(x)\}$ to interpolate onto a regular grid; then choose M so that the arithmetic average of these values matches the arithmetic average of the data. Note that if Z(x) is logNormal (whether multivariate logNormal or not) then the $Z(v_i)$ values cannot be logNormal.

The indicator transform is a second kind of non-linear transformation that is often useful for spatial data. Let

$$\begin{aligned} I(x; a) = 1, &\quad Z(x) \le a \\ 0, &\quad Z(x) > a \end{aligned} \tag{31}$$

The value "a" is often referred to as a "cutoff" value (because of applications in mining). For each choice of "a" there is an indicator transform. Note that $E\{I(a; x)\} = P[Z(x) \le a]$. Linear estimators such as given by Eq. 23 are not particularly useful

for estimating probability distributions. The indicator transform provides for a way to use Eq. 23 to do this. One disadvantage of using the indicator transform is that it will be necessary to use a number of values of "a" depending on degree of smoothness desired in the resulting estimated probability distribution function, hence requiring the modeling of a variogram for each cutoff value and solving the system of equations given by Eq. 24a and 24b for each cutoff. At a non-data point

$$I^*(x; a) = \sum_{[i=1,\ldots,n]} \lambda_i I(x_i; a) \tag{32}$$

is an estimator of I(x; a). Whereas the transformed data are all 0's or 1's, Eq. 32 does not ensure the same for I*(x; a) since the weights are not restricted between 0, 1. It is somewhat more useful to consider estimating

$$\phi(v; a) = (1/v)\int_v I(x; a)\, dx \tag{33}$$

which might be interpreted as the proportion of v for which the point values are less than or equal to a. By estimating Eq. 33 for multiple choices of a, a better appraisal of the variability of point values within V is obtained than is given by Eq. 6. Both Eq. 6 and Eq. 33 provide a way to classify pixels or subregions within an image. An estimator of the same form as that in Eq. 32 can be used to estimate $\phi(v; a)$ but a slightly modified set of equations is used to determine the coefficients in the estimator. Because there is a dependence between the I(x; a) for different values of a and similarly for the $\phi(v; a)$, some form of joint estimation is to be preferred. One approach to this and other joint estimation problems is considered in the next section.

SPATIAL INTERVARIABLE CORRELATION

In many applications there are multiple variables of interest, which might correspond to multiple data layers. Let these be denoted by $Z_1(x),\ldots, Z_m(x)$. Then in addition to the spatial variability and correlation of each $Z_k(x)$, one can consider the correlation of pairs of these as well as the spatial intervariable correlations.

The use of principal components analysis (PCA) to remove noise from multiband images is a well-known technique and is based on the premise that the noise term corresponds to a relatively small part of the total variance. Correspondence analysis (CA) might also be used for the same purpose, since it has the advantage of providing an R and a Q mode analysis simultaneously. An example of an application of CA to an environmental data set is given in Avila and Myers (1991). However, both PCA and CA only incorporate intervariable correlations. That is, they consider similarities and dissimilarities between $Z_j(x)$, $Z_k(x)$; they do not consider such between $Z_j(x + h)$, $Z_k(x)$. There are at least three ways to quantify the latter correlations. If the $Z_j(x)$ are second order stationary then use

$$C_{jk}(h) = \mathrm{Cov}\{Z_j(x+h), Z_k(x)\} \tag{34}$$

Note that although $C_{jk}(h) = C_{kj}(-h)$, $C_{jk}(h)$ is not symmetric. Under a slightly weaker stationarity assumption there are two possibilities; the cross-variogram and the pseudo cross-variogram

$$\gamma_{st}(h) = 0.5\,\mathrm{Cov}\{Z_s(x+h) - Z_s(x), Z_t(x+h) - Z_t(x)\} \tag{35}$$

In this case $\gamma_{st}(h) = \gamma_{st}(-h)$, i.e., the cross-variogram is symmetric. Alternatively there is the pseudo cross-variogram

