# Use of Geostatistical Algorithms to Model the Uncertainty in Soil Lead Concentrations at Skeet and Trap Ranges

William Thayer[1,2] , Philip E. Goodrum[2], Gary L. Diamond[2] and James M. Hassett[1]

[1]Syracuse Research Corporation, 6225 Running Ridge Road, North Syracuse, NY 13212,
[2]Faculty of Environmental and Forest Engineering, Bray Hall, SUNY-College of Environmental Science and Forestry, Syracuse, NY 13210.

## Introduction

Remediation decisions at superfund sites require estimates of contaminant concentrations. Skeet and trap ranges demonstrate highly skewed soil lead concentrations, and therefore provide an opportunity to test methods used to map the spatial distribution of contaminant concentrations. We examine the utility of two geostatistically based mapping methods, and comment on their usefulness in assessing the uncertainty of estimates of soil lead concentrations at a skeet and trap range. The first three figures describe the soil lead data at a skeet and trap range, and introduce the issues under investigation.

## Background

Kriging provides a weighted least-squares estimate of the concentration at an unsampled location. The kriging variance can be used to estimate a distribution for the uncertainty in the estimated concentration. The kriging variance is calculated from the variogram (Figure 5) which is estimated from the site data. The kriging variance depends upon the geometry of the data used to estimate the concentration at the unsampled location

The kriging variance does not consider the local variability, i.e., the variability of the data used in the estimation. The kriging variance is assumed to be constant throughout the site. Exploratory data analysis indicates the assumption of constant variance is invalid for this site (Figure 3). Non-constant variance (heteroscedasticity), together with sample clustering in areas with high concentrations (Figure 1) results in the overestimation of the variance at short distances (Goovaerts 1997).

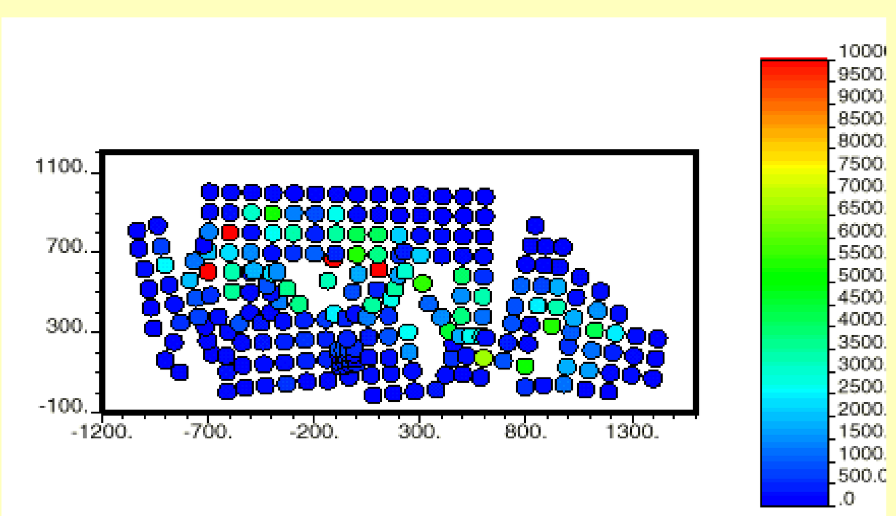


*Figure 1. Schematic site map of a skeet and trap range. The shooting stands are located at 0,0 and the firing direction is in a semicircle towards the top of the page. Sample locations are shown with the concentration of lead indicated by the legend. Note that the lead concentration is less than 500 ppm throughout much of the site, particularly near the limits of the sampling grid. Concentrations greater than 1000 ppm are found primarily near the center of the site. Concentrations greater than 10,000 ppm were measured in the four samples indicated in red.*

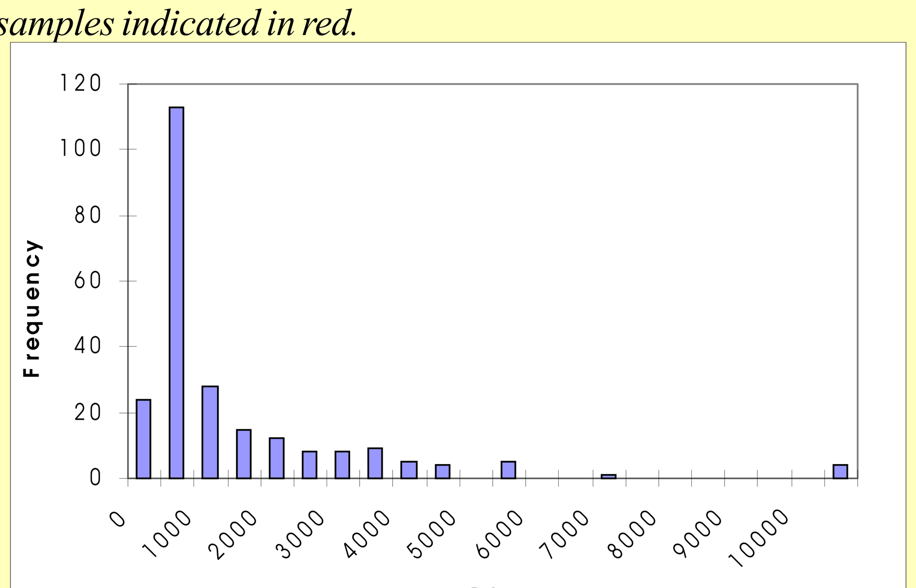



*Figure 2. The histogram of the site data shows the lead concentrations are positively skewed, which is typical of contaminant concentrations measured in soil at hazardous waste sites.*

