Introduction

In everyday life, information is often presented in the form of a "map". A typical example would be a map showing roads, cities, points of interest, etc. in a particular state. Another example would be a map that shows elevation above sea level for a given region. In both cases, for all practical purposes, the maps are "deterministic". That is, the information represented in the map is known with certainty. However, in many cases, the information needed to produce deterministic maps is not available — it is either too expensive or too time consuming (or both) to collect all the necessary data.

For example, we may want to know how rainfall varies over a given region of the earth. However, we cannot possibly measure the amount of rainfall everywhere within even a relatively small area (say a few square miles). Similarly, the thickness of a coal seam may be of great interest because of its economic importance for a mining company. But, since the coal seam is buried within the ground (often at great depths), there is no practical way of measuring the seam thickness at every point in the deposit.

As in many other problems, sampling can come to the rescue. For mapping rainfall, sampling stations (which essentially consist of "buckets") can be located throughout the region. For mapping the thickness of a coal seam, sampling entails drilling "core" holes, removing the core and measuring the thickness of the seam. The samples are located ideally on a (more or less) regularly spaced grid throughout the area of the deposit. These problems differ from the more routine applications of sampling in that the location of the sample in space plays a major role. Therefore, the usual analysis will generally not work.

In what follows, I will try to outline a general approach to producing maps based on limited sample information. It should be clear that the problem of producing estimates for a region in space away from any sample is completely general and can be applied to many areas outside of meteorology and mining. The basic theory behind this approach was developed by Krige and by Matheron in the early 1950's. Much of the subsequent development and refinement has been carried on abroad and was generally unknown in the United States until a few years ago.

General Problem Definition

The common goal in map making and related activities using limited sample data can be stated as follows:

Given a discrete sample measurements \( Z(X_1), Z(X_2), \ldots, Z(X_n) \) located in space at positions \( X_1, X_2, \ldots, X_n \) for some variable of interest \( Z \), estimate:

1) the value of the variable \( Z \) at point \( X_0 \), i.e., \( Z(X_0) \)
2) the average value of \( Z \) over the region \( D \) or \( Z(D) \), and
3) the reliability of these estimates.

Note that the region \( D \) is completely general. It could be any irregular shape or possibly a square or rectangular block. To produce a map from a finite set of discrete samples, a set of points on a regular grid is first estimated. (The method of estimation will be of primary concern in this discussion.) Then contour lines are created by a suitable method of interpolation. To further simplify the discussion, we will generally concentrate on point estimation and exclude average estimates for regions.

For our discussion, we will assume that the samples are point samples. (Often in the literature, point samples are called punctual samples. Also, the general term support is used in place of sample.) This assumption is generally justified if the region we are estimating is large compared to the size of the samples. This will allow a less complicated presentation of the basic ideas of spatial statistics. In cases where the simplification of point samples cannot be justified, the theory easily generalizes.

Ad Hoc Methods Versus Spatial Statistical Theory

Since there are virtually an infinite number of possible estimators, a question arises: what is the "best" type of estimator to use? To determine "best" we need some acceptable criterion.

The criterion that is generally used is the estimation variance. Estimation variance is a measure of how well the predicted value of the point or region agrees with the actual point or region on average. Besides comparing estimators in terms of estimation variance, it is highly desirable that the "best" method be a product of a general theory that describes and explains why and to what extent a particular method works.

The theory of spatial statistics presents a method of estimation generally known as kriging. The justification and interpretation of the method is determined by what is known as a structural analysis. We will see that structural analysis is simply and essentially an in-depth analysis of the data that allows us to determine the spatial law that governs a particular process.

Methods such as inverse distance, polygonal, triangulation, and trend surfaces, have all been used extensively in the past. All of these methods share a number of properties. For example, they are all more or less "ad hoc"
procedures. That is, they are not based on any general theory of data analysis. Most of these methods, when viewed through a general theory of data analysis, are seen to be much less desirable than the methods of spatial statistics. Many of these methods in some ways are counter-intuitive to the notions developed later in this discussion.

