Gregg Babish
GIS Specialist

Ecological Research Division
Environmental Conservation Branch
2365 Albert Street, Room 300
Regina, Saskatchewan S4P 4K1

Edition 4.60

March 1, 2000

Environment Canada

Geostatistics Without Tears

A practical Guide to Geostatistics, Variograms and Kriging

GEOSTATISTICS WITHOUT TEARS

A Practical Guide to Geostatistics, Variograms and Kriging

by Gregg Babish
GIS Specialist
Environment Canada
illustrations by Mark Gilchrist
Environment Canada

"In the dawn of the new millenium, an ever-increasing number of companies are using geostatistics. Today, this new science, based on numerous disciplines, has well and truly come into its own." *(http://www.geovariances.fr/)*

PREFACE

Geostatistics is a branch of applied statistics developed by George Matheron of the Centre de Morphologie Mathématique, Fontainebleaue, France. Geostatistics was originally developed for the mining industry to estimate changes in ore grade. Today, the principles of geostatistics are applied in a number of scientific disciplines.

Geostatistics is used to estimate a data value for locations that cannot be sampled directly by examining data taken at locations that can be sampled. The ability to predict the unknown locations is influenced by the spatial arrangement of the known sample locations. If a model of spatial correlation can be established, this model can be used to interpolate data values at the unknown locations.

The term geostatistics typically refers to spatial estimation using a technique known as kriging. Kriging is essentially a weighted moving average method for estimation based on known observations. The results of kriging are dependent upon the distance between the sampling locations and the point of estimation. The effectiveness of kriging depends upon how well the observation points represent the phenomenon under study and the appropriate selection of the model parameters.

It has been said that " . . . there is to be no geostatistics without tears" (Oliver and Webster, 1991). The intent of this document is to provide the reader a clear introduction and understanding of surface interpolation methods, geostatistics, development of the variogram, semivariogram models, development of the kriged surface, and the affects of kriging parameter changes. This report was developed to serve as supporting documentation for a hands-on training course on geostatistics that is presented by the author. Hopefully, no tears will be shed.

INTRODUCTION

The creation of data surfaces is a frequent requirement for environmental Geographic Information System (GIS) applications. Surface models can be used for a variety of purposes including interpolating between actual data measurements, identifying data anomalies and establishing confidence intervals around predictions. These surfaces include development of digital elevation models (DEMs), digital terrain models (DTMs), population densities, groundwater tables, contamination plumes, soil properties, vegetation types, and climatic measurements.

The surfaces created are then used to provide information for subsequent analysis such as flooding scenarios, population change, distribution of contaminants, groundwater modeling, and weather change. The accuracy of subsequent analyses depends directly on the accuracy of the surfaces created in the early stages of analyses.

Data to create these surfaces are usually collected through field sampling and surveying. As a rule, data collection can only be conducted at a limited number of point locations due to limited resources and high costs. In order to generate a continuous surface, such as migratory bird population counts across a region, some type of interpolation method must be used to estimate data values for those locations where no samples or measurements were taken.

INTERPOLATION METHODS

The procedure of estimating the value of properties at unsampled sites within the area covered by existing point observations is called interpolation. The best interpolation method to use depends on the data, the objectives of the study, as well as the ease of generating a surface versus accuracy of the developed surface. There is no interpolation method that guarantees the best results for all data sets. Interpolation methods that are commonly available to the GIS user are discussed below.

Splines

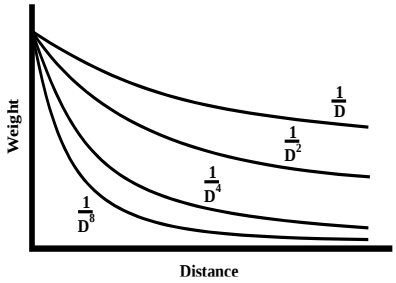
Although their mathematical equivalent is now used, splines started out being used as flexible pieces of wood or metal. These were originally used to fit curvilinearly smooth shapes when the mathematics and/or tools were not available to machine the shapes directly (boat hulls and airplane wings). Splines are useful in fitting a smooth line or surface to irregularly spaced data points. The most common spline is the cubic spline. Some variations of the spline method include minimum curvature, thin plate spline, and completely regularized splines.

A spline is an interpolation of a curve between measured data points with a series of functions. The splined curve goes through the data points exactly and assumes a curved shape elsewhere. Some supposedly “spline” programs are line smoothing functions, not spline functions. These programs smooth a line by generally going through the midpoints on a line segment, not the

endpoints. A true spline will keep the original points on the line. If line smoothing is what is required, these scripts will work. If a spline is required, be careful.

There are a number of known problems with splines. Splines are poor for surfaces which show marked fluctuations, causing wild oscillations in the spline (climatic variables for example). Extrapolating beyond the edges of the data domain often yields erratic results. This is because there is no information beyond the data to constrain the extrapolation and splines will grow to large values (positive or negative). Closely spaced data points can develop embolisms. However, if a smooth surface is required, splines can frequently give a good, usable smooth fit to a data set. Splines tend to emphasize trend rather than anomalies.

Inverse Distance Weighted (IDW)



The inverse distance method is fundamentally a function that averages the surrounding points (Jones, et. al., 1986). Weighting is assigned to observational data points through the use of a weighting power that controls how the weighting factor drop off as the distance from the point increases. The greater the weighting power, the less effect further away points have on the interpolation.

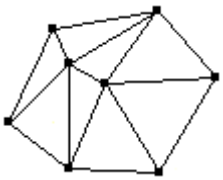
Simplicity, speed of calculation, programming ease, and reasonable results for many types of data are some of the advantages associated with inverse distance weighted interpolation. The number of surrounding points used to interpolate the grid value can significantly change the map output. Because inverse-distance is an averaging (smoothing) technique it cannot interpolate above or below the surrounding data, which tends to generate flat areas. Distance-weighted average preserves sample data values, and is therefore an exact interpolation technique.

The inverse distance interpolation method is easily affected by the uneven distribution of data points since an equal weight is assigned to each of the data points even if they are a cluster. IDW is fast, but has a tendency to generate "bull's eyes" around the actual data points, and tends to emphasize both trends and anomalies.

This method is sometimes called the nearest neighbor weighted interpolation. Other inverse distance methods are the Shepard's Method that uses an inverse distance weighted least squares method, and the inverse distance to a power interpolator.

Although distance-weighted methods underlie many of the techniques in use, they are far from ideal. IDW is probably always the worst technique to implement (Goodchild, 1999).

Triangulation



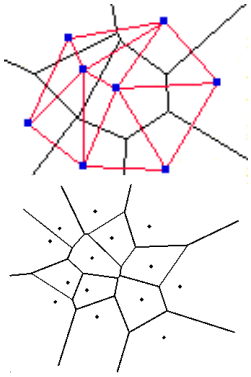
When a set of arbitrarily distributed data points is connected in a triangular network, the resulting lattice work is irregular, and is called a Triangular Irregular Network (TIN). The TIN is a vector data structure. Various techniques are available for the construction of TINs using a computer. In

some software, the TIN is generated and then used to create a continuous raster surface model.

Triangulation is a fast, flexible and popular interpolation method that offers several advantages. When a surface has significant relief features, triangulation generates more accurate surface representation by maintaining breakline features such as stream channels, shorelines, and relief features. Triangulation can represent a surface with fewer data points than grid based methods. Triangulation works best when data points are evenly distributed over an area. Sparse data sets result in distinct triangular facets.

The main disadvantage of triangulation is that the generated surfaces are not smooth and may have a jagged appearance. Because triangulation is an exact method – the original data points are honored exactly – triangulation is generally not suitable for extrapolation beyond the observed data points.

The Voronoi diagram, sometimes named Dirichlet tessellation, is one of the most fundamental and useful constructs defined by irregular lattices. The term tessellation means to break an area into pieces or tiles. A Voronoi polygon is defined by lines that bisect the lines between a center point and its surrounding points. The bisecting lines and the connection lines are perpendicular to each other. Geographical interest in the Voronoi diagram originates from the climatologist A. H. Thiessen's use of Thiessen polygons to define regions that surround unevenly distributed weather stations.



Delaunay triangulation, named after B. Delaunay, is closely related to the Voronoi diagram. Using the Voronoi diagram as a basis, Delaunay triangulation can be constructed by drawing lines between the points in adjacent polygons. Delaunay triangulation is the most common triangulation technique.

Triangular models have been used in terrain modeling since the 1970s, and are sometimes used to create Digital Terrain Models (DTMs). Commercial systems using TIN began to appear in the 1980's as contouring packages and some were embedded in GIS packages.

Certain types of terrain are very effectively divided into triangles with plane facets. This is particularly true with fluvially-eroded landscapes. However, other landscapes, such as glaciated ones, are not well represented by flat triangles. Triangles work best in areas with sharp breaks in slope, where TIN edges can be aligned with breaks, such as along ridges or channels.

Due to limitations of computers and the complexity of TIN data structures, triangular models have been overshadowed by gridded models. Today the use of triangular models is increasing.

Trend Surface Analysis

Trend surfaces are typically used to determine whether spatial trends exist in a data set, rather than to create a surface model to be used in further analyses. Trend surfaces may also be used to describe and remove broad trends from data sets so more local influences may be better understood. Because the resulting surface is an ideal mathematical model, it is very smooth and is free from local detail.

The idea behind the trend surface interpolation is to fit a least-squares surface to observational data points by using polynomial regression. Trend fits one of three mathematically-defined ideal surface models, linear, quadratic, or cubic, to a point data set.

"Trend" analysis is the geology profession's name for a mathematical method of separating map data into two components - that of a regional nature, and local fluctuation (Davis, 1973). Trend surface analysis is the most widely used global surface-fitting procedure. A global interpolator derives the surface model by considering all the data points at one. The resulting surface give a “best fit” for the entire sample data, but may provide a very poor fit in particular locations. A local interpolator, calculates new values for unknown locations by using the values of known locations that are nearby.

Trend surface analysis is an inexact interpolator. Since trend surface is a smoothing and approximate method, the generated surface rarely pass through original data points.

The advantage of trend analysis is that it is easy to understand the way the surfaces are estimated. It can be used to show broad features of the observational data points. Trend surfaces are appropriate for estimating a grid matrix in certain circumstances. If the raw data are statistical in nature, with perhaps more than one value at an observation point, trend surfaces provide statistically optimal estimates of a linear model that describes their spatial distribution.

There are a number of disadvantages to a global fit procedure. The most obvious of these is the extreme simplicity of a polynomial surface as compared to most natural surfaces. Trend surface analysis is highly affected by the extreme values and uneven distribution of observational data points. The problem is further complicated by the fact that some data points are more informative than others. For example, in interpolating topographic surfaces, the data points taken from the peaks and depressions are more significant than the points taken from level or sloping locations. Polynomial trend surfaces also have a tendency to estimate higher or lower values in areas where there are no data points, such as along the edges of maps. All surface estimation procedures have difficulty extrapolating beyond the area of data control, but trend surfaces seem especially prone to the generation of seriously exaggerated estimates.

Kriging

Kriging is a weighted moving averaging method of interpolation, devised by geostatisticians to provide the best local estimate of the mean value of a regionalised variable. Kriging is an advanced application of Gauss’ Theory of Errors. The main difference between kriging and a simple distance-weighted average is that kriging allows flexibility in defining the interpolation model.

Measurements that change at every location they are measured at are called regionalized variables. Regionalised variables describe phenomena with geographical distribution, such as surface elevation, geophysical measurements, soil contamination, and groundwater quality. Regionalised variables are variables that fall between random variables and completely deterministic variables. Regionalised variables recognize the fact that

properties measured in space (or time) follow an irregular pattern that cannot be described by a mathematical function.

Kriging is done in two parts. First, the sample semivariance is used to estimate the shape of the variogram (a curve that represents the semivariance as a function of distance). The variogram describes the spatial relationship between the data points.

Second, the estimated semivariance function is used to determine the weights needed to define the contribution of each sampled point to the interpolation. Sample points close to the point for which an estimated value is to be generated contribute the most to the interpolation.

Although other interpolation methods may be used, kriging is typically applied since kriging overcomes many of the shortcomings of traditional interpolation methods. Interpolation techniques such as inverse distance and triangulation do not take into account a model of the spatial process (the variogram).

The results from kriging are generally of higher quality and more realistic compared to techniques such as inverse distance and triangulation. Kriging can provide a measure of the error or uncertainty of the contoured surface. Since the estimation variances can be mapped, the confidence placed in the estimates can be calculated.

A variance plot provides information on how well the interpolated values fits the overall model that was defined by the user. The plot of the variance can be used as a diagnostic tool to refine the model. The goal is to develop a model with an even distribution of variance that is a close as possible to zero.

The estimation variance can also be used to determine where more data is needed if future sampling is planned. Kriging reduces the extreme weighting of values caused by irregular distribution of sample points, especially those which might result from clustering.

For example, where samples are strongly autocorrelated over small sampling intervals, kriging can show where to add sample points to bring estimation precision to a desirable level in sparsely sampled regions. Conversely, where samples are weakly autocorrelated over small intervals, kriging can show that additional sampling in a given region will add little additional precision to interpolation estimates.

Computationally, it is often desirable to choose a limited number of sample points to perform the interpolation, since the weights from distant sample points are often negligible and the error associated with leaving them out is minor. Kriging is not a suitable method for data sets which have spikes, abrupt changes or breaklines. Kriging works best with data sets which have regions of densely scattered data and regions of lightly scattered data. Kriging can take into account redundant data (clustering) and possible anistropies.

HISTORY OF KRIGING

Variation of crop yields (Mercer & Hall, 1911) was one of the earliest records of geostatistical application. They found that the plot-to-plot variance decreased as the size of the plots increased, up to some limit, beyond which the variance stabilized. It was noted that there was strong correlation between crop yields on adjacent plots (spatial dependence).

In the 1940s, the mining industry was using the sample mean of core-sample assays to estimate the average grade in a prospective mining block. Those estimates were then used to mine selectively. This method of estimation does not provide a very good estimate of total recoverable ore. Local variability can make or break a mining venture because concentrations of high-grade ore are easier and hence more profitable to mine, while regions of low-grade ore should be ignored.

In the early 1950's, Danie.G. Krige, a South African mining engineer working on the Witwatersrand looked at this problem and his work culminated in a benchmark paper (Krige, 1951). Professor Herbert Sichel is credited with the first significant use of mathematical statistics for mineral evaluation on the South African gold mines and worked extensively on problems associated with gold and diamonds until his death in 1995.

Georges Matheron, an engineer with Ecoles des Mines, Centre de Morphologie Mathématique, Fontainebleaue, France, became aware of Krige's approach to ore reserves calculation. Out of Matheron's research came a spatial interpolation method which he called kriging (pronounced "kree-ging" or "kree-jing") in honor of Dr. Krige. Matheron's first English paper on this matter appeared in 1963 (Matheron, 1963a).

While Matheron was developing his theory of prediction in France, the meteorologist L.S. Gandin in the Soviet Union was doing remarkably similar work (Gandin, 1963).