$$\gamma_{st}(h) = 0.5\,\mathrm{Var}\{Z_s(x+h) - Z_t(x)\} \tag{36}$$

which is not symmetric. Both of the latter generalize the variogram given in Eq. 11. A vector form of Eq. 23 known as the cokriging estimator allows joint estimation of the various components. In the counterparts of Eq. 24a and 24b the variograms must be replaced by matrix valued functions. The diagonal entries are variograms and the off diagonal entries are cross-variograms (or pseudo cross-variograms). In this case it is necessary for the matrix valued function to satisfy certain conditional positive definiteness conditions. Details are found in Myers (1982, 1988, 1991c, 1991d, 1992). This estimator and the corresponding system of equations are adaptable to the possibility that data is missing for some components at some data locations; see Myers (1988). For an application to an environmental data set, see Myers (1989b). Carr and Myers (1984) and Glass et al. (1988) consider applications of cokriging to remotely sensed data and to multiband images. Modification for non-point data or for the estimation of spatial averages is analogous to that for the single variable case. Because of the greater difficulty in modeling cross-variograms (or pseudo cross-variograms) in addition to the separate variograms, it may be desirable to consider ways to avoid this. One possible solution is to use PCA to generate "uncorrelated" components; then these new components are treated separately, variograms are estimated and modeled. After interpolation, the original variables are reconstructed. There are at least three disadvantages: first, the lack of correlation generated by the use of PCA does not guarantee that the cross-variograms are identically zero since the latter is a model property; secondly, while the minimized estimation variances can be computed for the interpolated "components" there is no way to utilize these to generate minimized estimation variances for the interpolation of the original variables; finally, PCA does not explicitly take into account the support of the data as is possible in the statistical interpolation process. Nevertheless it may be useful to examine the cross-variograms (or pseudo cross-variograms) of the principal components generated by PCA. It is also possible to utilize PCA to aid in modeling the cross-variograms. Myers and Carr (1984) give an example to compare these two approaches.

Diagonalization: Generalized PCA

Recall that the "components" generated by PCA are linear combinations of the original variables. Write the original variables as a $1 \times m$ vector, $\mathbf{Z}(x) = [Z_1(x),\ldots, Z_m(x)]$. Suppose that each component is a linear combination of uncorrelated components, i.e,

$$\mathbf{Z}(x) = \mathbf{Y}(x)A \tag{37}$$

where $\mathbf{Y}(x)$ is $1 \times q$ vector and A is a $q \times m$ matrix. Since the components of $\mathbf{Y}(x)$ are uncorrelated, the matrix variogram of $\mathbf{Y}(x)$ is diagonal (the off diagonal entries are all zeros). Then the matrix variogram of $\mathbf{Z}(x)$ is given by

$$\gamma_Z(h) = A^T\gamma_Y(h)A \tag{38}$$

In PCA it is only necessary to diagonalize a single constant matrix (i.e., find the eigenvalues and the corresponding eigenvectors) but $\gamma_Z(h)$ is not a single matrix. In general given a matrix function there may not be a matrix A leading to the form given in Eq. 38. Even if $q = m$, A may not be invertible; hence finding A is not quite the same as diagonalizing $\gamma_Z(h)$. In practice, $\gamma_Z(h)$ need only be computed for a finite number of different values of h and hence it is only necessary to generate a "common" diagonalization for these values. The assumption that such a diagonalization exists is a moderately strong one, and as yet easily checked sufficient conditions are not known. However, a near diagonalization may be possible. Let $h_1,\ldots,h_t$ be the values of h for which $\gamma_Z(h)$ must be computed. The objective is to find a matrix B such that

$$B^T\gamma_Z(h_1)B, \ldots, B^T\gamma_Z(h_t)B \tag{39}$$

are "nearly" diagonal. "Nearly" could be evaluated in terms of the sums of the squares of the off-diagonal entries in the matrices in Eq. 39. Switzer and Green (1984) used this approach to analyze the Silver Bell TM image data but only considered $h = 1$.

SMOOTHING

Interpolation

All interpolation techniques smooth the data, i.e., the variance of the set of interpolated values is always less than the variance of the original data. This will occur even if it is assumed that e(x) is identically zero in Eq. 1. This property can

be used to smooth an image even when no interpolation is necessary, for example when all the pixel values are known in an image or when the values are known at every point of a regular grid. As an ad-hoc smoother, an image is sometimes subjected to a low-pass filter by replacing each original value by a weighted average of the values of nearby pixels (or grid points). However, the weighting in the average is somewhat arbitrarily determined, so an alternative is to use the variogram to determine the weighting. This approach has been discussed by Ma and Royer (1988) wherein each data value in turn is temporarily suppressed and the value at that location is interpolated from nearby locations. By using the variogram, which is modeled from the data, the filter is determined by the spatial variability of the specific image. By using the vector interpolation technique described above, this form of smoothing could be applied to a multiband image, smoothing all bands simultaneously, and incorporating the intervariable spatial variability.

Noise Removal

If in the model given by Eq. 1, Y(x) is taken to be identically zero then the smoothing objective is to remove the e(x) term. Ordinarily the noise component is assumed to be uncorrelated spatially and the most common method for its removal is by some form of regression of Z(x) on m(x).