How to avoid being over-critical of these ad hoc methods, it must be pointed out that they do share two important properties with the spatial statistical method of kriging. All of these methods are linear estimators. In addition has a component that is a random situation. That is to say, the values of the variable are less alike the closer they are, a

Characterizing Spatial Correlation
We have already mentioned the usual assumption of most methods used to produce point and region estimates of variables distributed in space: as two locations are closer together, the values of the variables tend to be more alike. One of the major drawbacks of the ad hoc methods is that the degree and extent of this assumed "alikeness" is independent of the nature of the variable and the sample collected. If we are to have a method that is more than ad hoc, we must be able to determine for a given variable and data set what the pattern of this "alikeness" is. We will now develop the method used by spatial statistics. Later we will discuss the connection with the more common and equivalent approaches used in more traditional areas of statistics.

Regionalized Variables
Often when we collect samples, the actual location of the sample is ignored in the analysis. For example, to calculate average height of males in the United States, the actual location of each person in the sample is not usually needed. However, if we are interested in the population density for the country, then the locations will be important.

Any variable that, for analysis purposes, is distributed in space is called a regionalized variable. For example, the ground level (relative to sea level) may be denoted as g(x, y) to show the dependence on the location (x, y). The idea of a regionalized variable does not necessarily involve any probabilistic or statistical interpretation. For the most part, the ground level is perfectly well determined at a given point in time.

However, for many variables of interest, a statistical interpretation is useful and possible. In any application, the assumption is tentatively made that the regionalized variable has a component that is a random function. If this is plausible, it is borne out in the subsequent analysis. This approach is similar to other statistical analyses. For example, in linear regression analysis, we may tentatively assume a linear dependence between variables and independent and normally distributed errors. All of these assumptions are checked and if needed, they are appropriately modified.

The Empirical Variogram Function
How can we characterize the "alikeness" of measurements separated by a certain distance in space? The idea of correlation immediately arises. For example, an autocorrelation function based on distance (rather than time) would seem appropriate. This is indeed a correct and logical approach. However, for a number of reasons (some accidental), the approach of the early developers of spatial statistics was to work with essentially the inverse of correlation. That is, the variation of sample values a given distance apart is studied. (We will see the connection between these two approaches and their equivalence later.)

An illustration of the fact that "alikeness" and variation are inverses is shown above. As the points in space are separated by larger distances and the "alikeness" decreases, the "variation" will increase. In precisely quantifying these terms, the correlation (or covariance) replaces the concept of "alikeness" and the variance replaces the concept of "variation". The important point to note is that however we describe the relation between the values of a variable at two different locations, this relation depends on the distance between the two locations. We shall, however, generally work in terms of "variation". Therefore, we define the empirical variogram function (Journel and Huijbregts, 1978, p. 207) as:

\[ \gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [g(x_i + h) - g(x_i)]^2 \]

In words, this function can be interpreted as the variance of the n(h) differences between pairs of sample measurements separated by a distance of h, falling in the same direction a. This computation is tedious (if done by hand) but much simpler than it may appear. Suppose we have a set of measurements essentially located on a regular grid as shown below. Let us assume that the samples are generally spaced one unit apart. To calculate the variogram function in the horizontal direction (m=0) for h-1, we must calculate all the differences between pairs of samples one unit apart in the horizontal direction as shown below.
Thus, $\gamma(\alpha=0, h=1) = \frac{(-1)^2 + (-1)^2 + (-1)^2 + (-1)^2}{2} = \frac{(-1)^2 + (-1)^2}{2}$, or 1.32 units squared.

In other words, the empirical variogram function is simply the average squared difference between sample measurements a distance $h$ units apart in a particular direction $\alpha$. (The average squared difference in this case can be shown to be exactly equivalent to the variance of the differences of measurements a distance $h$ units apart.)

Similarly, we can calculate the variogram function for $h=2$ and $\alpha=0$. Thus, $\gamma(\alpha=0, h=2) = \frac{(-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2}{2} = \frac{(-1)^2 + (-1)^2}{2}$, or 4.06 units squared.

In a similar manner, $\gamma$ can be calculated for any direction $\alpha$.

If we carried out this procedure for a number of $h$ values with $\alpha$ held constant, we could then construct a graph of the variogram function in a given direction. This has been done for the two values already calculated and is shown below:

![Graph of $\gamma(\alpha=0, h)$]

Some regionalized variables will have variogram functions that are quite different depending on the direction. The regionalized variable is then said to be anisotropic. If the empirical variograms in different directions show roughly the same pattern, we can consider the regionalized variable to be isotropic. This can occur either because the spatial variation is the same in all directions, or the difference in different directions was relatively small and obscured by sampling variation. In either case, if we can consider the variograms in different directions as representing isotropic regionalized variables, they can be combined to produce an "average" variogram independent of direction.