By the middle of the 1970s, knowledge of geostatistics had begun to spread beyond the minerals industry. During the late 1970s standard geostatistical techniques proved suited to describe the spatial distribution of soil. Kriging was applied to solving hydrological problems for point and areal rainfall estimation, and the design of rainfall measurement networks, and simulations (Delhomme, 1978).

During the 1980s and 1990s more applications of geostatistics appeared in the fields of meteorology, aerial distribution of acid rain and aerial contaminants, surface and subsurface hydrology, air and water pollution, disease outbreaks, and migratory bird population estimates.

GEOSTATISTICS

Conventional (univariate) statistics are based on random, independent variables which assume zero continuity and allow for no extension of each data value. This makes it theoretically impossible to estimate individual cells or points in a map as all of the points should be given the same overall mean value.

Geostatistics assumes that adjoining points are correlated to each other spatially. This continuity is measured and then used for obtaining estimates. The relationship expressing the spatial correlation is measured in a function called the variogram.

The rationale behind geostatistics and spatial interpolation is the observation that points closer together in space are more likely to have similar values than points further away – Tobler's Law of Geography (Tobler, 1979). Tobler's law is often cited as the First Law of Geography, where "everything is related to everything else, but near things are more related than distant things." This is also know as autocorrelation, samples collected close to one another are often more similar than samples collected further away, whether in space or in time.

As describe above, geostatistics is basically a two step process:

1. define the degree of autocorrelation among the measured data points, and once the spatial or temporal dependency is established,

2. interpolate values for points not measured using kriging algorithms based on the degree of autocorrelation.

An understanding of how data values vary as a function of distance allows interpolation of values at unsampled locations. Conventional statistics cannot take this into account. Two distributions might have the same mean and variance, but differ in the way they are correlated with each other.

Kriging methods may be used as either a global or local interpolator. However, local implementation is most often used. Kriging preserves sample data values and is therefore an exact interpolator.

VARIOGRAMS

Geostatistics has the ability to quantify the correlation between any two values separated by a distance (known as the lag distance) and use this information to make predictions at unsampled locations. Semivariance is a measure of the degree of spatial correlation among sample data points as a function of the distance and direction between the sample data points. The variogram controls the way that kriging weights are assigned to points during interpolation, and consequently controls the quality of the results.

The semivariogram, or simply the variogram, is a graph describing the expected difference in value between pairs of samples within a given orientation. The semivariance increases with distance until at a certain distance away the semivariance will equal the variance around the average value, and will therefore no longer increase, causing a flat region to occur on the semivariogram called the sill. From the point of interest to the distance where the sill begins is termed the range of the regionalised variable. Within this range, locations are related to each other, and all known samples contained in this region (the neighborhood) must be considered when estimating unknown points.

The shape of the semivariogram function is typically one of four forms: linear, spherical, exponential, or Gaussian. Since there are only limited observational data points available, it is impossible to compute the absolute semivariogram. The semivariograms is estimated from observed data values by selecting the function which best fits the observed data points. This process involves a great deal of interpretation, and may require a number of trial and error computations.

CALCULATING THE VARIOGRAM

An estimate must be made of the distance required to travel before data points separated by that much distance are uncorrelated. This information is presented in the form of the variogram – the semivariance versus the lag distance. Other than the work involved, it is not really difficult to calculate and plot the semivariogram.

The first step is to define a lag increment (h), which is the spacing between any two points in the data. For the data shown in Figure 1, a 1 kilometer lag increment is arbitrarily defined. For each pair separated by 1 kilometer, calculate the difference, and then square the difference. Sum up all the differences and divide by twice the number of pairs. This gives the similarity measure $\gamma(h)$ for the variogram for that particular lag increment or distance. The same is done for other lag distances of 2 kilometers, 3 kilometers, 4 kilometers, 5 kilometers, and 6 kilometers.

$$\gamma(h) = \left(\sum_{i=1}^{N} [Z(x_i) - Z(x_i + h)]^2 \right) / 2N$$

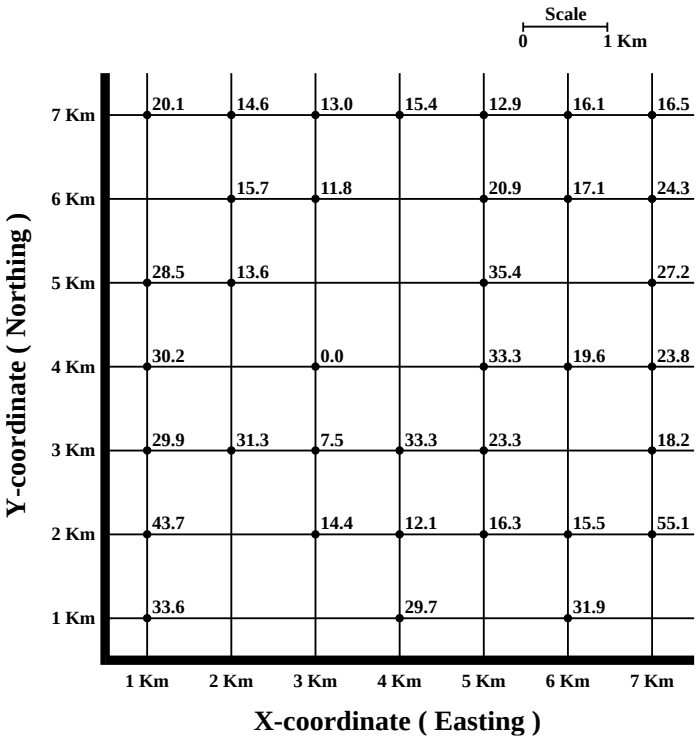


Figure 1: A two-dimensional spatial region with regionalised variable observations (Z).

First Lag Distance[1]

Twenty (N = 20) pairs are found for lag distance h = 1 (**Figure 2a**). The pairs in the east-west direction and their calculated differences:

20.1 - 14.6 = 5.5	14.6 - 13 = 1.6	13 - 15.4 = -2.4	15.4 - 12.9 = 2.5
12.9 - 16.1 = -3.2	16.1 - 16.5 = -0.4	15.7 - 11.8 = 3.9	20.9 - 17.1 = 3.8
17.1 - 24.3 = -7.2	28.5 - 13.6 = 14.9	33.6 - 19.6 = 13.7	19.6 - 23.8 = -4.2
29.9 - 31.3 = -1.4	31.3 - 7.5 = 23.8	7.5 - 33.3 = -25.8	33.3 - 23.3 = 10
14.4 - 12.1 = 2.3	12.1 - 16.3 = -4.2	16.3 - 15.5 = 0.8	15.5 -55.1 = -39.6

The sum of the squared differences:

30.25 + 2.56 + 5.76 + 6.25 + 10.24 + 0.16 + 15.21 + 14.44 + 51.84 + 222.01 + 187.69 + 17.64 + 1.96 + 566.44 + 665.64 + 100.00 + 5.29 + 17.64 + .64 + 1568.2 = 3489.82

The gamma for the first lag distance is:

$\gamma(1) = 3489.82/(2*20) = 87.25$

Second Lag Distance

The lag distance h is incremented such that $h = h + h = 2h$. In the case of this example, 2 kilometers. A search is conducted for pairs at a lag distance of 2 kilometers, oriented in the east-west direction. All pairs used for the h = 1 search and calculation are not included in the new lag interval.

Twenty (N = 20) pairs are found for lag distance h = 2 (**Figure 2b**). The pairs and their calculated differences:

20.1 - 13.0 = 7.1	14.6 - 15.4 = -0.8	13.0 - 12.9 = 0.1	15.4 - 16.1 = -0.7
12.9 - 16.5 = -3.6	11.8 - 20.9 = -9.1	20.9 - 17.1 = -3.4	35.4 - 27.2 = 8.2
30.2 - 0 = 30.2	0 - 33.3 = -33.3	33.3 - 23.8 = 9.5	29.9 - 7.5 = 22.4
31.3 - 33.3 = -2.0	7.5 - 23.3 = -15.8	23.3 - 18.2 = 5.1	43.7 - 14.4 = 29.3
14.4 - 16.3 = -1.9	12.1 - 15.5 = -3.4	16.3 - 55.1 =-38.8	29.7 - 31.9 = -2.2

The sum of the squared differences:

50.41 + 0.64 + 0.1 + 0.49 + 12.96 + 82.81 + 11.56 + 67.24 + 912.04 + 1108.89 + 90.25 + 501.76 + 4.00 + 249.64 + 26.01 + 858.49 + 3.61 + 11.56 + 1505.40 + 4.84 = 5502.65

The gamma for the second lag distance is:

$\gamma(2) = 5502.65/(2*20) = 137.57$

[1] The example variogram calculations have been adapted from Carr (1995).

Third Lag Distance

Pairs of observations oriented parallel to the east-west direction and with lag distances of 3 kilometers are now listed.

Fifteen (N = 15) pairs are found for lag distance h = 3 (**Figure 2c**). The pairs and their calculated differences:

20.1 - 15.4 = 4.7	14.6 - 12.9 = 1.7	13.0 - 16.1 = -3.1	15.4 - 16.5 = -1.1
15.7 - 20.9 = -5.2	11.8 - 17.1 = -5.3	13.6 - 35.4 = -21.8	0 - 19.6 = -19.6
29.9 - 33.3 = -3.4	31.3 - 23.3 = 8.0	33.3 - 18.2 = 15.1	43.7 - 12.1 = 31.6
14.4 - 15.5 = -1.1	12.1 - 55.1 = -43.0	33.6 - 29.7= 3.9	

The sum of the squared differences:

22.09 + 2.89 + 9.61 + 1.21 + 27.04 + 28.09 + 475.24 + 384.16 + 11.56 + 64.00 + 228.01 + 998.56 + 1.21 + 1849.00 + 15.21 = 4117.88

The gamma for the third lag distance is:

$\gamma(3) = 4117.88/(2*15) = 137.26$

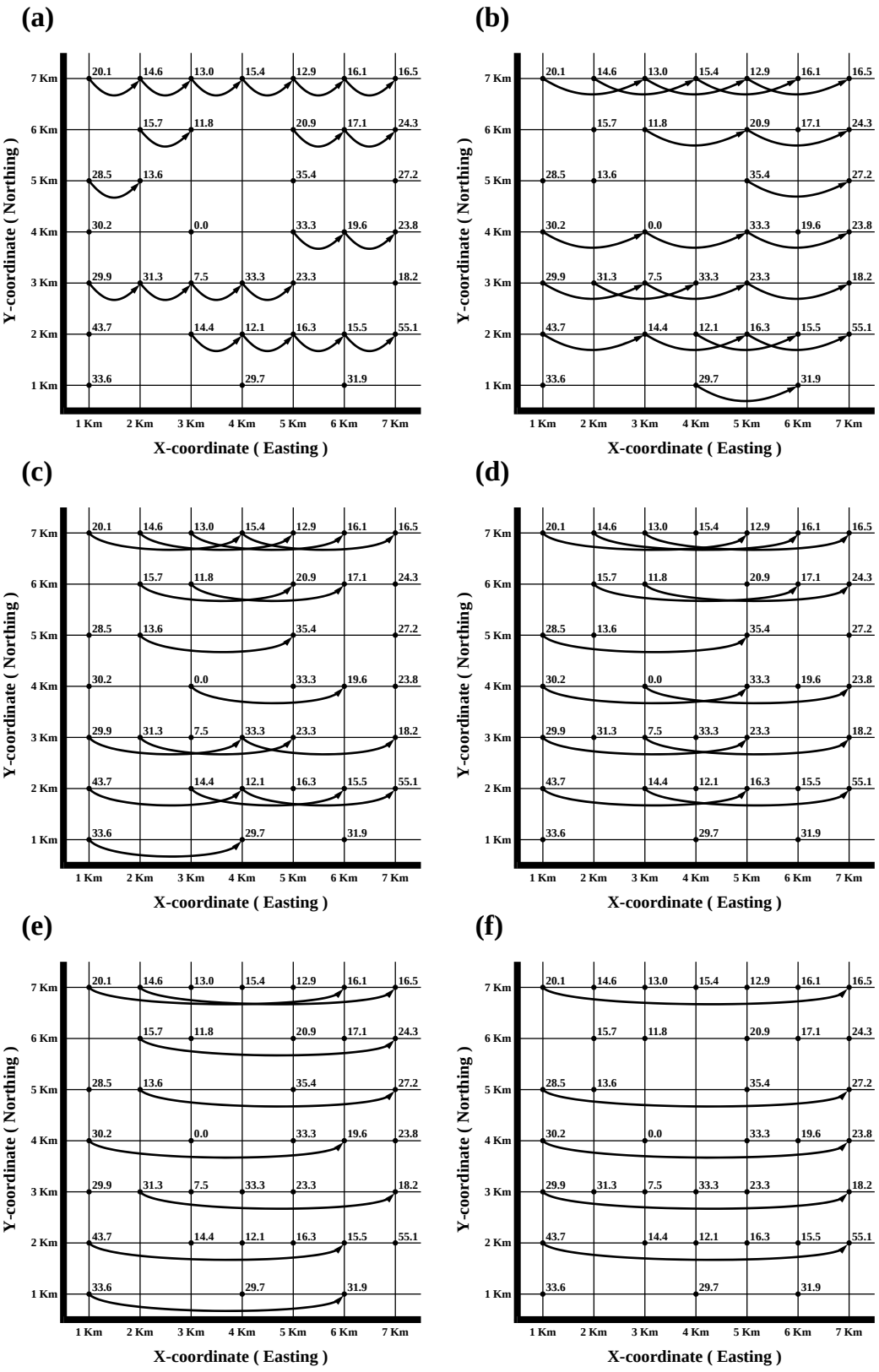


Figure 2: Transects used to estimate the semi-variances at varying lag distances (*h*)

Fourth Lag Distance

Pairs of observations oriented parallel to the east-west direction and with lag distances of 4 kilometers are now listed.

Twelve (N = 12) pairs are found for lag distance h = 4 (**Figure 2d**). The pairs and their calculated differences:

20.1 - 12.9 = 7.2	14.6 - 16.1 = -1.5	13.0 - 16.5 = -3.5	15.7 - 17.1 = -1.4
11.8 - 24.3 = -12.5	28.5 - 35.4 = -6.9	30.2 - 33.3 = -3.1	0 - 23.8 = -23.8
29.9 - 23.3 = 6.6	7.5 - 18.2 = -10.7	43.7 - 16.3 = 27.4	14.4 - 55.1 = -40.7

The sum of the squared differences:

51.84 + 2.25 + 12.25 + 1.96 + 156.25 + 47.61 + 9.61 + 566.44 + 43.56 + 114.49 + 750.76 + 1656.49 = 3413.51

The gamma for the fourth lag distance is:

$\gamma(4) = 3413.51/(2*12) = 142.23$

Fifth Lag Distance

Pairs of observations oriented parallel to the east-west direction and with lag distances of 5 kilometers are now listed.