It was noted above that PCA is sometimes used to "remove" noise from a multiband image by deleting those principal components with small eigenvalues, i.e., small variances, then reconstructing the image using only the remaining components. There is a degree of subjectivity in determining how small is small. Note that the components generated by PCA are not affected by an arbitrary rearrangement of the locations of the pixels in the image, and hence, important information reflected in the image is not utilized.

A Combination

In the preceding discussion of interpolation, it was assumed that the error term e(x) in Eq. 1 was identically zero. A more general objective might be to estimate $Y(x) + m(x)$ at data locations and to interpolate this sum at non-data locations. The smoothing spline is a modification of the thin plate spline that combines both interpolation and noise removal. Since the thin plate spline is a special case of the interpolator given in Eq. 20 it is reasonable to expect that a generalization of the smoothing spline would have a similar form, but where the coefficients are obtained from a modification of Eq. 21a and 21b. It is somewhat easier to do this by beginning with the statistical form given in Eq. 23. This is discussed in Myers (1994a) and results in only a slight modification of the equations in Eq. 24a and 24b. The variance of the noise term must be known or estimated from the data. The modified form of Eq. 21a and 21b is then obtained from the modified version of Eq. 24a and 24b. This may be extended to the multiband case where the vector interpolator discussed above is used.

FRACTALS AND FRACTAL DIMENSION

As was noted earlier the exponent in a power model variogram is related to the fractal dimension. This relationship has been exploited in both Carr and Benzer (1991) and Jaggi et al. (1993) as a means of estimating the fractal dimension. This relationship can be interpreted intuitively in a useful manner and it can also be related to the identity given in Eq. 9. The theory of fractals is built on the concept of self-similarity, whereas the fractal dimension can be interpreted as relating to the roughness of a curve or surface, Lam and Quattrochi (1992), Jaggi et al. (1993). Let x be a point in k-dimensional space, V a region in this space, and Z(x) a function defined on V; then the pair (x, Z(x)) determines a "surface" in k+1-dimensional space. One measure of the "roughness" of this surface is simply

$$(1/V)\int_V [Z(x) - Z(V)]^2 \, dx$$

where $Z(V) = (1/V)\int_V Z(x)\,dx$ which are simply Eq. 2 and 4. Z(V) might be interpreted as the "area" of the surface measured at a minute scale. Alternatively, Z(V) is the average of local areas, i.e., if V is partitioned into disjoint, congruent subelements v_i, then $Z(V) = (v/V)\Sigma Z(v_i)$. Think of these as the measurements made in a ruler-step process where each $Z(v_i)$ is not known exactly but is estimated by units of size v_i. In particular consider a 1-dimensional analog, i.e., x in 1-dimensional space. Then V is an interval; suppose now that V is partitioned into subintervals of length v and let $x_1,\ldots,x_m$ be the end-points of the subintervals, m = V/v. Then the length of the curve is approximated by

$$\sum\left[(x_i - x_{i+1})^2 + (Z(x_i) - Z(x_{i+1}))^2\right]^{0.5}$$

which is clearly related to

$$\sum\left[(Z(x_i) - Z(x_{i+1}))^2\right]^{0.5}$$

and in turn to

$$\sum(Z(x_i) - Z(x_{i+1}))^2$$

which is essentially the sample variogram for lag v. The variation in the measurements of the $Z(v_i)$'s from the true values might be represented by interpreting Z(x) as a random function. It is known that a fractional Brownian motion has a power model variogram. Now the identity in Eq. 9 can be reinterpreted. Averaging Z(x) over the elemental units v "smoothes" the surface but leaves discontinuities; the triangular prism method described in Jaggi et al. (1993) can be thought of as a way

to obtain the smoothing but retain continuity. As the surface is smoothed, the fractal dimension decreases; this is seen in the identity because the regularized variogram Eq. 17 is smoothed as the regularizing unit is increased in size.

SUMMARY

The quantification of spatial variability and spatial correlation is seen to be the key to relating different data models as well as multiscaled data layers. Within-pixel and between-pixel variation are related to the total within-image variation. This quantification may be considered for each layer separately or it may be determined for multiband images. A distinction is made between empirical measures and theoretical models of spatial variability and spatial correlation; theoretical models are important for interpolation and for estimation of spatial averages. The relationship between and the distinction between interpolation and smoothing are discussed.

REFERENCES

Avila, F. and Myers, D.E., Corrrespondence analysis applied to environmental data sets: a study of Chatauqua lake sediments, *Chemometr. Intell. Lab. Sys.*, 11, 229–249, 1991.