A simple curved line has been fit to suggest what the value of $\gamma$ is for values of $h$ not equal to 1 or 2. For the population from which the samples have been chosen, there is a theoretical variogram function that "controls" the process that generates the samples. The reason for computing the empirical (or sample) variogram function is to get an idea of how this theoretical variogram behaves.

**Stationarity**

If we refer back to the graph of the variogram for $\alpha=0$, we notice that the variogram was interpolated to pass through zero. Because of the way $\gamma$ is calculated — on a fixed set of measured values — $\gamma(h=0)$ must be exactly equal to zero. The measured values are "fixed" because the regionalized variables that we will study will necessarily be concerned with can only be sampled once at a given location. (Many phenomena fit into this category.) In these cases, in order to use statistical methods to get an idea of the spatial law governing the regionalized variable, we must work under the assumption that the "alikeness" of observations will depend only on the relative distance between the locations of the observations, and not their absolute locations. This means that the random function part of the regionalized variable must be weakly stationary.

For phenomena that can be repeatedly sampled at the same location, the variance of the difference between measurements at two sample locations can be estimated directly from the observed differences between the paired observations. When the measured values can be obtained only once, the variance of the difference between observations at two locations $h$ units apart is estimated by observing similar pairs of observations $h$ units apart on (at least part of) the region being studied. For the latter approach to be valid, the assumption of weak stationarity must be reasonable.

**Structural Analysis**

Before we can hope to make optimal estimates for a regionalized variable and qualify the estimates with some quantitative measure of reliability, we must be able to at least approximately describe the spatial law that governs the particular regionalized variable. The steps necessary to do this make up what is called a structural analysis (Journel and Huijbregts, 1979, pp.12, 168-235). The first step after the initial sampling campaign and simple descriptive data analysis would be to compute an empirical variogram function. We would then attempt to fit an appropriate theoretical model to the observed variogram. We will describe this process and two frequently used theoretical variogram models.

**Theoretical Variogram Models**

The empirical variogram is based on a sample of observations from a certain population.
This population has some spatial law that controls the spatial correlation of the values of the regionalized variable and can often be summarized by a theoretical or population variogram. The empirical variogram will vary depending on the particular set of samples chosen. However, the empirical variogram will usually give at least an approximate estimate of the population variogram if the number of sample pairs n(h) for a given distance is greater than about 30 (Journel and Huijbregts, 1979, p. 194). Once the empirical variogram is constructed, we will then want to fit an appropriate model or function to the data that is our best estimate of the population variogram.

Now not every function can represent a variogram. Obviously, the values of the regionalized variable must be positive or zero. Also, y(0) must equal zero. Even beyond this, there are certain restrictions on possible models that exist because the variograms of the regionalized variable must be positive or zero. For example, not all polynomial functions can be variogram functions. We will study two possible models that are used frequently in practice. (For an elementary treatment on fitting theoretical variograms, see Clark, 1979, Chapter 2.)

Spherical Variogram Models

The spherical model (David, 1977, p. 102) is given by:

\[
y(h) = \begin{cases} 
0 & \text{if } h = 0 \\
C_0 + C \left( \frac{3h^2 - 1h^3}{2h^3} \right) & \text{if } 0 < h < a \\
C_0 + C & \text{if } h > a 
\end{cases}
\]

where \( C_0 \) is the nugget effect (i.e., the limiting value of \( y(h) \) as \( h \) approaches zero). The value \( C_0 + C \) is called the sill. The value \( a \) is called the range of the spherical variogram and is the length of distance necessary for the variance to stop increasing and become constant.

Models that tend to level off at a sill are often referred to as transition models. The values of a regionalized variable that can be represented as a transition model are uncorrelated if separated by a distance larger than the range \( a \). The sill value \( C_0 + C \) is approximately equal to the variance of point samples. The spherical model is probably one of the most often encountered and used models.

Pure Nugget Effect Model

When the regionalized variable is completely random, (either because it is very discontinuous or because of measurement error, or both), the appropriate variogram model is a pure nugget effect model (Journel and Huijbregts, 1979, pp. 152-153). This model is given by

\[
y(h) = \begin{cases} 
0 & \text{if } h = 0 \\
C_0 & \text{if } h > 0 
\end{cases}
\]

Theoretically, no matter how small \( h > 0 \) becomes, \( y \) still equals \( C_0 \). For most real phenomena, for very small positive values of \( h \), \( y \) usually does decrease. However, if \( y \) does decrease for \( h < m \), where \( m \) is very, very small, the actual situation is often adequately represented by a pure nugget effect model. Note that the pure nugget effect model occurs as a degenerative case for a number of models, including the spherical, when \( C \) equals zero.