Eight (N = 8) pairs are found for lag distance h = 5 (**Figure 2e**). The pairs and their calculated differences:

20.1 - 16.1 = 4.0	14.6 - 16.5 = -1.9	15.7 - 24.3 = -8.6	13.6 - 27.2 = -13.6
30.2 - 19.6 = 10.6	31.3 - 18.2 = 13.1	43.7 - 15.5 = 28.2	33.6 - 31.9 = 1.7

The sum of the squared differences:

16.00 + 3.61 + 73.96 + 184.96 + 112.36 + 171.61 + 795.24 + 2.89 = 1360.63

The gamma for the fifth lag distance is:

$\gamma(5) = 1360.63/(2*8) = 85.04$

Sixth Lag Distance

Pairs of observations oriented parallel to the east-west direction and with lag distances of 6 kilometers are now listed.

Five (N = 5) pairs are found for lag distance h = 6 (**Figure 2f**). The pairs and their calculated differences:

20.1 - 16.5 = 3.6	28.5 - 2.7.2 = 1.3	30.2 - 23.8 = 6.4	29.9 - 18.2 = 11.7
43.7 - 55.1 = -11.4			

The sum of the squared differences:

12.96 + 1.69 + 40.96 + 136.89 + 129.96 = 322.46

The gamma for the sixth lag distance is:

$\gamma(6)= 322.46/(2*5) = 32.25$

Graphing the Semivariogram

Summary of the east-west directional semivariogram calculations for the data represented in Figures 1 and 2:

Interval of *h*	Number of Pairs	$\gamma(h)$
1	20	87.25
2	20	137.57
3	15	137.26
4	12	142.23
5	8	85.04
6	5	32.25

The similarity measurements are plotted with the x-axis being the lag distance (*h)*, and the y-axis being $\gamma(h)$. These values are plotted in the semivariogram shown in Figure 3. A line is fitted through the plotted points to assess the spatial correlation. The result is called the experimental variogram, or simply the variogram. The initial value of *h* is called the class size. In this example, the class size was selected to be the minimum sampling resolution of 1 kilometer.

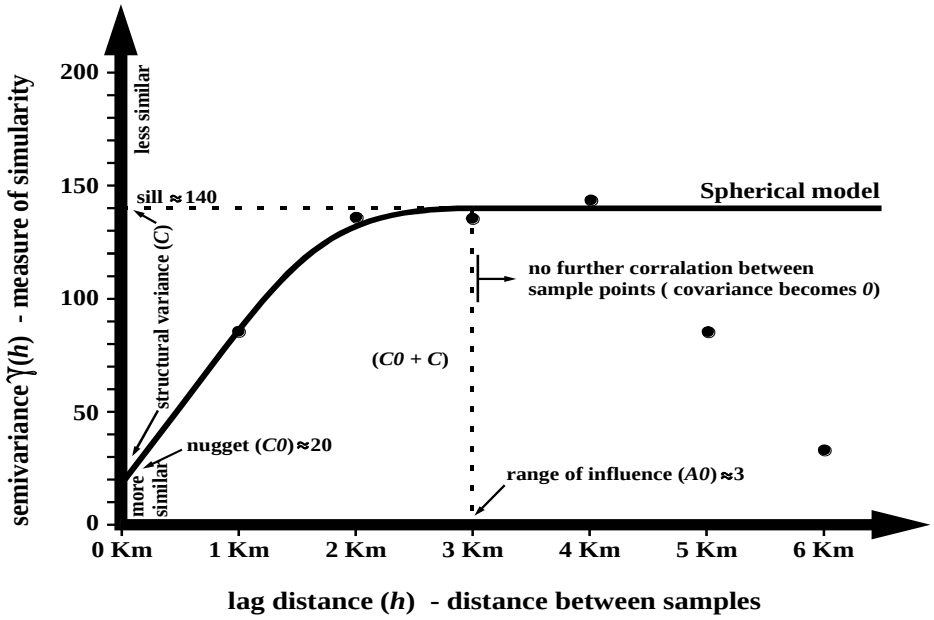


Figure 3: Characteristics of a semivariogram.

Characteristics of the semivariogram

The shape of the plot for the example semivariogram shown in Figure 3 is an ideal representation of the semivariogram observed for many regionalised phenomena and is known as a spherical semivariogram shape. A typical semivariogram has three characteristics – the sill, the range, and the nugget.

Sill

When the distance between sample points is zero, the value at each point is compared with itself. Therefore, all the differences are zero, and the calculated semivariance for $\gamma(0)$ should be zero. If h is a small distance, the points being compared tend to be very similar, and the semivariance will be a small value. As the distance h increases, the points being compared are less and less closely related to each other and their differences become larger, resulting in larger values of $\gamma(h)$. In most cases involving environmental data, spatial variability between sample pairs increases as the separation distance increases.

At some distance the points being compared are so far apart that they are not related to each other, and their squared differences become equal in magnitude to the variance around the average value. The semivariance no longer increases and the semivariogram develops a flat region called the sill. The sill indicates that there is no further correlation between pairs of samples. The higher the sill value, the higher the prediction variances.

Range

Essentially, the range is the sampling distance between points where the semivariance reaches its maximum point, called the sill. The distance at which the rising variogram reaches the sill is called the range of influence, or simply the range and is symbolized by *A0*. The range defines a neighborhood within which all locations are related to one another. At distances greater than the range, any two pairs of values are independent of each other. Samples that are separated by distances in excess of this range are said to be spatially independent (not correlated).

The range signifies the distance beyond which sample data should not be considered in the interpolation process when selecting points that define a local neighborhood. The range can be used to identify patterns of spatial independence. A larger range means more continuous behavior, and the variable is very well correlated in space. Predictions result in fairly smooth maps of the variable of interest. The range typically tends to increase as more and better data become available.

Nugget effect

By extrapolating the plot of the semivariogram back to the origin ($h = 0$), the nugget value is found – the value of the semivariogram at a zero lag distance. This is called the nugget effect or nugget variance and is represented by *C0*.

Ideally, the nugget effect should be zero as two samples from the same point should have the same value. If the intercept is greater than zero then a random or unstructured component of variation at $h = 0$ is present. Although $\gamma(0) = 0$ by definition, such a variogram is seldom obtained in practice.

The nugget effect shows the apparent failure of the semivariogram to go through the origin, indicating a variable that is highly variable over distances less than the sampling interval. The nugget effect is related to the underlying characteristics of the spatial data and can have two components. The nugget

effect shows the pure random variation in population density (white noise) or it may be associated with errors in the sampling process. Analytical error, for example, will introduce some degree of uncertainty into the data. A nugget structure increases the variability uniformly across the entire graph because it is not related to distance or direction of separation.

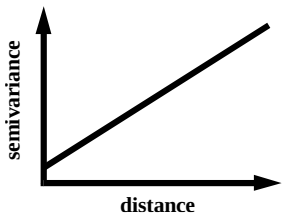
Species of North American birds censused on the breeding bird survey vary considerably in their nugget variances (Villard and Maurer, 1996). Temperate migrants tend to have higher nugget variances than Neotropical migrants, and species that use a number of different habitat types have higher nugget variances than species using a single type of habitat.

For estimates of abundances obtained from censuses like the breeding bird survey, part of the nugget variance could be due to the inherent variability among observers in counting birds at different breeding bird survey routes. The second component is due to the discontinuous nature of the process in space. For birds, this would be related to the continuity of the habitat in which they were being counted. Since the habitat of a species is generally discontinuous in space, this is a likely component of the nugget variance for data like the breeding bird survey.

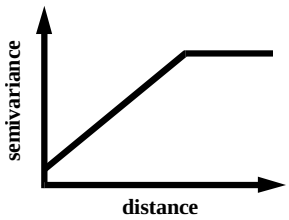
Semivariogram Models

There is an art to determining the correct variogram model to use. Semivariances are estimated at discrete values of *h*, whereas the true variogram is continuous. The estimates are subject to error, and unless a large sample is taken (several hundred points) the experimental variogram will appear erratic (as in our example variogram plotted in Figure 3). It is necessary to fit some kind of model to the sample values to represent the true variogram for a region. To consider the effect of the model on predictions, think in terms of the shape of the variogram for early lags – the better correlated the variable the smoother looking the resulting maps. There are only a few simple models that are appropriate to use.

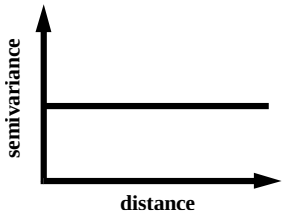
Linear



The linear model describes a straight line variogram. The linear model applies when the spatial variability observed for a variable keeps increasing linearly with distance and never levels off. The range is defined arbitrarily to be the distance interval for the last lag class in the variogram. Because the range is an arbitrary value it should not be compared directly with the ranges of other models. There is no sill. The sill is the calculated semivariance for the arbitrarily defined range. The linear form indicates moderate continuity.

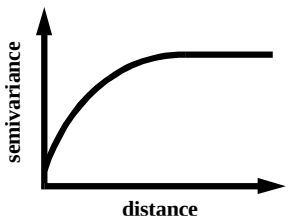


The linear/sill model is similar to the linear model except that at some distance (the range), pairs of points will no longer be autocorrelated and the variogram will reach a sill. This model should not be used unless the variation is limited to one dimension (Webster, 1985).



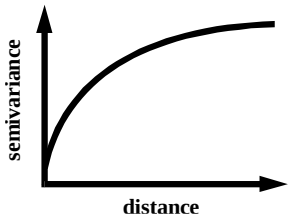
The horizontal form is produced by a random variable having no spatial autocorrelation. Horizontal variograms, are pure nugget, they indicate that there is no spatial dependence among the observations at the scale of sampling. Measurement errors, or sampling sites that are too far apart and thus spatially independent or both may lead to noisy semivariograms that appear as pure nugget effect. Choosing a pure nugget effect model is an extreme modeling decision that precludes usage of kriging.

Spherical



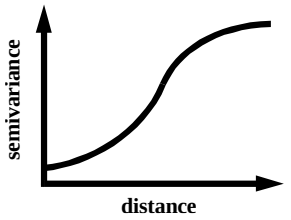
The spherical model shows a progressive decrease of spatial dependence of the variable when the distance between them decreases. The dependence fades away altogether when the model reaches a sill. The spherical model is a modified quadratic function for which at some distance (the range), pairs of points will no longer be autocorrelated and the semivariogram reaches a sill (becomes asymptote). The spherical model is a commonly used model.

Exponential



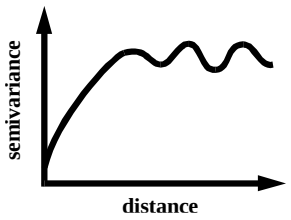
The exponential model applies when the spatial dependence decreases exponentially with increasing distance; this dependence disappear completely only at infinite distance. The exponential model is similar to the spherical model in that it approaches the sill gradually, but different from the spherical model in the rate at which the sill is approached and in the fact that a definite sill is never actually achieved. The exponential model is a commonly used model.

Gaussian



The gaussian or hyperbolic model is similar to the exponential model but assumes a gradual rise for the y-intercept. The spatial dependence vanishes only at an infinite distance. The main feature of this model is its parabolic shape at the origin. It expresses a rather smooth spatial variation of the variables. The range in the gaussian model is not the range but rather a parameter used in the model to provide range. Range to 95% of the sill in the gaussian model can be estimated as 3 times the range.

Hole Effect



The hole effect variogram, also called the holesin model, suggests repetition in the variation that is neither wholly random nor periodic. Semivariograms from breeding bird survey data typically do not indicate the existence of an asymptotic maximum semivariance (Maurer 1994). Rather, they indicate that the maximum semivariance occurs at intermediate distances, and that points farther apart than this distance are actually more highly correlated.

This increasing correlation with large distances separating sampling sites can be explained by the characteristic distribution of abundances across a geographic range. Generally a species is most abundant in one or a few places near the center of its geographic range, and its abundance decreases away from these points (Maurer and Villard, 1994). Abundances at points relatively large distances away from one another should be positively correlated because they represent peripheral sites in the geographic range.

Abundance may be determined by a number of factors, each operating at a different scale. For example, large-scale patterns in climate variation across a continent may determine large-scale patterns of vegetation productivity, and consequently, variation in habitat for birds. Small-scale variation exists among sites located in the same biotic province, such as the physiographic regions used in the breeding bird survey. Sites that occur within the same physiographic regions, for example, might have abundances more similar than those that fall in different regions.

Alternate Variogram Analysis

Alternate variograms to provide an estimate of spatial variability can be plotted. These alternate varigorams are sometimes less sensitive to outliers, skewed distributions, or clustered data than ordinary variograms and may help in the recognition of data structure when the ordinary variogram is too noisy.

The relative variogram is analogous to the relative standard deviation often used to measure analytical variability. Relative variograms are unitless (decimal fraction squares). When modeled and used for kriging the relative kriging standard deviations must be multiplied by the estimated values to be comparable with kriging standard deviations produced with ordinary variogram models.

The madogram calculates the mean absolute difference. Avoiding the squared term results in a more robust measure of the correlation structure. Madograms are not true variograms because they are not based upon squared differences. In general, kriging with madogram models is not recommended.

The non-ergodic variogram is based on estimates of covariance rather than variance (Srivastiva, 1987). The non-ergodic variograms have the same units (measurement units squared) as ordinary variograms and may be modeled and used for kriging in the same way.

Directional Calculations

One purpose of the semivariogram calculation is to identify spatial correlation behavior that depends on direction (anisotropy). Once trends have been accounted for (or assumed not to exist), all other variation is assumed to be a function of distance.

It is common to begin variogram modeling using the omnidirectional variogram. The omnidirectional variogram pairs all samples to each other regardless of what direction they are oriented. To graph it requires specifying only two parameters: lag width and number of lags.

The omnidirectional variogram is unaffected by the direction of values. It is therefore useful for exploring the overall continuity behavior of the sample data and the spatial dependency pattern of the measured surface. The omnidirectional sample variogram curve is also useful for detecting or confirming lag distance parameters that show a clear continuity pattern before calculating other variograms.

Spatial continuity often varies with direction. The directional variogram is a form of exploratory data analysis based on the selection of sample pairs according to directional limits and tolerances. In practice, viewing many variograms is necessary to infer a spatial continuity structure. A collection of variograms best reveals the composite spatial continuity structure of the sample data.

Typically, four variograms with different orientations are constructed. These orientations are usually E/W, NE/SW, N/S, and NW/SE. The directional convention used in geostatistics are E/W is 0 degrees, NE/SW is 45 degrees,

N/S is 90 degrees and NW/SE is -45 degrees (or 135 degrees). The direction component is set at one of these directions, plus or minus a deviation angle.

Sometimes eight spatial directions analysis is used to yield more directional resolution. Large data sets are usually required to yield enough information for the eight directional calculations.

If the directional variograms are reasonably similar, then the spatial correlation can be assumed to be isotropic in which case the variability between samples would be a function of distance only. If they are not the same, then the distribution is said to anisotropic – spatial correlation varies with direction and distance. In this case directional trends are important and this knowledge should be incorporated into monitoring design and analysis.

Another type of anisotropy is zonal anisotropy, in which the ranges of semivariograms over all spatial directions are approximately the same, but the sill values are different.

Two primary types of kriging exist. Ordinary kriging (simple kriging) assumes that the surface has a constant mean, no underlying trend and that all variation is statistical. Universal kriging is used if there is a gradual trend in the data. Universal kriging combines trend surface analysis with ordinary kriging (Bonham-Carter, 1994). Universal kriging is more difficult to apply, but estimates are obtained in the same manner.

