GEOSTATISTICAL MAPPING FOR HAZARDOUS WASTE SITES

Richard L. Satkin[1] and Aris P. Georgakakos[2]

AUTHORS: [1]Engineering-Science, Inc., 57 Executive Park South, NE, Atlanta, Georgia 30329, 404/325-0770; and [2]Associate Professor, Georgia Institute of Technology, School of Civil Engineering, 225 North Avenue, NW, Atlanta, Georgia 30332, 404/894-2000.
REFERENCE: Proceedings of the 1991 Georgia Water Resources Conference, held March 19 and 20, 1991 at The University of Georgia, Kathryn J. Hatcher, Editor, Institute of Natural Resources, The University of Georgia, Athens, Georgia.

ABSTRACT

Three commonly used techniques (inverse distance squared, linear interpolation, and kriging) for interpolating spatial data points were compared on the basis of robustness and accuracy. The results demonstrate that 1) kriging is a more robust contouring method; 2) kriging provides explicit measures of estimation accuracy; and 3) kriging balances the risks of misclassification (false positive and false negative error rates).

INTRODUCTION TO KRIGING

Geostatistical interpolation or kriging is a method for optimally estimating a quantity which is distributed in space and is measured at a network of points (Marsily, 1986). Kriging is applicable for defining contaminated areas in soils or groundwater, estimating piezometric surfaces, or mapping any spatially distributed parameter. The issue of delineating contaminated hazardous waste sites is relevant in Georgia as it is throughout the United States.

Geostatistical mapping is essentially completed in two steps (Figure 1): first is the identification of the correlation properties of the data; and second is the generation of estimates at each grid point.

The correlation function is a measure of the joint variability of data points at different locations. Suppose one is interested in the concentration of a chemical in soils. As the distance between two sampling points decreases, one would expect a higher correlation between concentration values. But as the distance increases, this correlation weakens and after a certain separation distance, called the range of influence, the concentration values are independent from one another. The spatial correlation properties of the data can be quantified by the correlation function or the variogram.

The second step is the generation of concentration estimates for each grid point in a user-defined network. The kriging estimates represent a weighted average of nearby measurements (Figure 1), where the λ_i weights are determined based on the variogram values at different distances. Clearly, points, X_1, X_3, and X_5, which are closer to the unsampled location at X_0, and having a higher

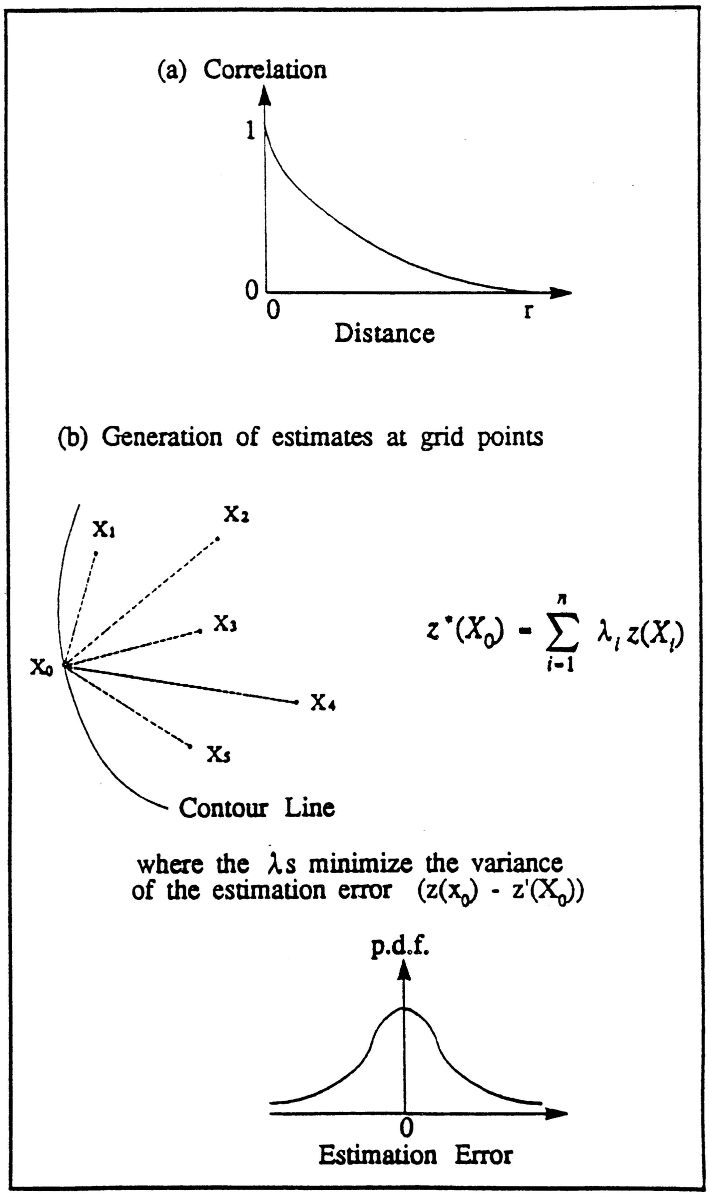

Figure 1. Process of Geostatistical Interpolation (Kriging).

correlation to X_0 than X_2 and X_4, will have a greater influence on the estimation of X_0. The λ_i weights are optimal in the sense that they minimize the variance of the estimation error. The estimation error is defined as the difference of the true value from the estimated one. The λ_i weights are chosen to make the average estimation error zero, indicating unbiasedness.