Stationarity

We have already discussed the concept of stationary regionalized random variables. If we have a phenomenon that is stationary, then we can easily transform the variogram function into the equivalent autocovariance function by the following (Olness, 1975, p. 37):

\[ C(h) = C(0) - y(h) \]

where \( C(h) \) is the autocovariance function value at a distance of \( h \).

\( C(0) \) is the autocovariance at zero distance. Hence \( C(0) \) is the simple a priori variance of the random function and is equal to the sill value of the variogram function. \( C(0) \) exists if the phenomenon is stationary. We can derive the autocorrelation function by dividing the previous equation by \( C(0) \):

\[ \rho(h) = \frac{C(h)}{C(0)} = 1 - \frac{y(h)}{C(0)} \]

Stationary phenomena are characterized by having variogram functions that tend to be stable (constant) for large distances \( h \). On the other hand, if the variogram tends to increase as \( h \) increases, then the phenomenon is non-stationary. (Although non-stationary phenomena occur frequently, space does not allow a complete treatment here.)

Linear Estimator

A linear estimator is simply a weighted average where the weights sum to one. That is, if \( \hat{Z}(x_0) \) denotes a linear estimate of the true value \( \hat{z} \) at location \( X_0 \), and \( w_i \) denotes the ith weight, then

\[
\hat{Z}(x_0) = \sum w_i \hat{Z}(x_i),
\]

with \( \sum w_i = 1 \).
All the differences between linear methods only involve choosing the weights \( w_i \). The general problem of estimating an unknown point or region value from a set of sample observations is often referred to as the process of extension. That is, we must in some manner "extend" the known sample values into unknown areas.

### The Kriging Estimator

The kriging estimator is a linear estimator. The weights for the ad hoc procedures are for the most part mathematically developed on purely geometric considerations. The ad hoc procedures do not use any information from the structural analysis. In contrast, kriging estimators depend on the particular phenomenon because the weights take advantage of the structural information. The weights are also constructed so that the reliability of the estimates is maximized.

To see how this is accomplished, imagine trying to estimate \( Z(X_0) \) from the four observations \( Z(X_1), Z(X_2), Z(X_3), \) and \( Z(X_4) \) illustrated earlier (see General Problem Definition). First, we would want to estimate the theoretical variogram function. As we know, the variogram summarizes much of the spatial correlation information concerning the phenomenon. (With only four observations it would be impossible to reliably estimate the variogram. In most real problems, we would ordinarily have to have more observations. However, if we cannot estimate the variogram, then it is very dangerous to assume that our estimates are reliable. For this illustration, we will assume that we know the variogram function, possibly because we have other observations that are not considered.)

Some of the observations are correlated to the unknown point \( Z(X_0) \). We want to know which ones are correlated and how strongly. Each of the observations is possibly correlated with other observations. This must also be taken into account. The variogram function gives this information. For instance, the variogram tells us the variance \( \gamma(h) \) of the difference between \( Z(X_0) \) and \( Z(X_1) \) where \( h \) is the distance between \( X_0 \) and \( X_1 \). This can easily be converted to a covariance (if the phenomenon is stationary) by a formula that we have seen before:

\[
C(h) = C(0) - \gamma(h).
\]

For example, suppose we want to know the covariance between sample 1 and sample 2. Let us call this \( \sigma_{12} \). First, find the distance \( d \) between the two sample locations. Then find the value of the theoretical variogram \( \gamma(d) \) that corresponds to this distance \( d \), i.e., \( \gamma(d) \). Then using the above formula, \( \sigma_{12} = C(d) - C(0) - \gamma(d) \).