Biondi, F., Myers, D.E., and Avery, C., Geostatistically modeling of stem size and increment in an old growth forest, *Can. J. For.*, 24, 1354–1368, 1994.

Burrough, P., 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, 1983a.

Burrough, P., Multiscale sources of spatial variation in soil. II. A non-Brownian fractal model and its application in soil survey, *J. Soil Sci.*, 34, 577–597, 1983b.

Carr, J.R. and Benzer, W.B., On the practice of estimating fractal dimension, *Math. Geol.*, 23, 945–958, 1991.

Carr, J.R. and Myers, D.E., Application of the theory of regionalized variables to the spatial analysis of LANDSAT data, in *Proc. 9th Pecora Symp. Remote Sensing*, IEEE Computer Society Press, 1984.

Cheng, Q. and Agterberg, F.P., Multifractal modeling and spatial point processes, *Math. Geol.*, 27, 831–846, 1995.

Clark, W.A.V. and Avery, K.L., The effects of data aggregation in statistical analysis, *Geogr. Anal.*, 8, 428–438, 1976.

Davis, F.W., Quattrochi, D.A., Ridd, M.K., Lam, N.S.-N., Walsh, S.J., Michaelson, J.C., Franklin, J., Stow, D.A., Johannsen, C.J., and Johnston, C.A., Environmental analysis using integrated GIS and remotely sensed data: some research needs and priorities, *Photogr. Eng. Remote Sensing*, 57, 689–697, 1991.

Fabbri, A.G., van der Meer, F. D., Valenzuela, C., and Kushigbor, C. A., Shape analysis and multispectral classification in geological remote sensing, *Math. Geol.*, 25, 773–794, 1993.

Glass, C.E., Carr, J.R., Yang, H.M., and Myers, D.E., Application of spatial statistics to analyzing multiple remote sensing data sets, in *Geotech. Appl. Remote Sensing Remote Data Trans., ASTM STP 967*, Johnson, A.I. and Peterson, C.B., Eds., 138–151, 1988.

Goodchild, M.F., Fractals and the accuracy of geographic measurements, *J. Math. Geol.*, 12, 85–98, 1980.

Goodchild, M., Haining, R., and Wise, S., et al., Integrating GIS and spatial data analysis: problem and possibilities, *Int. J. Geogr. Info. Sys.*, 6, 407–423, 1992.

Herman, G.T., Lent, A., and Rowland, S.W., ART: mathematics and applications, *J. Theoret. Biol.* 42, 1–32, 1973.

Jaggi, S., Quattrochi, D.A., and Lam, N.S.-N., Implementation and operation of three fractal measurement algorithms for analysis of remote sensing data, *Comp. Geosc.*, 19, 745–767, 1993.

Lam, N.S.-N. and Quattrochi, D.A., On the issues of scale, resolution and fractal analysis in the mapping sciences, *Prof. Geogr.* 44, 88–98, 1992.

Ma, Y.Y. and Royer, J.J., Local geostatistical filtering: application to remote sensing data, *Sci. de la Terre*, 27, 17–36, 1988.

Meijerink, A.M.J., de Brouwer, H.A.M., Mannaerts, C.M., and Valenzuela, C.R., *Introduction to the Use of Geographic Information Systems for Practical Hydrology*, UNESCO IHP-IV, Paris, 1994.

Modjeska, J.S. and Rawlings, J.O., Spatial correlation analysis of uniformity data, *Biometrics*, 39, 373–384, 1983.

Myers, D.E., Matrix formulation of cokriging, *Math. Geol.*, 14, 249–257, 1982.

Myers, D.E., Multivariate geostatistics for environmental monitoring, *Sci. de la Terre*, 27, 411–427, 1988.

Myers, D.E., To be or not to be...stationary: that is the question, *Math. Geol.*, 21, 347–362, 1989a.

Myers, D.E., Borden field data and multivariate geostatistics, in *Hydraulic Engineering*, Ports, M.A., Ed., American Society of Civil Engineers, 795–800, 1989b.

Myers, D.E., Spatial-Temporal Geostatistical Modeling in Hydrology, in *Proc. Int. Workshop Geostat. Meth. Recent Dev. Appl. Surface Subsurface Hydrol.*, Karlsruhe, Germany, 1990.

Myers, D.E., On variogram estimation, in *Proc. 1st Int. Conf. Stat. Comp.*, Vol II, American Sciences Press, 261–281, 1991a.

Myers, D.E., Interpolation and estimation with spatially located data, *Chemometr. Intell. Lab. Sys.*, 11, 209–228, 1991b.