The directional variograms are used to further explore and improve the emerging description of spatial continuity. Directional variograms should be examined with the intent of revealing the most information, not of making the best looking models.

KRIGING

If measurements have been made at scattered sampling points and the form of the semivariogram is known, it is possible to estimate the value of the surface at any unsampled location.

Kriging uses the information from the semivariogram to find an optimal set of weights that are used in the estimation of the surface at unsampled locations. The kriging process involves the construction of a weighted moving average equation which is used to estimate the true value of a regionalised variable at specific locations. Since the semivariogram is a function of distance, the weights change according to the geographic arrangements of the samples. Kriging assigns low weights to distant samples and vice versa, but also takes into account the relative position of the samples to each other.

Calculating Kriging Weights[2]

For the sake of simplicity, assume that the data in Figure 1 are associated with isotropic spatial correlation (no directional trend) and that the semivariogram in Figure 3 can be modeled using a spherical model.

$\gamma(h) = C0 + C\,[1.5(h/A0) - 0.5(h/A0)^3]$ for $h \le A0$

$\gamma(h) = C0 + C$ for $h > A0$

where:

h = the lag distance interval
C0 = nugget variance
C = structural variance
A0 = range

Example semivariogram model parameters:

nugget (*CO*) = 20.0
sill = 140.0
structural variance (C) = 120
range (*A0*) = 3.0

$\gamma(h) = 20.0 + 120.0\,(1.5h/3 - 0.5h^3/27), \quad 0 < h < 3$

$\gamma(h) = 140.0, \quad$ when $h >= 3$

$\gamma(0) = 0$

An estimate can be made for the value of *Z* at the unsampled location 3,5 of Figure 1. To obtain this estimate, a decision is made to use six of the closest data locations, arbitrarily numbered 1 to 6.

i	X	Y	Z	Distance (km) from location (3,5)
1	2	6	15.7	1.414
2	3	6	11.8	1
3	1	5	28.5	2
4	2	5	13.6	1
5	5	5	35.4	2
6	3	4	0.0	1

The intersample covariance matrix must be calculated. This matrix is symmetrical, that is, $cov_{ij}(I,J) = cov_{ij}(J,I)$. The distance between locations X_i and X_j is the same as between X_j and X_i. The intersample covariance matirix is diagonalized. Because the matrix is symmetric, only the entries for one side of the diagonal must be calculated explicitly. Entries for the other side can be filled by the symmetrical relationship. A few steps are required to obtain an estimate at location X = 3, Y = 5, from Figure 1.

Step 1, calculation of distance to X_i to X_j for the computation of covariances. Because $\gamma(0) = 0$, then value for each diagonal entry is zero.

location	(2,6)	(3,6)	(1,5)	(2,5)	(5,5)	(3.4)
(2,6)	0.000	1.000	1.414	1.000	3.162	2.236
(3,6)	1.000	0.000	2.236	1.414	2.236	2.000
(1,5)	1.414	2.236	0.000	1.000	4.000	2.236

[2] The example calculation of the kriging weights has been adapted from Carr (1995).

(2,5)	1.000	1.414	1.000	0.000	3.000	1.414
(5,5)	3.162	2.236	4.000	3.000	0.000	2.236
(3,4)	2.236	2.000	2.236	1.414	2.236	0.000

Step 2, using the spherical semivariogram model parameters as described above, the semivariogram $\gamma(h)$ values are calculated for each of these coordinate pairs. For example, for the location pair of (2,6) (3,6):

$$\begin{aligned}\gamma(h) &= 20 + 120.0\ (1.5(1)/3 - 0.5(1^3)/27)\\ &= 20 + 120\ (0.5 - .0185185)\\ &= 20 + 120\ (.4814815)\\ &= 20 + 57.77778\\ &= 77.77778\end{aligned}$$

and rounding to two decimals gives: $\gamma(h) = 77.78$. Each diagonal entry $\gamma_{ij}(I,I)$, is the semivariance for a zero lag distance, which is zero ($\gamma(0) = 0$).

location	(2,6)	(3,6)	(1,5)	(2,5)	(5,5)	(3,4)
(2,6)	0	77.78	98.56	77.78	140.00	129.32
(3,6)	77.78	0	129.32	98.56	129.32	122.22
(1,5)	98.56	129.32	0	77.78	140.00	129.32
(2,5)	77.78	98.56	77.78	0	140.00	98.56
(5,5)	140.00	129.32	140.00	140.00	0	129.32
(3,4)	129.32	122.22	129.32	98.56	129.32	0

Step 3, calculating the covariance:

$$\text{cov}(h_{ij}) = \text{sill} - \gamma(h_{ij})$$

Because, the covariance of the zero lag distance is equal to the sill minus the semivariance, and the semivariance for a zero lag distance by definition is 0, each diagonal entry of the covariance matrix is equal to the sill value of the semivariogram. In this example the sill is 140.0.

location	(2,6)	(3,6)	(1,5)	(2,5)	(5,5)	(3,4)
(2,6)	140.00	62.22	41.44	62.22	0	10.68
(3,6)	62.22	140.00	10.68	41.44	10.68	17.78
(1,5)	41.44	10.68	140.00	62.22	0	10.68
(2,5)	62.22	41.44	62.22	140.00	0	41.44
(5,5)	0	10.68	0	0	140.00	10.68
(3,4)	10.68	17.78	10.68	41.44	10.68	140.00

These covariances are a function of distance between the six closest data locations and the estimation location.

i	X	Y	Z	Distance (km) to Estimation Point	Covariance
1	2	6	15.7	1.414	41.44
2	3	6	11.8	1	62.22
3	1	5	28.5	2	17.78
4	2	5	13.6	1	62.22
5	5	5	35.4	2	17.78
6	3	4	0	1	62.22

With all calculations required to assemble the matrix to solve for the kriging weights completed, the matrix is as follows:

location	(2,6)	(3,6)	(1,5)	(2,5)	(5,5)	(3,4)	Distance Covariance
(2,6)	140.00	62.22	41.44	62.22	0	10.68	41.44
(3,6)	62.22	140.00	10.68	41.44	10.68	17.78	62.22
(1,5)	41.44	10.68	140.00	62.22	0	10.68	17.78
(2,5)	62.22	41.44	62.22	140.00	0	41.44	62.22
(5,5)	0	10.68	0	0	140.00	10.68	17.78
(3,4)	10.68	17.78	10.68	41.44	10.68	140.00	62.22

An extra row , and an extra column, are added. The additional row assures that kriging weights sum to 1. The additional column maps intersample covariances to point-sample covariances using the Langrangian multiplier.

140.00	62.22	41.44	62.22	0	10.68	1	λ_1		41.44
62.22	140.00	10.68	41.44	10.68	17.78	1	λ_2		62.22
41.44	10.68	140.00	62.22	0	10.68	1	λ_3		17.78
62.22	41.44	62.22	140.00	0	41.44	1	λ_4	=	62.22
0	10.68	0	0	140.00	10.68	1	λ_5		17.78
10.68	17.78	10.68	41.44	10.68	140.00	1	λ_6		62.22
1	1	1	1	1	1	0	μ		1

Gaussian elimination, is used to solve for kriging weights.

λ_1		0.036
λ_2		0.316
λ_3		-0.039
λ_4	=	0.267
λ_5		0.090
λ_6		0.331

Gaussian elimination is mathematical method to solve simultaneous equations. Simultaneous equations are two or more algebraic functions having a common solution. In Gaussian elimination, unknowns associated with simultaneous equations are determined by solving one of the equations for one of the unknowns, then substituting this solution into all remaining equations. This is done repeatedly until only one unknown remains, at which point it is solved. Then, using this solution, other unknowns are solved through back calculation. A detailed discussion of Gaussian elimination is beyond the scope of this report. A good example of solving linear equations by Gaussian elimination is described by Calter (1990), pages 304 - 306.

In the example, solve for:

$$Z_{(3,5)} = \lambda(1)*Z_{(2,6)} + \lambda(2)*Z_{(3,6)} + \lambda(3)*Z_{(1,5)} + \lambda(4)*Z_{(2,5)} + \lambda(5)*Z_{(5,5)} + \lambda(6)*Z_{(3,4)}$$

An estimate for $Z_{(3,5)}$ is computed:

$Z_{(3,5)} = 0.036*15.7 + 0.316*11.8 - 0.039*38.5 + 0.267*13.6 + 0.090*35.4 + 0.331*0 = 9.6$

The Exact Interpolation Characteristics of Kriging[3]

If the interpolation algorithm adjusts the interpolated grid so that the interpolated surface agrees with the original data values, it is called an exact interpolator. If the algorithm does not force the grid to honor the original data points, it is called an inexact interpolator.

Choosing an exact interpolator over an inexact interpolator is a matter of philosophy not mathematics. If the data is believed to have random error or uncertainty, use an inexact interpolator. If the data is believed to be absolutely accurate and precise, and would be identical upon replication, use and exact interpolator. In a sense, using inexact interpolation shows a mistrust of the data; using an exact interpolation shows a mistrust of the algorithm. Your choice.

Kriging is an exact interpolator. The values of known data points are never altered and their location is never changed. Should a known data point fall exactly on a location in the target matrix, that value will be present in the final matrix. Most often the coordinates of few, if any data values, match the coordinates of an element in the target matrix. In that case all or nearly all values are estimated.

To accept that kriging is an exact interpolator you have to assume that there is no measurement error (replicate measurements) and that any sharp discontinuity at the origin (the nugget effect) is caused by small scale variations that cannot be resolved by the sample spacing.

To demonstrate the exact interpolation characteristics of kriging, we can use our example, an estimate can be made at the location (2,6) which coincides with a sample location having the data value of 15.7. In this case, using the seven closest data locations for estimation, including the location (2,6), the matrix system for kriging is as follows, using the semivariogram model:

140.00	41.44	62.22	41.44	62.22	62.22	41.44	1	λ_1		140.00
41.44	140.00	62.22	17.78	10.68	10.68	17.78	1	λ_2		41.44
62.22	62.22	140.00	62.22	41.44	17.78	10.68	1	λ_3		62.22
41.44	17.78	62.22	140.00	62.22	10.68	0.58	1	λ_4		41.44
62.22	10.68	41.44	52.22	140.00	41.44	10.68	1	λ_5	=	62.22
62.22	10.68	17.78	10.68	41.44	140.00	62.22	1	λ_6		62.22

[3] The example exact interpolation calculations have been adapted from Carr (1995).

41.44	17.78	10.68	0.58	10.68	62.22	140	1	λ_7	62.22
1	1	1	1	1		1	0	μ	1

Gaussian elimination is once again used to solved for the kriging weights yielding:

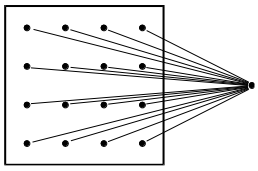
$\gamma 1$		1
$\gamma 2$		0
$\gamma 3$	=	0
$\gamma 4$		0
$\gamma 5$		0
$\gamma 6$		0

Using these weights for estimation, of course, yields an estimate exactly equal to the data value at the sample location, which coincides with the estimation location.

Punctual and Block Kriging

In the example calculation, the estimate of the regionalised variable obtained from kriging is a point, or punctual estimate. The estimate is valid only for the discrete point. 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.

A significant problem with punctual kriging is that it will not work unless the variable being mapped is stationary. In the presence of a trend a linear estimator is no longer unbiased. The computed estimates will be systematically shifted upwards or downwards from the true values depending upon the arrangements of the control points and the direction of dip of the surface. Punctual kriging is the most common method used in environmental engineering.



In some estimation applications, an estimate is desired that represents the average value of *Z* over a definable area. In this case, the estimate is referred to as average, or block kriging. By computing block estimates, the nugget effect, which may be due either to measurement error or very short range variation can be avoided. The interpolated surface from block kriging is smoother, and the longer range variation can be detected more easily. Consequently block estimates appear more reliable than those for points.

The choice of block versus punctual kriging should be made on the basis of sampling design and variate characteristics. If samples were taken to represent an area around the actual sample point (e.g. if soil samples from an area around the sampling coordinate were composited before analysis), then block kriging may be more appropriate than punctual. If samples were taken to represent point values in a field, or in time, then punctual kriging may be more appropriate.

Block interpolation (Burgess and Webster 1980) may be more appropriate than punctual interpolation where average values of properties are more

meaningful than exact single-point values, especially where spatial or temporal dependence is weak.

Block kriging is usually recommended for most environmental applications. Point kriging usually provides estimates very similar to those from block kriging, but if a point being estimated happens to coincide with a sampled location, the estimate is set equal to the sample value.

KRIGING PARAMETERS

Semivariograms are statistical measures that assume the input sample data are normally distributed and that local neighborhood means and standard deviations show no trends. Each sample data set must be assessed for conformity to these assumptions. Transformations of the data, editing of the data set, and the selection of different statistical estimators of spatial variability are all used to cope with data sets that diverge from the assumptions.

Sampling and Measurement Error

Knowledge of the sampling process along with the environmental processes influencing the sample attributes is critical to our interpretation of the characteristics of the data distribution. At the very least, saptial continuity detection is a tricky and time consuming process. Ancillary information and experience are the best guides for making inferences from the variogram. A good understanding of the underlying phenomenon and how it has been sampled is essential for effectively setting the parameters of the variogram model.

Measurement error affects individual samples, and may affect all samples because of the quality of the measuring device used. Measurement error also may be proportionately systematic, and the samples approximation to the true value may increase as the measured value increases. A sampling scheme, on the other hand, may systematically under or over sample natural heterogeneities of the geospatial phenomenon if not well matched. The same issues of precision, bias, and randomness in sampling for statistics apply to geostatistics as well.

Trends

Trends in data sets should be minimal when making inferences about spatial continuity patterns. A tend in the data has the effect of either a small or large upward shift in the variogram. Strong trending suggests that the process affecting the data distribution is nonlinear or multivariate. The variogram may remain resistant and stable under strong trending while the effect in the covariogram can be terrible (Cressie, 1993). If the trend is not constant across distance, then the variogram is poor or unrepresentative. Therefore, the data may need to be detrended.

When the local means and variability of the sample data are nonstationary, the first task is to identify the degree or pattern of the nonstationarity and decide whether to accommodate it. Does the nonstationarity disturb and

physically interfere with identifying the continuity structure? If nonstationarity intereacts with preferential clustering of data samples, then the variogram may be more sensitive.

When trending occurs, preferably the source of the trend is known, which will help influence the decision on how to handle it. Trending occurs for measurement reasons, sampling design, or actual physical qualities of the sampled object. Climatic data may follow a strong directional pattern. Increasing rainfall from west to east leads inevitably to constantly increasing local means in the same direction.

There are several quantitative management tools for trends. One possibility is to remove the trend component. The trend component can be subtracted from the samples before calculation of variogram measures and identification of the spatial continuity. The trend surface can be added back after kriging (Agterberg, 1974) or can be used as data input in universal kriging methods (Pebesma and Wesseling, 1998). Despite criticisms that the detrended data ofted have a different covariance function or variogram from the original data, they do have the same generalized covariance function, a simplified component of the covariance function which is swamped by redundant variability imposed by a trend (Kitanidis, 1993).