One way of computing the kriging weights is to set up several simultaneous equations involving variances and covariances which when solved yield the weights with minimum estimation variance. For the above example with four observations, the equations are (David, 1979, p. 204):

\[
\begin{align*}
\sigma_{11}w_1 + \sigma_{12}w_2 + \sigma_{13}w_3 + \sigma_{14}w_4 + \mu &= \sigma_{01} \\
\sigma_{21}w_1 + \sigma_{22}w_2 + \sigma_{23}w_3 + \sigma_{24}w_4 + \mu &= \sigma_{02} \\
\sigma_{31}w_1 + \sigma_{32}w_2 + \sigma_{33}w_3 + \sigma_{34}w_4 + \mu &= \sigma_{03} \\
\sigma_{41}w_1 + \sigma_{42}w_2 + \sigma_{43}w_3 + \sigma_{44}w_4 + \mu &= \sigma_{04} \\
w_1 + w_2 + w_3 + w_4 &= 1
\end{align*}
\]

We would solve the above system for the four unknown weights \( w_1, w_2, w_3, \) and \( w_4 \). The \( \sigma_{ij} \)'s are all known values taken from the variogram as earlier discussed. Note that when \( \gamma(h) = 0 \) and \( \sigma(0) \), the a priori variance. (Also note that computations using the variogram are simplified somewhat because we are assuming point samples.) The Lagrangian multiplier \( \mu \) which is used to force the estimation variance to a minimum, is an unknown and must also be found. The last of the five equations simply represents the fact that the sum of the weights must be one, which we must have if the kriging estimator is to be an unbiased estimator. Because we have five unknowns and five equations, we can get a unique answer for the weights (and the Lagrangian multiplier).

As it turns out, the kriging estimator then has three important properties:

1) it is a simple linear estimator,
2) it is unbiased, and
3) out of all unbiased linear estimators, it has the smallest estimation variance, and hence the highest reliability.

Thus, the kriged estimator is known as a BLUE estimator, or best linear unbiased estimator.

### The Reliability of a Linear Estimator

Fortunately, it is an easy matter to calculate the estimation variance of any linear estimator. Although kriging will always produce the lowest estimation variance, we can still compute the estimation variance of such methods as inverse distance, triangulation, polygonal, etc. All that we need to know is the variogram and the weights. The general formula for the estimation variance for any linear estimator (for the stationary case where the observations are point samples and the region to be estimated is a point) is:

\[
de = \sigma^2 - 2 \sum_i \sigma_{0i} w_i + \sum_i \sum_j w_i w_j \sigma_{ij}
\]

(David, 1979, p. 243).

### Estimating the Thickness of a Coal Seam

A North American coal deposit was sampled by \( n = 162 \) drill cores haphazardly scattered across a portion of the deposit. The minimum distance between almost all drill hole core samples was about 2,000 feet. When the cores were removed, the lengths of the coal seam were estimated in feet. The empirical variogram was constructed and is shown below.

A theoretical spherical model was fitted to the empirical variogram. The nugget effect \( C_0 \) was estimated at about 26 square feet, the sill \( C_0 + C \) at about 2.50 square feet, and the range at approximately 28,000 feet. Note that the
VARIOGRAM FOR COAL SEAM THICKNESS

N = 162

SPHERICAL VARIOGRAM MODEL FITTED
NUGGET EFFECT = .26, SILL = 2.50
RANGE = 28 THOUSAND FEET
GAMMA IN SQUARE FEET

nugget effect is about 10% of the sill. The relatively large seeming lack of continuity in the seam is probably due to measurement error exaggerating the true amount of discontinuity.
(Also, the nugget effect estimate is a linear extrapolation from the existing data, since only a very few inter-sample distances were actually less than 2,000 feet.) Because the variogram is relatively flat from about 28,000 feet to 80,000 feet, and from an analysis of possible drift, the assumption of stationarity over the region in question appears reasonable.

Based on the above analysis, an estimated map for a small area of the deposit was created using kriging and is shown below. (The black dots represent the locations of drill hole samples.)

Coal Seam Thickness in Feet

A map showing the square root of the estimation variance corresponding to the above coal seam thickness map was also calculated and is shown below. Note that areas that are farther away from samples generally have larger values for the square root of the estimation variance.

Square Root of the Estimation Variance in Feet

Widely Spaced Rainfall Sampling Stations

Monthly rainfall data was collected for 9 sampling stations in the New York state area from 1965 to 1973 (Water Resources Data for New York). Making some assumptions about stationarity over time, an empirical variogram was constructed using n = 817 rainfall measurements in inches and is shown below. Because the min-

VARIOGRAM FOR RAINFALL DATA
NEW YORK STATE (1965-73) N = 817

PURE NUGGET EFFECT VARIOGRAM MODEL FITTED
\[ \gamma(h) = 2.88 \text{ SQUARED INCHES} \]

Minimum inter-station distance is more than 40 miles, a pure nugget effect model is displayed.

References