Myers, D.E., Multivariate, multidimensional smoothing, in *Spatial Stat. Imaging: Proc. AMS-IMS-SIAM Joint Summer Res. Conf.*, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, CA, 275–285, 1991c.

Myers, D.E., Pseudo-cross variograms, positive definiteness and cokriging, *Math. Geol.*, 23, 805–816, 1991d.

Myers, D.E., Cokriging, radial basis functions and the role of positive definiteness, *Comp. Math. Appl.*, 24, 139–145, 1992.

Myers, D.E., Change of support and transformations, in *Geostatistics for the Next Century*, Dimitrakopoulos, R., Ed., Kluwer Academic, Dordrecht, 253–258, 1993a.

Myers, D.E., Spatial Statistics, in *Environmental Modeling with GIS*, Goodchild, M.F., Parks, B.O., and Steyaert, L.T., Eds., *Proc. 1st Int. Conf./Workshop Integrating GIS Environ. Modeling*, Oxford University Press, Oxford, 401–404, 1993b.

Myers, D.E., Statistical methods for interpolation of spatial data, *J. Appl. Sci. Comp.*, 1, 283–318, 1994a.

Myers, D.E., Spatial interpolation: an overview, *Geoderma*, 62, 17–28, 1994b.

Myers, D.E. and Carr, J.R., Cokriging and principal component analysis: bentonite data revisited, *Sci. de la Terre*, 21, 65–77, 1985.

Nelson, M., Felix-Gastelum, R., Orum, T.V., Stowell, L.J., and Myers, D.E., Geographic information systems and geostatistics in the design and validation of regional plant virus management systems, *Phytopathology*, 84, 898–905, 1994.

Parker, H., The volume-variance relationship: a useful tool for mine planning, *Eng. Mining J.*, October, 106–123, 1979.

Pettitt, A.N. and McBratney, A.B., Sampling designs for estimating spatial variance components, *Appl. Stat.*, 42, 185–209, 1993.

Pickett, S.T.A. and Cadenasso, M.L., Landscape ecology: spatial heterogeneity in ecological systems, *Science*, 269, 331–334, 1995.

Rhodes, H.R. and Myers, D.E., GRASS used in the Geostatistical Analysis of Lakewater data from the ELS-Phase I, in *Environ. Modeling with GIS*, Goodchild, M.F., Parks, B.O., and Steyaert, L.T., Eds., *Proc. 1st International Conf./Workshop Integrating GIS Environ. Modeling*, Oxford University Press, Oxford, 436–448, 1993.

Root, T.L. and Schneider, S.H., Ecology and climate: research strategies and implications, *Science*, 269, 334–341, 1995.

Rouhani, S. and Myers, D.E., Problems in space-time kriging of hydrogeological data, *Math. Geol.*, 22, 611–623, 1990.

Switzer, P. and Green, A.A., *Min/Max Autocorrelation Factors for Multivariate Spatial Imagery*, Tech. Rpt. No. 4, Dept. Statistics, Stanford University, 1984.

Townshend, J.R.G. and Justice, C.O., Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations, *Int. J. Remote Sensing*, 9, 187–236, 1988.

Van Wesenbeck, I.J. and Kachenoshi, R.G., Spatial scale dependence of in situ solute transport, *Soil Sci. Soc. Am. J.*, 55, 3–7, 1991.

Warrick, A.W. and Myers, D.E., Optimization of sampling locations for variogram calculations, *Water Resour. Res.*, 23, 496–500, 1987.

Webster, R. and Oliver, M.A., How large a sample is needed to estimate the regional variogram adequately? in *Geostatistics Troia '92*, Soares, A., Ed., Vol. 1, Kluwer Academic, Dordrecht, 155–166, 1993.

Woodcock, C.E., Variograms and Spatial Variation in Remotely Sensed Images, in *International Geoscience and Remote Sensing Symposium*, Univ. Mass. Press, Amherst, 1078–1083, 1985.

Woodcock, C.E., Strahler, A.H., and Jupp, D.L.B., The use of variograms in remote sensing: I. Scene models and simulated images, *Remote Sensing Environ.*, 25, 323–348, 1988.

Yates, S.R. et al., Spatial variability of remotely sensed surface temperatures at field scale, *Soil Sci. Soc. Am. J.*, 52, 40–45, 1988.

Zhang, R., Warrick, A., and Myers, D.E., Variance as a function of sample support size, *Math. Geol.*, 22, 107–121, 1990.

Zhang, R., Myers, D.E., and Warrick, A.W., Soil heterogeneity and plot shape effect, *Geoderma*, 62, 183–197, 1994.