Co-Located Samples

If the samples represent replicates of the same sample collected at the same time, an accepted practice is to average the samples and only include the average in the data set. If the sample represents a different element or a different time of collection an accepted practice is to change the location coordinates slightly.

Spatial Outliers

Detecting special features, characteristics, or abnormalities of the data set are the first steps of exploratory data analysis which will influence variability/continuity interpretations of the variogram. Such characteristics that potentially influence the interpretation of spatial continuity are the clustering of sample values, breaks in the data distribution, spatial or directional trends in the local means, or erratic values. It is important to identify and distinguish problems or characteristics of individual samples from the problems or characteristics of the sampling scheme overall.

Special attention should be given to abnormal values appearing in the data set since these values will seriously distort the interpolation process. The best way to detect the abnormal values is to map or visually display the data set. Abnormal values will usually appear as spatial outliers.

Map display also provides additional insights about the data set, how the samples are spatially distributed (uniform or clustered), their general pattern, trend and the extreme high or low values associated with the data set. The spatial location of extreme values is helpful in detecting erroneous data. An isolated extreme value may be suspicious, but this may not be sufficient to justify its removal.

Depending on the number of points that deviate from surrounding points by more than the usual variance suggests is reasonable, using a higher nugget effect, a larger search radius, or a larger number of points in the neighborhood may be preferable.

Care should be taken before deciding that a given outlier is actually a bad data point. In some cases outlier values are obvious. There are a variety of methods that can be used to treat such a point, including inclusion, accommodation, replacement and rejection. The method to use depends on the objectives of the interpolation (depiction versus prediction; trend versus anomalies) and the nature of the data point. The decision to discard extreme values must be made with care, and data should be dismissed only if they are clearly wrong.

It is possible to set a missing value indicator prior to surface generation in some software packages. Missing values are ignored during analyses. The missing value indicator can be set to a value such as -99. Some programs can set the missing values to permanent (blank data cells) or have the ability to mark some values as temporarily missing.

Extreme Values

When extreme values or skewed distributions have an adverse impact on variogram calculations, they can be managed in several ways. Extreme values can be identified as errors and then eliminating them from the sample data set. The data set can be stratified by grouping extreme values together and treating them as a separate statistical population. The extreme values can be accepted, and their effect on the distribution can be mathematically moderated.

There are two sets of mathematical methods. The first is to transform the data set before applying statistical measures. The other is to apply robust statistical measures that are better designed mathematically to reduce the effect of outliers or a small number of samples.

Censored Data

Censored data is data having a <, > or estimated qualifier, as is typical with water quality data. An accepted practice is to convert not detected (ND) values to one-half of the detection limit; to convert less than detection (<DL) values to the detection limit; and to convert estimated (E) values to the value estimated.

Data Set Subdivision

Sample data often represent a phenomenon with a wide range in attribute values. Spatial clusters or strings of large or small data values may reflect the presence of another process or event. A subset of samples may represent the natural clustering of small areas of high or low concentrations which may be a realistic component of the overall distribution. One could consider treating these values as belonging to another statistical population.

If certain values represent a significant spatial cluster, it may make sense to model their variograms separately by first separating the sample data into separate distributions before applying statistics. The decision to split the data into more homogeneous subsets should be based on physical considerations and the density of the available data. Each subset should have enough data to allow the reliable inference of statistics for each population.

On the other hand, one may have to accept the realities of insufficient sampling if this is what accounts for the appearance of data clustering. If the data are indeed representative of the sampled process, removing or separating samples to improve the variogram does not make sense. One could instead remove the data pair that negatively impacts the calculation for a particular lag. In the h-scatterplot, the par will appear as an outlier. The solution, however, is not to remove the individual points. They have been paired with every other sample value too and hold useful information for other lags.

Sample and Class Size

Semivariogram calculation and interpretation is partly an art, and so the selection of sample and class size is somewhat arbitrary and subjective.

Variograms computed from small samples almost always appear erratic, and in many instances the principal cause is too few data. It has been shown (Webster and Oliver, 1992) that variograms computed on fewer than 50 sample points are of little value and that at least 100 data are needed. Their experiments suggest that for a normally distributed isotropic variable a variogram computed from a sample of 150 data might often be satisfactory, while one derived from 225 data will usually be reliable.

A class size that is too large forces data pairs of widely varying separation distances into a single lag increment. A class size that is too small often results in a relatively few pairs contributing to any one lag increment resulting in a “noisy” semivariogram plot. Class size is selected by experimenting (increasing or decreasing the value) and noting its influence on the semivariogram plot.

Optimally, a class size should be found where approximately the same number of pairs result for each lag increment. For some data sets, such a class size cannot be found due to sampling geometry. For these data sets a class size should be sought yielding an interpretable semivariogram (showing a clear sill value and range).

It is common that the sill will be reached in less than 20 lags (Myers, 1997). Typically, a lag distance slightly less than the mean or median distance is sufficient for the initial detection of a structure. A recommended (Oliver, Webster and Gerrard, 1989) 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. In general, 30 to 50 data pairs is consider a practical minimum number of pairs for obtaining a reliable estimate of the variogram at any lag (Armstrong, 1984 and Russo, 1984). The goal is to find lag distances that generate a relatively smooth continuous structure. Typically variograms should generally be limited to a maximum distance equal to half the sampled distance.

The number of distinct pairs in a lag neighborhood ideally would be at least 30 pairs. In statistics, this approximates the number of pairs needed to produce a normal distribution if drawing samples from a random distribution. Large tolerances can ensure large numbers of samples, but also may lead to unstable variograms that mask the continuity structure sought for a user-specified direction. If the angular tolerance is set to 180 degrees when specifying a direction for neighborhood selection, this is in effect, plotting an omnidirectional variogram.

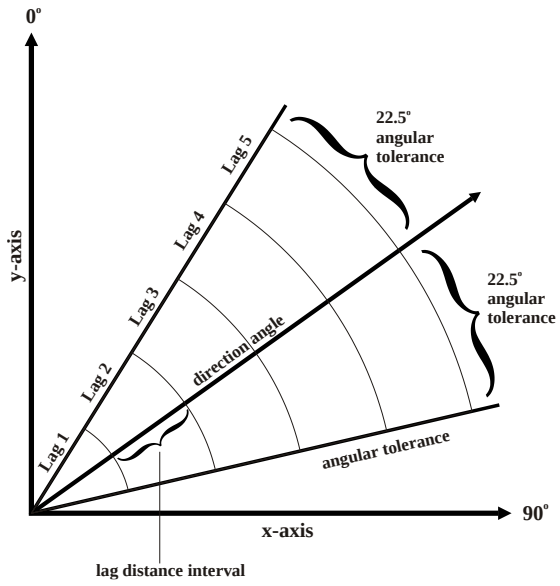
While less of an issue for omnidirectional than directional variograms, checking the number of pairs that occur within each lag is necessary to judge whether a sufficient sample is present. If the total sample is small, the lag division may affect the average of the variogram values significantly, producing erratic results. The advantage of the omnidriectional variogram is that it contains more sample pairs and more likely, the clearest "average" structure. If the structure of the variogram does not become evident, the lag distance parameters may be poor or there may be erratic pairs. Analysis of the data histogram and the exploration of *h*-scatterplots of each lag are some ways to explore and identify such problems.

If a regionalized phenomenon is known or suspected to be composed of several components, each having a different range (in the semivariogram sense), use differing class size values to resolve the spatial correlation of each component. If data are irregularly spaced or clustered, it may be necessary to set irregular width lag intervals. In this situation, a suggestion is to use lags with small separation distances at short distances and use lags with increasingly larger separation distances at longer distances. Changing the width of each lag may help improve detection of continuity.

Lag and Direction Tolerance

The distance between data pairs for the variogram are, in actuality, imprecise. Variogram values represent averages based on bounded separation distances (a range of distances) between samples limited by directional tolerances (or a range of directions).

If the data are irregularly spaced, there is no way to get pairs of values exactly 1 kilometer apart or exactly 2 kilometers apart. Normally the lag tolerance is defined to be half the lag increment used. For a lag increment of 1 kilometer, the lag tolerance would be ± 500 meters. Use all pairs separated 1 kilometer ± 500 meters apart. This ensures that there will be enough pairs for a particular lag distance. When the variogram is plotted, the lag distance will be the average of all the distances between pairs falling within that tolerance, so it wouldn't be 1 kilometer. The bigger the tolerance, the more pairs that can be defined, and the smoother looking the variogram. The smaller the tolerance, the fewer the number of pairs and the less smooth the variogram will appear.



Directional variograms analysis required additional lag parameters as well. Lag interval only separate samples by their separation distance. Distinguishing pairs by their separation in a particular direction narrows the selection process even further. In the same way that a distance interval is used to set a range of distances for a lag, an angular tolerance, and in some software a bandwidth are used for setting a range of angles.

By defining a tolerance for the direction, directional variograms can be calculated. Normally, an azimuth or direction to analyze is chosen, (for example, 45 degrees). Then all pairs in this direction, plus some tolerance on either side of the azimuth (for example, 22.5°) are taken and the variogram calculated. An angular tolerance is the range of angles for grouping data pairs on either side of the specified direction. So 22.5° on either side of 47° , for example, constitutes an angular range from 24.5° to 69.5° , or a total of 45° A tolerance of 90 degrees means an omnidirectional variogram, regardless of azimuth. The bigger the tolerance, the more pairs obtained.

Directional tolerance must be chosen carefully and the process of iteratively examining the *h*-scatterplots at different lag widths or angular tolerances is essential. Using wide tolerance angles can over generalize by incorporating more pairs. This is especially true when anisotropy is extreme.

Projection and Coordinate System

Sample data collected over large areal extents can pose challenges because measuring distances across large areas will cause problems in variogram reliability. A latitude/longitude coordinate reference system is not suitable. Ultimately, separation distances should be calculated using spherical distances for large areal expanses. A Lambert Conformal Conic projection can be used to reduce errors in distance calculation, but this may still create problems for the largest separation distances. Another option would be to break the sample data into separate smaller sections and to project each independently to reduce distortions.

Nugget Effect

The relative nugget effect tends to increase with the lag tolerance and with data scarcity. Typically, the relative nugget effect decreases as more and better data become available. The way in which the sample variogram behaves at near zero separation distances is critical in describing the spatial continuity.

A smoothing effect may be obtained by increasing the nugget, while leaving the total sill and the range the same. Consequently the ratio of the nugget effect C0 to total sill C0 + C increases. The effect is that points close to the grid node get a relatively lesser weight, whereas points farther away get relatively higher weight. The extreme would be to weight all points equally (pure nugget effect model).

Local scale variability in some sample data, like soil attributes, may create a natural nugget effect that should be modeled. While behavior of the sample variogram at shorter lag distances is hard to view because of fewer samples at close distances, it is one of the most important areas to accurately identify. Sometimes, sample pairs at relatively long separation distances are grouped with relatively shorter distance pairs and this creates an artificial nugget effect. Whether caused by sampling "noise" or true variability, nugget effects must be assessed carefully.

Search Neighborhood

Kriging seldom uses all of the sampled data. The more data is used in kriging, the bigger the matrix that would have to be inverted, and the more time consuming the kriging process. Using a search neighborhood, limits our estimates to just the data within some predefined radius of our point of estimation. This is called the search radius. The search radius does not need to be a circle or a sphere. To account for possible anistropies in the data the search radius can be elliptical or ellipsoidal (in 3D)

In some cases it is beneficial to use all of the data. This results in a unique neighborhood (all the data is used, for example, if the number of samples is less than 100). The advantage is that the kriging matrix is inverted only one. For more than 100 samples, it is recommended to use a plain search neighborhood. For kriging with finite ranges, sparse matrix techniques can also be used with a unique neighborhood and with a lot of data without excessive computer cost. For kriging without finite ranges, search neighborhoods would still be required.

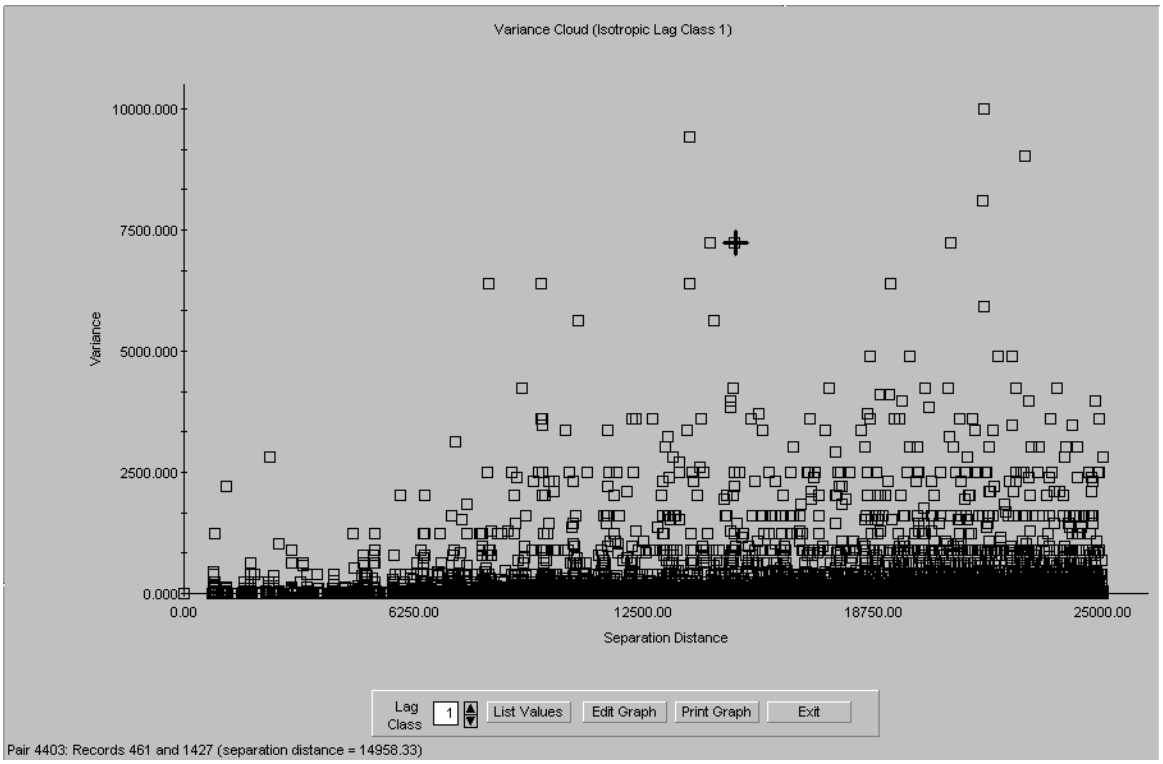
Also implicit in the use of the search neighborhood is that we are assuming stationarity only within the search window (quasi-stationarity). For example, this can be useful if there is a trend in the data. By assuming a small enough search window, such a trend component can be ignored in the estimation.

Variogram Model

If a smoother variogram model is used, map details are smoothed further (Davis, 1986). A smoother model can be a function of higher differentiability. For example, a Gaussian model as opposed to a spherical model with a steeper slope, or a model with a larger range or a higher nugget effect. It has been shown (Warnes, 1986) that the influence of changing variogram parameters on the kriged map is not drastic for the exponential variogram model but is substantial for the Gaussian model.