Two advantages of geostatistics over traditional contouring packages are 1) incorporates correlation structure of each data set; and 2) provides the variance of the estimation error and, therefore, allows for the computation of confidence intervals.

CASE STUDY

The objective of this investigation was to compare three commonly used interpolation methods for defining the areal extent of surface soil contamination. The first technique is a traditional inverse distance squared (IDS) method, the second is kriging, and third is linear interpolation. The IDS method provides grid estimates based on the neighboring measurements weighted according to the inverse of the distance squared. The software packages used were Geopack (EPA, 1990), a public domain software endorsed by EPA, and Surfer® (Golden software, 1989).

Suppose the cleanup level for this site was 5 ppm. Figure 2 shows the areas to be remediated by the two methods. The main observation is that the critical area determined by kriging is significantly smaller than the area delineated by the IDS method. This is a general result and is attributed to the data normalization process required by kriging. When data exhibit skewness toward the higher values, the critical areas are over-estimated by mapping methods which do not require data normalization.

Because of the difference in critical areas predicted by each contouring method, the question then arises which is the most accurate method. To address this issue we used two verification procedures. The first was to test the robustness of the contouring methods in relation to the number of data points used in the analysis. The second was to evaluate the accuracy of the areal extent estimates.

Comparison Based on Robustness

With regard to robustness,contours based on 100 percent of the data, 80 percent, 60 percent, 40 percent, and 20 percent of data were generated by three methods (kriging, IDS, and linear interpolation). An average point density for 100 percent data would correspond to a rectangular sampling network of approximately 30 by 30 feet; 80 percent corresponds to a sample density of 40 by 40 feet; 60 percent, 50 by 50 feet; 40 percent, 60 by 60 feet; and 20 percent, 70 by 70 feet. The data eliminated in each step were determined randomly. Table 1 summarizes the results by showing the areal extent of the 5 ppm contour lines. As expected, when the number of data points decrease, the estimated critical areas eventually decrease as well. Ultimately, the critical area approaches zero when no data points are available.

------ Kriging
______ Inverse Distance Squared

ESTIMATED AREAS ABOVE 5 PPM

Kriging- 1,064,770 FT²   Inverse Distance Squared-1,512,020 FT²

Figure 2. Comparison of 5 ppm Contour for Two Methods: Kriging and Inverse Distance Squared.

To evaluate the robustness of the methods in sparser data sets, we compared the areal extent differences with the base case where all of the data were included. A robust method is one which yields estimates which do not deviate significantly from the base case despite the reduction in the number of data points.

	% Data	Estimated Area (sq ft)		Normalized Bias/Mean Squared Error (MSE)
Kriging	100	1,049,980		1.00
	80	1,064,770		1.01
	60	1,028,590		0.98
	40	982,989		0.94
	20	839,831		0.80
			MSE	0.11
Inverse Distance Squared	100	1,518,310		1.00
	80	1,512,020		1.00
	60	1,455,930		0.96
	40	1,313,550		0.87
	20	1,074,830		0.71
			MSE	0.16
Linear	100	1,608,430		1.00
	80	1,583,710		0.98
	60	1,519,620		0.94
	40	1,376,510		0.86
	20	1,091,580		0.68
			MSE	0.18

Table 1. Summary of Contaminated Areal Extent Using Different Contouring Methods.

To illustrate the results, we also report the areal differences normalized to the base case (Table 1). The methods are evaluated according to the mean squared error (MSE) which is the average squared difference from 1.0 in each case. The kriging MSE is consistently lower than both the linear and the IDS contouring methods. This result indicates that kriging is a more reliable estimator of the true contour location. The linear and IDS methods perform comparably.

Another observation is that the base case results for each of the contouring methods are significantly different. For example, the 5 ppm areal extents differ by 30 percent for kriging and IDS methods. This difference will appreciably add to the remediation cost.

Comparison Based on Accuracy

With regard to the areal extent accuracy, the 5 ppm contour lines were compared to the data above and below 5 ppm. For kriging, 3 percent of the data above 5 ppm were not contained within the 5 ppm contour line (false negative), versus only 0.25 percent and 1 percent for the IDS and linear interpolation methods, respectively. This is something to be expected given the significantly larger areal extent of contamination that the IDS and linear interpolation methods predict. However, comparing the 5 ppm contour lines, with the data less than 5 ppm, kriging failed 2 percent of the time (false positive), versus 12 percent and 13 percent for the IDS and linear interpolation methods, respectively.

CONCLUSIONS

This case study demonstrates that kriging is a more robust contouring method than IDS or linear interpolation. Statistically, kriging attempts to maintain minimal comparable false negative and false positive errors. The probabilities of these errors cannot be simultaneously minimized. Since one error type decreases at the expense of the other, kriging-generated contours represent a statistically balanced approach. Please note, another version of kriging called probability kriging (Journel, 1988) can provide isopleths at any false negative or false positive error rate desired.

LITERATURE CITED

EPA, 1990. Geostatistics for Waste Management: A User's Manual for the GEOPACK (Version 1.0) Geostatistical Software System. EPA/600/8-90/004. Ada, Oklahoma.

Golden Software, Inc. 1989. Surfer (Version 4). Golden, Colorado.

Journel, A.G. 1988. Nonparametric geostatistics for risk and additional sampling assessment. *In:* Principles of Environmental Sampling, L.H. Keith (Editor). Amercian Chemical Society, pp. 45-72.

Marsily, G. 1986. Quantitative Hydrogeology. Academic Press, Inc., New York, New York.