Variance Cloud

The variance cloud or variogram cloud is a graph of the variances for all individual pairs of point in an autocorrelation analysis. Each sample data point is matched with each and every other sample data point and produces a variogram value for each resulting pair. The result is displayed by locating variogram values according to their separation distance and direction (the separation vector). In the example, the cloud is specific to the first lag class of the isotropic variogram, and the mouse in on pair 4403, made up of data records 461 and 1427, which are separated by 14958.33 meters.



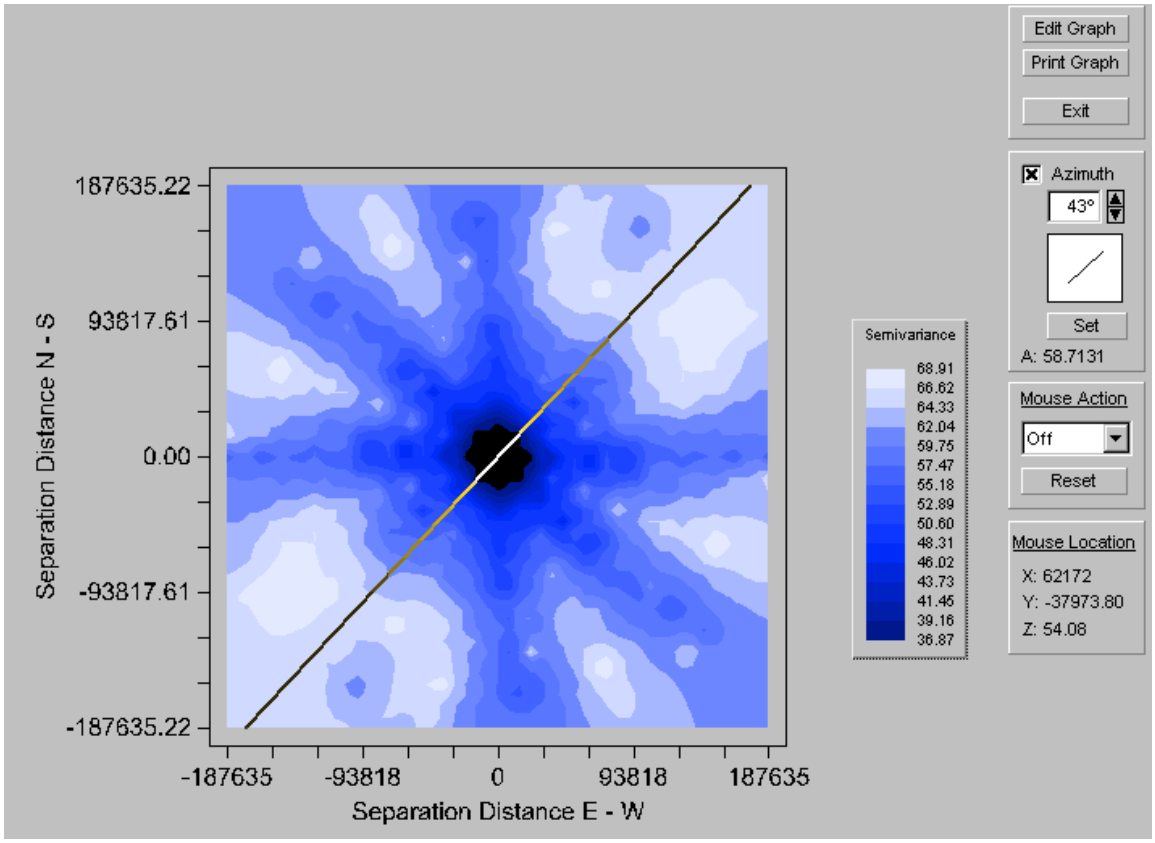
The variance cloud is particularly useful for discovering outliers that may inappropriately skew the average value for a lag class. The variance cloud is specific to both direction (isotropic or a specific anisotropic direction) and to a particular lag class.

Variogram Map

The variogram map or anisotropic semivariance surface provides a visual picture of semivariance in every compass direction. The variogram surface is a representation of statistical space based on the variogram cloud.

A transect in any single direction (e.g. 43 degrees) is equivalent to the variogram in that direction. The surface (z-axis) is semivariance; the x and y axes are separation distances in E-W and N-S directions, respectivley. The center of the graph corresponds to the origin of the variogram $\gamma(h)=0$ for every direction, from which lag distances increase outwardly in all directions. Each location on the graph represents an approximate average of the pairs' semivariance for the set of pair separation distances and directions.

The variogram map provide a visual guide to establishing the appropriate principal axis for defining the anisotropic variogram model.



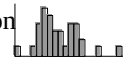
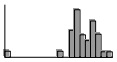
Data Transformations

Sometimes the best representation of a regionalized variable cannot be obtained by using the original data units. If the data are positively skewed, as is common with contamination concentrations, a logarithmic transformation of the data is generally beneficial. The purpose of data transformation is to moderate or reduce the effects of imperfect distributions in the data. A lognormal tranformation can reduce the erratic behavior of variogram models and make the continuity structure more apparent.

Statistical evaluation is to understand statistical properties associated with the data. These properties may include distribution, location, spread, and shape of the data set. There are many tools available in the univariate description statistics which can be used to describe these properties. For example, the frequency table and corresponding histogram can be used to describe how often observed values fall within certain intervals or classes. Probability plots can be used to determine how close the distribution of the data set is to a Gaussian or normal distribution.

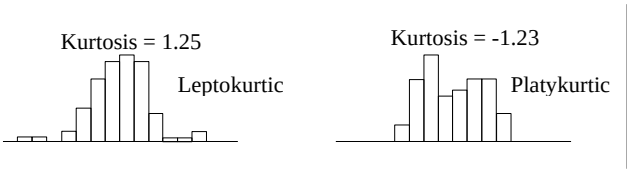
Summary statistics provide a valuable summary of information contained in the histogram. It includes three groups: measures of location, measures of spread, and measures of shape.

Summary Statistics

	Untransformed	Transformed
mean	22.383	2.9600
std deviation	10.822	0.6650
sample variance	117.112	0.4423
minimum value	0.00	0.000
maximum value	55.10	4.009
n (n missing)	36 (0)	36 (0)
frequency distibution		
skewness (se)	0.73 (0.39)	-2.39 (0.39)
kurtosis (se)	0.89 (0.77)	9.03 (0.77)

Measures of location describe where various parts of the distribution lie. The center of distribution is described by the mean, the median, and mode. The location of other parts of the distribution are given by various quantiles. Measures of spread describe the variability. This group includes the variance, the deviation and the interquartile range. Measures of shape include coefficient of skewness and coefficient of variation.

Skewness measures the degree to which a distribution is asymmetric. If a distribution's mean and median do not coincide, the distribution is skewed. If the distribution's mean is greater than (to the right of) the median, the distribution is said to be skewed to the right. The tail of the distribution points to the right, and the mound of data is on the left. This is referred to as positive skew. The opposite is referred to as a negative skew. For a symmetric distribution, the skewness is zero. A simple formula for skewness is skewness = (mean - median) / standard deviation.



Kurtosis is another measure of a distribution's departure from normality. A distribution may have no skew - that is, may be symmetrical - yet still not be normal. Kurtosis measures how flat or peaked the distribution is. Or stated in another way, kurtosis measures how thin or thick the tails of the distribution are. Distributions that have thicker than normal tails are referred to as leptokurtic (kurtosis is greater than 3). Distributions with thinner than normal tails are platykurtic (kurtosis is less than 3). Distributions with normal tails are mesokurtic (kurtosis is 3). The two distributions illustrated here have the same variance, approximately the same skew, but differ markedly in kurtosis.

The computer can generate measures of skewness and kurtosis, as well as standard errors for them. Simply divide the skew (or kurtosis) figure by the standard error of the skew (kurtosis). Interpret the result like a Z score. That is, if it is greater than 1.96 or less than -1.96 (for alpha = .05), the data is significantly skewed (kurtotic).

There is no requirement for data to be normally distributed for use in the computation of semivariograms, or predictions mad by kriging. However kriging is a linear estimator and is sensitive to a few large samples, which may bias the results. Notably, the variogram may become unstructured, close to a pure nugget effect. Normality is rare in mining and environmental data.

Whether a distribution is normal, log-normal, or something else has no particular geostatistical significance. However, it is often more difficult to interpret variograms for highly skewed distributions such as the log-normal, and in such cases it may be useful to also compute variograms on log-transformed data.

Unfortunately, there are consequences of data transformation. The problem lies in the back-transform of results. For square root or logarithmic transforms, the back-transform through squaring or exponentiation tends to

exaggerate any error associated with the interpolation. Such exaggerations of errors is most dramatic for extreme values. Skewed data may be transformed to a normal distribution by a non-linear transformation (i.e., logarithmic). However, when the data are back-transformed to real values, the unbiased property of the kriging estimates is lost.

The appropriateness of data transformation should be made carefully and should call for more than a quick look at the shape of the sample histogram and the simple desire to make that sample histogram symmetric.

Lognormal Kriging

One reason for computing a histogram prior to geostatistical analysis is to identify the presence of unusual data values in the data set. In some cases measurement error, recording error, accidents, and so on, can result in unusual data values. Of course, unusual data values may be an inherent attribute of the phenomenon under study. Unusual data values have a detrimental influence on geostatistical analysis, especially on the computation of the semivariogram. For this reason, unusual data values must be identified and a decision must be made about how to treat them.

Another reason for computing a histogram for spatial data, prior to geostatistical analysis, is to identify the most likely distribution of the data. For normally distributed data, kriging is the best linear, unbiased estimator for the data. Kriging errors are normally distributed, and kriging has minimum variance of estimation error. For distinctly non-normal data sets, kriging may still be used, but it is not necessarily the best linear, unbiased estimator. A data transformation may be considered for distinctly non-normal data sets if the deviation of the data set from a normal distribution is determined to be problematic for estimation.

If the data distribution is not examined by a histogram prior to geostatistical analysis, severe errors may occur. The application of kriging to a strongly skewed data set will result in estimates whose distribution is more normal. In other words, the distribution of the estimates will not match that of the original data.

For data associated with decidedly non-normal distributions, a data transformation should be considered. Transformation is necessary to make the distribution more symmetrical and to remove the trend in variance. In a log-normal distribution variance is proportional to the mean squared. Thus, in high-density areas, the variance is higher than in low density areas. After log-transformation the variance become uniform.

Many environmental parameters are lognormally distributed, and in such cases, lognormal kriging will often be useful. Lognormal kriging is a general term for kriging with the logarithms of the data. Lognormal kriging is performed in the same manner as simple and universal kriging, except that the logarithms of the data are used instead of the original data. The resulting estimates are thus expressed in logarithmic units. To obtain estimates in the original units, the transformation must be reversed, and both the logarithmic estimate and variance are needed to transform the estimate back to the original units.

When a transformation is chosen, after analysis of the transformed data the output data are customarily (but not necessarily) back-transformed to the original data domain. Three potential backtransformations are none, exp(z), or weighted. If zero values are present, use transformation: log(N+1). Offset values are subtracted from the backtransformed values. The weighted backtransformation is a complex backtransformation that more closely approximates true population statistics than simple backtransformations. (Haan, 1977 and Krige, 1981).

Kriging a log-transformed variable does introduce a bias. Simply kriging a log-transformed data set would introduce a bias, because in the kriging, we're assuming that the prediction is the mode/mean/median of our minimized error distribution. But, by transforming the transformed value back to its original value, it wouldn't correspond to the mean anymore (probably near the median)! Therein lies the bias.

The lognormal transform is referred to as a "sledgehammer" approach to data (Myers, 1997). It can reduce significantly or even eliminate true distribution characteristics of the data. It is also believed that log transforming data often goes too far, especially if data are not lognormally distributed originally (Cressie and Hawkins, 1980). They propose adjusting variability measures instead. Specifically, they propose a robust variogram estimator as a way of including extreme values while lessening their impact.

Indicator Kriging

Indicator kriging is the ordinary kriging of indicators. Instead of kriging continuous variables, indicator (binary) variables are kriged. These variables take a value of 0 or 1. The 0's may represent values below a cut off, and the 1's represent values above a cut off. The resulting variogram of these 0's and 1's is the indicator variogram. An advantage of the indicator transform involves minimizing the influence of outliers or to contend with skewed distributions.

These values can be viewed as probabilities of occurrence, with a higher value denoting a higher probability or vice versa. An indicator is like a threshold value. The final result of indicator kriging is a set of probabilities that a grid cell exceeds a specific set of cut-off values with the values ranging between 0 and 1. This method is especially useful when a decision will be made based on a specific cutoff, or if the probability of exceeding some threshold is the final use of the interpolated data.

The indicator variogram is calculated on the transformed data samples. The result for each lag is a direct measure of how often paired samples separated by vector *h* have different attributes. The ranges and shapes of the directional indicator variograms reflect the spatial continuity patter for the specified threshold.

Although sampling error is not taken into account, the indicator transform can improve estimation of spatial continuity parameters in two ways. By reducing the scale of variogram, the behavior of sample variograms frequently improves. Only the position of the threshold value relative to each data point is considered in the summary statistics produced for each lag.

Being an arithmetic average, kriging can be drastically affected by even a single outlier. Deleting outliers is one approach to dealing with unusual values, but this may not be desirable because data are thrown away. Alternatively an indicator transform may be applied to the data, because no matter how unusual the data values are, they are transformed to either 0 or 1. Since the indicators are 0 or 1, the indicator variogram is very well behaved and resistant to outliers. Indicator variograms have a linear relationship (Lemmar, 1985).

When the threshold level does not correspond well with continuity structure (when a skewed proportion of the data fall into one level relative to the other), the indicator variogram will not produce good results. Experience (Myers, 1997) shows transforming the data into indicators using the median of the distribution as the threshold can provide a variogram revealing the approximate description of variability, which at times is a reasonable approximation.

When successful, the indicator transform potentially improves the overall estimation of continuity structure, and ultimately, the fit of the model selected for interpolation. While indicator variograms allow one to visualize the spatial continuity structure of different strata of data, it is also possible to customize the kriging interpolation to multiple strata. The kriging process can provide an independent summary of the statistical consistency of each spatial continuity model's fit to each stratum of data.

Co-Kriging

When studying two or more regionalized variables which are correlated with each other, a technique called co-kriging can be used. Co-kriging, uses one set of data to help explain and improve the description of variability in another. Co-kriging assumes that the second data set is highly correlated with the primary data set to be interpolated.

For example, when interpolating a rainfall surface from point rainfall data, incorporating a highly correlated variable such as elevation could help improve the estimation of rainfall. In such a case where the correlation is known, sampled elevation data could be used to help in the prediction of a rainfall surface, especially where rainfall sampling is space. In this case co-kriging can be attempted. To perform cokriging, we need to model not only the variograms of the secondary and primary data, but also the cross-variograms between the primary and secondary data.

Probability kriging or co-kriging is a more sophisticated version of indicator kriging. Whereas indicator kriging uses only information from the indicator values, probability kriging uses the additional piece of information about the distribution. The regionalized variables are sorted by increasing value and assigned their cumulative frequency. This is referred to as the rank order transform. The conditional distribution function is then estimated.

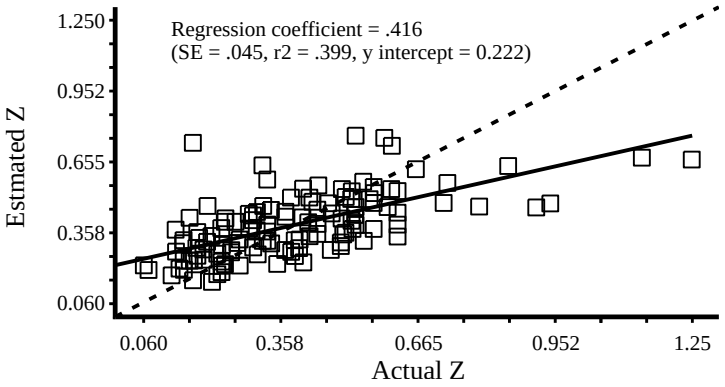
The weights are determined by the geostatistical technique called co-kriging. The use of co-kriging requires the variogram of the indicator variable, the variogram of the rank order transform variable and the cross variogram between the indicator and transformed variable. Since the co-kriging must be done for a discrete number of cutoffs, there are a large number of variograms required to implement probability kriging. If ten cutoff values are used, then

twenty-one variograms are necessary. Since probability kriging uses additional information it should theoretically produce better estimates than indicator kriging. (Knudsen, 1987).

Cross-Validation

Cross-validation is used to compare the impact of different models on interpolation results. In cross-validation, a variogram model and a search neighborhood are specified. Data values are then kriged at each sampled location, assuming that particular sample is missing. Then the kriged values and the true values are compared. The difference between these two values is called the cross-validated residual. Cross-validation techniques can be used to evaluate the impact of different kriging parameters on the interpolation results.

Before proceeding with correlation analysis it is often useful to examine the graphical representation of the data. The plot of the measurements of one variable against the corresponding measurements of the second variable will result in a scatter diagram. The smaller the dispersion of the points in the diagram, the more closely the variables are related. The shape of the pattern indicates the nature of the relationship. If the shape is a straight line, the relation is said to be linear.



If all estimated values are the same as the true value the scatter plot would be a 45-degree straight line. The distance each point is from the 45 degree line is the estimation error for that sample value. The closer the cloud of points are to this line, the better the estimation. In addition, the correlation coefficient can also be used as a reliable index for measuring how close the interpolated values are to the true values.

Ideally, points should plot close to a 45-degree line passing through the origin. Systematic deviations from the line indicate lack-of-fit with the model. Individual points deviating from the line may be considered outliers.

These residuals can be used to assess the performance of the model assumptions. Is the variogram model appropriate? Is the search neighborhood the right size? Is the data being over- or under-estimated? Is a trend model using universal kriging appropriate? Are prediction errors comparable? Cross validation is used to evaluate inconsistencies, and to ultimately decide whether to modify or reject the model, not to evaluate prediction accuracy. Cross-validation does not prove that the variogram model is correct, merely that it is not altogether incorrect.

Correlation Analysis

Correlation analysis is used to determine the manner in which one (an independent) variable affects another (dependent variable). In correlation

analysis, the coefficient of correlation, coefficient of determination, and standard error of estimate provide assessments of the reliability.

Coefficient of Correlation

Correlation is the linear association between two random variables X and Y. It is usually measured by a correlation coefficient, such as Pearson's r, such that the value of the coefficient ranges from -1 to 1. A positive value of r means that the association is positive. If X increases, the value of Y tends to increase linearly, and if X decreases, the value of Y tends to decrease linearly. A negative value of r means that the association is negative. The larger r is in absolute value, the stronger the linear association between X and Y. If r is 0, X and Y are said to be uncorrelated, with no linear association between X and Y. Independent variables are always uncorrelated, but uncorrelated variables need not be independent.

A value reasonably close to +1 or -1 can occur even though the scatter plot has a much more pronounced curve than linear pattern, because a curve often can be well approximated by a straight line over a wide range of value. A zero correlation does not necessarily mean there is no relationship between X and Y. It means there is no linear relationship. It is always a good idea to have a scatter plot available to help with the interpretation.

Deciding if *r* is "good enough" is an interpretive skill developed through experience. As a rule of thumb, if r> 0.95, there is good evidence for positive linear correlation, if r< -0.95, there is good evidence for negative linear correlation, and if -0.95 < r< 0.95, the data is either not linear, or "noisy". As an informal rule of thumb, call the relationship strong is r is greater than +0.8 or less than (more negative than) -0.8, weak if r is between -0.5 and 0.5, and moderate otherwise.

Correlation (Pearson’s or any other) measures the extent of association, but association does not imply causation. If frequently happens that two variables are highly correlated not because one is causally related to the other but because they are both strongly related to a third variable.

Coefficient of Determination

In a simple correlation, the proportion of the variation of the dependent variable that is explained by the linear regression equation is the coefficient of determination r^2, which is the square of the correlation coefficient r.

The coefficient of determination is a measure of the proportion of the variance of the dependent variable that is explained by the regression equation. If the coefficient of determination $r^2 = .399$, then 39.9% of the variance in the dependent variable is explainable by the variance in the independent variable. The better the fit, the larger the r^2. A correlation coefficient of one (perfect correlation) is rarely, if ever, achieved. R^2 should be at least 0.6 and preferably between 0.8 and 1.0 for a reliable surface model.

Standard Error of Estimate

The standard deviation of the residuals (the difference between the observed and predicted values). The standard error of estimate (SE), the Root Mean Square Error is a measure of the variation of the points about the regression

line and is analogous to the standard deviation. However, the standard deviation measures the scatter about the arithmetical mean of a sample whereas the standard error of estimate measures the scatter of the dependent variable about the regression line.

The size of the standard error of estimate provides a measure of the scatter of the observed data about the regression line. However, it does not indicate how much improvement has been obtained by estimating the dependent variable from the regression curve, rather than assuming it equal to the mean of the dependent variable. The standard deviation is a measure of the error in assuming the dependent variable equal to its mean value. The ratio of the square of the standard error of estimate to the square of the standard deviation or variance of the dependent variable provides an estimate of this improvement. The coefficient of correlation measures the effectiveness of the regression.

If the standard error of estimate is equal to the standard deviation of the dependent variable, then the correlation coefficient is equal to zero. In this case the regression equation is no better for estimating the dependent variable than the mean of the dependent variable and in fact there is no correlation between the two variables. On the other hand, if the standard error of estimate is zero, then the correlation coefficient is equal to one. The regression equation therefore accounts for all the variance of the dependent variable and all the observed data will lie on the regression line. The standard error of estimate provides a measure of the closeness of the point to the curve relation, it does not show the degree of correlation.

The standard error of estimate should be less than about 30% of the acceptable range for values predicted from the surface model.

Measures of Autocorrelation

Crop-yield and soil-electric conductivity data mapped as a grid data structure produces a data set that continuously describes a mapped variable. This type of data structure is radically different from point-samples data, as they fully capture the spatial relationships throughout an entire area.

The analysis techniques for spatial dependency in such data involve a moving a “roving window” throughout the data grid. Map values for the center cell and its neighbors are retrieved from storage, and depending on the technique, the values are summarized. The window is then shifted over to the next cell, and the process is repeated until all map locations have been evaluated. Various methods are used to deal with incomplete windows that occur along map edges and areas of missing data. If two cells are close together and have similar values, then they’re considered spatially related; if their values are different, they’re considered unrelated or negatively related.

Geary’s C and Moran’s I are the most frequently used measures for determining spatial autocorrelation in mapped data.

Geary’s C compares the squared differences in values between the center cell and its adjacent neighbors to the overall difference based on the mean of all the values. If the adjacent differences are less , then there is a positive correlation. If the differences are more, then there is a negative correlation. And if the adjacent differences are about the same, then things are unrelated.

Moran’s I is a similar measure, but relates the product of the adjacent differences to the overall difference. The Moran's I statistic is similar in interpretation to the Pearson's Product Moment correlation statistic for independent samples in that both statistics range between -1.0 and 1.0 depending on the degree and direction of correlation.

The general interpretation of the Geary’s C and Moran’s I statistics can be summarized as follows:

autocorrelation	Moran’s I	Geary’s C
strong positive autocorrelation	$I > 0$	$0 < C < 1$
random distribution of values	$I = 0$	$C = 1$
strong negative autocorrelation	$I < 1$	$C > 1$

Practical Tips

- state the objective of the modeling exercise in clear quantitative terms
- select a sample size and a sample collection strategy that are consistent with the objectives
- try to obtain a sample size of 100 data points or more
- check for enough number of pairs at each lag distance (a minimum of 30 to 50 pairs)
- correct data errors
- manage extreme values or remove outliers
- possibly transform the data of skewed distributions (i.e, logarithmic transforms; indiator)
- truncate at half the maximum lag distance to ensure enough pairs
- use a larger lag tolerance to get more pairs and a smoother variogram
- start with an omnidirectional variogram before proceeding with directional variograms
- select an appropriate model that is appropriate for the data and the objectives
- use other variogram measures to take into account lag means and variances
- interpret the model to obtain the best reliable fit possible
- understand the uncertainty and the precision of the predictions
- make decisions that are sensible

- reevaluate the model with new data when environmental conditions change

Developing the interpolated surface is an iterative process of testing different distance and sample pair inclusion parameters, exploring the removal of unusual values, visualizing results based on transformed data samples, and switching variogram estimators. The more that is learned about the data, the better judgement can be made about which adjustments are realistic and most effective in determining a continuity pattern.

CONCLUSION

For many years geographers have had to be content with the qualitative description of variables over the earth's surface. Many techniques have been devised for interpolation and mapping of data. The natural properties of the earth's surface seem to behave essentially as spatially dependent random variables. The theory of regionalised variables describes the kind of variation that is characteristic of many properties of the earth's surface. All can be treated as spatially dependent random variables. Geostatistics is the practical application of this theory.

Kriging is an advanced technique that relies heavily on statistical theory and computer processing. In principle, kriging is the ideal interpolator. An additional bonus is that the method yield estimates of the errors associated with each interpolation.

Geostatistics has been around since the early part of the twentieth century, but only in the last few decades has it been applied to environmental studies. Since geostatistics has to potential to be a very useful tool for environmental research the technology will almost certainly become an essential part of the GIS tool kit in the beginning of the twenty-first century.

The information presented herein was developed to provide a basis of knowledge for those interested in using kriging and geostatistics to either design or to analyze the data from environmental monitoring surveys. Hopefully no tears were shed during this learning process.

Rather than simply seeking to produce a visually-pleasing interpolated surface, geostatistics provides a large collection of tools for exploring and understanding the nature of a data set. The bottom line is, do the geostatistical tools produce useful results? "Useful" has to be decided by the user. Across a wide spectrum of applications the answer seems to be yes but in specific instances the answer may be not because of a lack of data, difficulty in estimating or modeling the variogram, the sensitivity of any linear estimator to unusual data values, etc.

There are no "correct answers in geostatistics, only the opportunity to gain more knowledge about the data and the interopolated surface, and to improve on the surface model.

REFERENCES

Agterberg, F.P., 1974. Geomathematics, Elsevier Science Inc.

Armstrong, M., 1984. Common problems with universal Kriging, Math. Geol 16(1): 101 - 108.

Ashraf, M., Loftis, J.C., and Hubbard, K.G., 1997. Application of Geostatistics to Evaluate Partial Weather Station Networks. Agricultural & Forest Meteorology. 84(3-4): 225-271.

Acevedo. M.F., 1997. Geog 5190 Advanced Stats and Quantitative Methods, Lecture Outlines, Lab Guides and Assignments. http://www.geog.unt.edu/~acevedo/courses/590/week10/lectur10.htm.

Atkinson, P.M. and Lloyd, C.D. 1998. Mapping Precipitaion in Switzerland with Ordinary and Indicator Kriging. Journal of Geographic Information and Decision Analysis, Vol. 2, No. 2.: 72-86. ftp://ftp.geog.uwo.ca/SIC97/Atkinson/Atkinson.html

Bastin, G., Lorent, B., Doque, C., and Gevers, M., 1984 Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations, Water Resources Res. 20(4): 463 - 470.

Berry, J.K, 1999. Extending Spatial Dependency to Maps. Beyond Mapping, Geoworld, Vol 12 (1): 26-27.

Binkley, M.R., Carlson, R. and Lee, J., 1997. Spatial Interpolation of Lake Water Quality Data. Kent State University. http://humboldt.kent.edu/~binkley/poster.html.

Bogaert, P., P. Mahau, and F. Beckers. 1995. The Spatial Interpolation of Agro-Climatic Data – Cokriging Software and Source Code, Use' Manual, Version 1.0b. Unité de Biométrie, Faculté des Sciences Agronomiaques, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. Agrometeorology Series Working Paper, Number 12, FAO Rome, Italy.

Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists: Modelling with GIS, Pergamon, Tarrytown, New York.

Boomer, K.M. and Brazier, R.A., 1997. Geostatistics, Seismic Risk Analysis of Northern Arizona from Acceleration Contour Maps, Enhancing the Sampling Procedure through a Geostatitstical Analysis. http://www.geo.arizona.edu/geophysics/students/brazier/geo.html.

Brill, D.J. 1997. Skewness and Kurtosis - A Discussion http://www.uark.edu/plscinfo/pub/methods/help/skkrdis.htm

Bruce, J.P. and Clark, R.H. 1980. Introduction to Hydrometeorology. Pergamon Press, Ontario. 324 pages.

Burgess, T.M. and Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties I. The semivariogram and punctual kriging. Journal of Soil Science 31: 315-331.

Burgess, T.M. and Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties II. Block kriging. Journal of Soil Science 31: 333- 341.

Calter, P., 1990. Technical Mathematics with Calculus. Prentice Hall, Engelwoods Cliffs, New Jerey.

Carrat, F. and Valleron, A., 1992. Epidemiological Mapping Using the "Kriging" Method: Application to an Influenza-like Illness Epidemic in France. American Journal of Epidemiology, 135(11): 1293 - 1300.

Carr, James R., 1995. Numerical Analysis for the Geological Sciences. Prentice Hall, Englewoods Cliffs, New Jersey.

Chua, S. H. and Bras, R. L., 1982. Optimal estimators of mean areal precipitation in regions of orographic influence. J. Hydrology, 57: 23 - 48.

Chang, T.J. and Teoh, C.B., 1995. Use of the Kriging Method for Studying Characteristics of Ground Water Droughts. Water Resources Bulletin, American Water Resources Association, Vol. 31, No. 6: 1001- 1007.

Cox D.D., Cox L.H., and Ensor, K.B., 1997. Spatial Sampling and the Environment - Some Issues and Directions. Environmental & Ecological Statistics. 4(3): 219 -233.

Cressie, N. 1985. Fitting variogram models by weighted least squares. Mathematical Geology 17: 563 - 586.

Cressie N., 1990. The Origins of Kriging. Mathematical Geology, Vol. 22, No. 3, 1990: 239 - 252.

Cressie, N.A.C. 1991. Statistics for Spatial Data. John Wiley, New York.

Cressie, N., 1993. Statistics for Spatial Data. 2nd edition. John Wiley and Sons, Inc, New York.

Cressie, N.A.C. and D.M. Hawkins, 1980. Robust Estimation of the Variogram. Journal International Association Mathematical Geology 12:115-125.

Davis, J.C., 1973. Statistics and data analysis in geology. John Wiley and Sons, New York.

Davis, J.C., 1986. Statistics and Data Analysis in Geology, 2nd edition, Wiley, New York.

Delhomme, J.P., 1979. Kriging in the Hydrosciences, Advance in Water Res. 1(5): 251-266

Deraisme. J., 1997. Interpolation and Reality. Geovariances News Letter #7, page 3. http://www.geovariances.fr/publications/n1/7/page3.html.

Deutsch, C.V., 1996. Correcting for Negative Weights in Ordinary Kriging. Computers and Geosciences Vol. 22, No. 7: 765 - 773

Dorsel, D. and La Breche, T., 1998. Environmental Sampling & Monitoring Primer, Kriging. http://www.ce.vt.edu/enviro2/smprimer/kriging/kriging.html

Eastman, J., 1999. IDRISI32, Guide to GIS and Image Processing – Volume 2, p 139 and p 155-161. Clark Labs, Clark University.

Eastman, J., 1999. IDRISI32, Tutorial. 179-201. Clark Labs, Clark University.

Freek, D., 1997. Classification of High Spectral Resolution Imagery Using an Indicator Kriging Based Technique. International Institute for Aerospace Survey and Earth Sciences (ITC), Department of Earth Resources Surveys. 350 Boulevard 1945, P.0. Box 6, 7500 AA, Enschede, The Netherlands.

Gadin, L.S., 1963. Objective analysis of meteorological fields: Gidrometeorolgicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965, 238 pages.

Gamma Design. GS+ for Windows is a Windows 95/98/NT version of a geostatistical mapping program for the PC. http://www.gammadesign.com.

Gambolatic, G. and Volpi, G., 1979a. A conceptual deterministic analysis of the kriging technique in hydrology, Water Resources Res. 15(3): 625 - 629.

Genoton,, M.G. and Furrer, 5. 1998. Analysis of Rainfall Data by Simple Good Sense: is Spatial Statistics Worth the Trouble? Journal of Geographic Information and Decision Analysis, Vol. 2, No. 2.: 11 - 17. ftp://ftp.geog.uwo.ca/SIC97/Genton_1/Genton_1.html

Genton, M.G. and Furrer, R. 1998. Analysis of Rainfall Data by Robust Spatial Statistics using S+Spatialstats. Journal of Geographic Information and Decision Analysis, Vol. 2, No. 2.: 126 - 136. ftp://ftp.geog.uwo.ca/SIC97/Genton_2/Genton_2.html

Goodchild, M. 1999. Spatial Analysis and GIS. 1999 ESRI User Conference, Pre-Conference Seminar.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York. 483 pages.

Griffith, D.A. 1987. Spatial Autocorrelation. A primer. Association of American Geopgraphers, Washington D.C.

Griffith, D.A., Amrhein, C.G. and Desloges, J.R., 1991. Statistical Analysis for Geographers. Prentice Hall, Englewoods Cliffs, New Jersey.

Guocheng, P. Garrd, D. Moss, K. and Heiner, T., 1993. A Comparison Between Cokriging and Ordinary Kriging: Case Study with a Polymetallic Deposit. Mathematical Geology, Vol. 25, No. 3: 377 - 398.

Haan, C.T. 1977. Statistical Methods in Hydrology. Iowa State University Press, Ames, Iowa.

Hansen, R.O., 1993. Interpretive Gridding by Anisotropic Kriging. Geophysics, Vol. 58, No. 10: 1491 - 1497.

Herzfeld, U.C, Eriksson, M.G., and Homlund P., 1993. On the Influence of Kriging Parameters on the Cartographic Output – A study in Mapping Subglacial Topography. Mathematical Geology, Vol. 25, No. 7: 881 - 900.

Holawe, F. and Dutter, R. 1999. Geostatistical Study of Precipitation Series in Austria: Time and Space. Vienna University of Technology.

Hu, J., 1998. Methods of Generating Surfaces in Environmental GIS Applications. Bechtel Environmental Inc., 151 Lafayette Drive, Oak Ridge, Tennessee 37922. http://pasture.ecn.purdue.edu/~edu/~aggrass/esri95/to100/p089.html.

Ingram, P., 1998. An Introduction to Geostatistics. School of Earth Sciences, and Kriging. Macquarie University, Sydney, NSW, 2109, Australia. http://137.111.98.10/users/pingram/geostat.html and http://atlas.es.mq.edu.au/users/pringram/krig.html#hrgprcss.

Isaaks, E.H., 1989. Applied Geostatistics. Oxford University Press, New York. 561 pages.

Isaaks, E.H. and Srivastava, R.M. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Jones, T.A., Hamilton, D.E., and Johnson, C.R., 1986. Contouring Geological Surfaces With the Computer. Van Nostrand Reinhold, London, England.

Kitanidis, P., 1985. 'Minimum-variance quadratic estimation of covariances of regionalized variables.' Mathematical Geology, 17(2):195-208.

Kitanidis, P.K., 1997. Introduction to Geostatistics: Applications to Hydrogeology. Cambridge University Press, New York. 249 pages.

Kitterod NO. Langsholt, E. Wong, W.K. and Gottschalk, L., 1997. Geostatistic Interpolation of Soil Moisture. Nordic Hydrology. 28(4-5): 307 - 328

Knudsen, H.P. And Baafi, E.Y., 1987. Indicator Kriging and Other New Geostatistical Tools. Pacific Rim Congress 87.

Kohl, M. and Gertner, G., 1997. Geostatistics in Evaluating Forest Damage Surveys - Considerations on Methods for Describing Spatial Distributions. Forest Ecology & Management. 95(20): 131 - 140.

Krige, D.G., 1951. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. Journal of the Chemical and Metallurgical Mining Society of South Africa 52: 119-139.

Krige, D.G. 1966. Two Dimensional Weighted Moving Average Trend Surfaces for Ore-Evaluation. Journal of the South African Institute of Mining and Metallurgy 66:13-38.

Krige, D.G. 1981. Lognormal-de Wijsian geostatistics for ore evaluation. South African Institute of Mining and Metallurgy Monograph Series. Geostatistics I. South Africa Institute of Mining and Metallurgy, Johannesburg, South Africa.

Kufs, C. 1987. Trend-Surface Modeling of Groundwater Data. Management of Uncontrolled Hazardous Waste Sites. Silver Springs, MD. Hazardous Materials Control Research Institute. http://members.home.net/terrabyte/t_surf.pdf

Kufs, C. 1998. Interpolation for Fun and Profit. http://members.home.net/terrabyte/didyou.htm

Lemmar, I.C., 1985, The Mononodal Cutoff-Application to Robust Variography and Spatial Distribution Estimation, Ph.D. Thesis, Stanford University.

Little, L.S., Edwards, D. and Porter, D.E., 1997. Kriging in Estuaries - As the Crow Flies, Or As the Fish Swims. Journal of Experimental Marine Biology & Ecology. 213(1):1-11.

Lynch, S.D. and Schulze, R.E., 1997. Techniques for Estimating Areal Daily Rainfall. Department of Agricultural Engineering, University of Natal, Private Bag X01, Scottsville, 3209, Sourth Africa. http://www.ccwr.ac.za/~lynch2/p241.html.

Matheron, G., 1962. Traité de geostatisque appliqueée, vol. I: Memories du Bureau de Géologiques et Miniéres, no 14, Editions Techniq, Paris. 333 pages

Matheron, G., 1963a. Principles of Geostatistics. Economic Geology 58: 1246 - 1266.

Matheron, G., 1963b. Traité de geostatistique appliqée, vol. II, Le krigeage: Memories du Bureau de Recherches Géologiques et Miniéres, no 24. Editions Bureau de Recherche Géologiques et Miniéres. Paris. 171 pages.

Matheron, G. 1971. The theory of regionalized variables and its applications. Cahiers du Center de Morphologie Mathematique, Fontainbleau, No. 5.

Maurer, B.A., 1994. Geographical Population Analysis. Blackwell Scientific, Oxford, UK.

Maurer, B.A. and Villard, M.-A., 1994. Population Density: Geographic Variation in Abundance of North American birds. National Geographic Research and Exploration 10: 306 - 317.

Mercer, W.B. and Hall, A.D., 1911. Experimental error of field trials. Journal of Agricultural Science, Cambridge 4: 107 - 132.

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 - 639.

Mitasova, H. and Mitas, L. 1993. Interpolation by Regularized Spline with Tension: I. Theory and Implementation. Mathematical Geology, Vol. 25, No. 6: 641-655

Mitasova, H. and Hofierka, J. 1993. Interpolation by Regularized Spline with Tension: II. Application to Terrain Modeling and Surface Geometry Analysis. Mathematical Geology, Vol. 25, No. 6: 657-669

Myers, J., 1997. Geostatistical Error Management: Quantifying Uncertainty for Environmental Sampling and Mapping, Van Nostrand Reinhold, New York.

Myers, D.E. 1999. What is Geostatistics? http://www.u.arizona.edu/~donaldm/whatis.html

Oliver, M.A. and Webster, R. 1991. How Geostatistics Can Help You. Soil Use and Management. Volume 7, Number 4, December 1991: 206 - 217.

Oliver, M., Webster, R. and Gerrard, J., 1989. Geostatistics in Physical Geography. Part I: theory. Trans. Inst. Br. Geogr. N.S. 14: 259-269 (1989) ISSN:0020-2754.

Oliver, M. Webster, R. and Gerrard, J., 1989. Geostatistics in Physical Geography. Part II: applications. Trans. Inst. Br. Geogr. N.S. 14: 270 -286 (1989) ISSN:0020-2754.

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): 107 - 125.

Pebesma, E.J. and C.G. Wesseling, 1998. 'Gstat: a program for geostatistical modelling, prediction and simulation.' Computers & Geosciences, 24(1): 17-31.

Robertson, G.P. 1987. Geostatistics in Ecology: Interpolating with Known Variance. Ecology 68:744 - 748.

Rosenbaum, M., and Söderström, M. 1997. Geostatistics as an Aid to Mapping. http://www.esri.com/base/common/userconf/europroc96/PAPERS/PN11/PN11F.HTM.

Rossi, R.E., Mulla, D.J., Journel, A.G. and Franz, E.H. 1992. Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence. Ecological Monographs 62: 277 - 314.

Royle, A.G., 1979. A Practical Introduction To Geostatistics. Department of Mining and Mineral Sciences, University of Leeds. 103 pages.

Russo, D., 1984. Design of an optimal sampling network for estimating the variogram. Soil Science Society of America J. 48: 708 - 716.

Shibli, S.A.R., 1997. Geostatistics FAQ. http://curie.ei.jrc.it/faq/index.html

Sharov, A., 1996. Quantitative Population Ecology. http://www.gypsymoth.ento.vt.edu/~sharov/PopEcol/popecol.html.

Sokal, R.R. and Oden, N.L. 1978 Spatial Autocorrelation in Biology 1. Methodology 2. Some Biological Implications and Four Applications of Evolutionary and Ecological Interest. Biological Journal of the Linnean Society 10:199 228.

Solow, A.R., 1990. Geostatistical Cross-Validation: A Cautionary Note. Mathematical Geology, Vol. 22, No. 6: 637-639.

Srivastava, R.M., 1988. A Non-ergodic Framework for Variograms and Covariance Functions. SIMS Technical Report No. 114, Department of Applied Earth Sciences, Stanford University.

Sterk, G. and Stein, A., 1997. Mapping Wind-Blown Mass Transport by Modeling Variability in Space and Time. Soil Science Society of American Journal. 61(1): 232 - 239.

Surfer. 1995. Surfer for Windows, Version 6 User's Guide – Contouring and 3D Surface Mapping. Golden Software, Inc. Golden, Colorado, 80401-1866, U.S.A.

Syed Abdul Rahman Shibli., 1998. Geostatistics FAQ. AI-Geostats. Desa Pandan 55100 Kuala Lumpur, Malaysia. http://curie.ei.jrc.it/faq/index.html.

Thomas, G.S., 1997. Interactve Analysis and Modeling Semi-Variograms. Snowden Associates Pty Ltd. Australia. http://www.metal.ntua.gr/msslab/MineIT97/papers/GT67/GT67.html.

Tobler, Waldo 1979: A Transformational View of Cartography, The American Cartographer, 6, 101-106.

Todini, E., and Ferrarcsi, M., 1996. Influence of Parameter Estimation Uncertainty in Kriging. Journal of Hydrology 175: 555-566.

Tomczk, M. 1998. Spatial Interpolation and it Uncertainty Using Automated Anisotropic Inverse Distance Weigthing (IDW) - Cross-Validation/Jackknife Approach. Journal of Geopgraphic Information and Decision Analysis, Vol. 2, No. 2: 18-33. ftp://ftp.geog.uwo.ca/SIC97/Tomczak/Tomczak.html.

Villard, Marc-André and Maurer, B.A., 1996. Geostatistics as a Tool for Examining Hypothesized Declines in Migratory Songbirds. Ecology, 77(1), 1996: 59 - 68.

Wackernagel, H., 1994. Cokriging Versus Kriging in Regionalized Multivariate Data Analysis. Geoderma, 62 (1994): 83-92.

Wackernagel, H., 1995. Multivariate Geostatistics: An Introduction with Applications. Springer, Berlin, New York. 256 pages.

Warnes, J.J., 1986. A Sensitivity Analysis for Universal Kriging. Math. Geol, Vol. 18, No. 7: 653 - 676.

Webster, R., 1985. Quantitative spatial analysis of soil in the field. Advances in Soil Science Volume 3. Springer-Verlag, New York: 1-70

Webster, R. and Oliver, M.A., 1992. Sample Adequately to Estimate Variograms of Soil Properties. Journal of Soil Science, 43, 177 - 192.

Webster, R. Oliver, M.A., Muir, K.R. and Mann, J.R., 1994. Kriging the local Risk of a Rare Disease from a Register of Diagnoses. Geographic Analysis, 26(2) 168 - 185.

Weber, D. And Englund D., 1992. Evaluation and Comparison of Spatial Interpolators. Mathematical Geology, Vol. 24, No. 4: 381 - 391.

Wiener, N., 1949. Extrapolation, interpolation and smoothing of stationary time series: MIT Press Cambridge, Massachusetts, 158 pages.

Wingle, W.L., 1992. Examining Common Problems Associated with Various Contouring Methods, Particularly Inverse-Distance Methods, Using Shaded Relief Surfaces. Geothech '92 Conference Proceedings, Lakewood, Colorado. http://magma.mines.edu/fs_home/wwingle/pub/contour/.

Zirschky, J., 1985. Geostatistics for Environmental Monitoring and Survey Design. Environmental International, Vol. 11: 515 -524.

Zrinji, Z., and Burn, D.H., 1992. Application of Kriging to Surface Water Level Estimation. Canadian Journal of Civil Engineering, 19: 181 - 185.

INDEX

T

U

V

Z