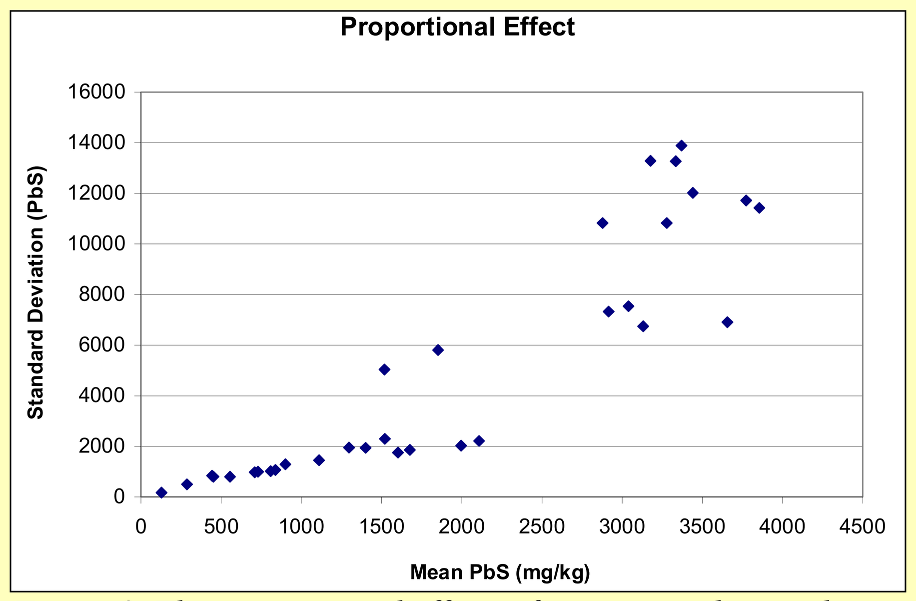



*Figure 3. The proportional effect refers to a condition where the variability of data changes across a site in relation to the change in the average concentration across a site. This figure was prepared by calculating the mean and standard deviation for the data contained in a 'window' that was 250x200 feet; the window was moved across the site resulting in 32 'moving window' estimates of the mean and standard deviation. The figure shows the variance is not constant across the site, a condition that is referred to as 'heteroscedasticity' (Goovaerts 1997).*

## Objectives

There are two objectives for this work:

1. Explore the use of data transformations to stabilize the variance of sample data from a skeet and trap range.
2. Determine the effect of variance stabilization on the distributions of uncertainty for soil concentrations that are estimated by kriging and simulation.

## Methods

Two transformations were used on the soil lead data: the lognormal and the normal score.

The untransformed and transformed data were used with kriging (Figures 7-9) and sequential gaussian simulation (SGS, Figures 10, 11) to produce maps of the probability of exceeding a risk-based threshold concentration of 1078 ppm. The risk threshold is based on the interim Adult Lead Model provided by the EPA.

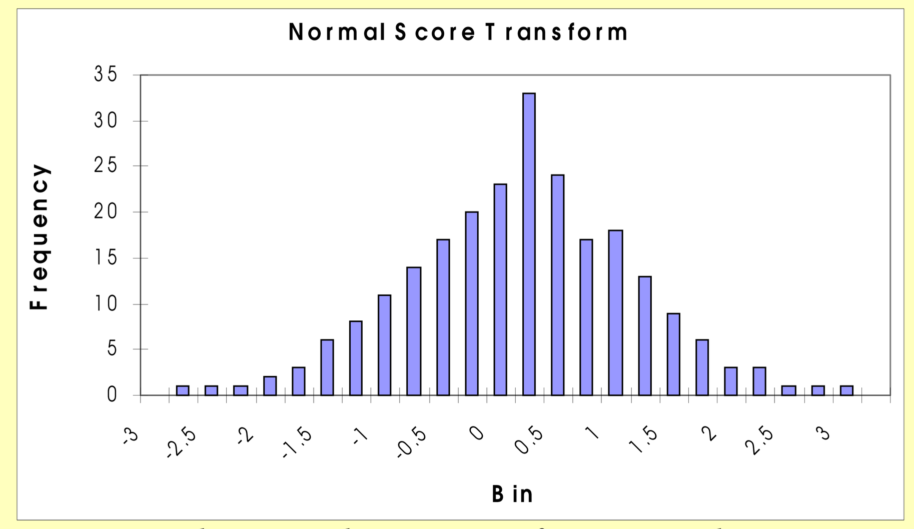



*Figure 4. The normal score transform is an alternative to the logtransform that is often used to increase the normality and stabilize the variance of data. The histogram indicates the normal distribution is a reasonable model for the normal score transform of the data. The logtransform was less successful for improving normality of the data; the normal score is preferred over the logtransform for this data.*

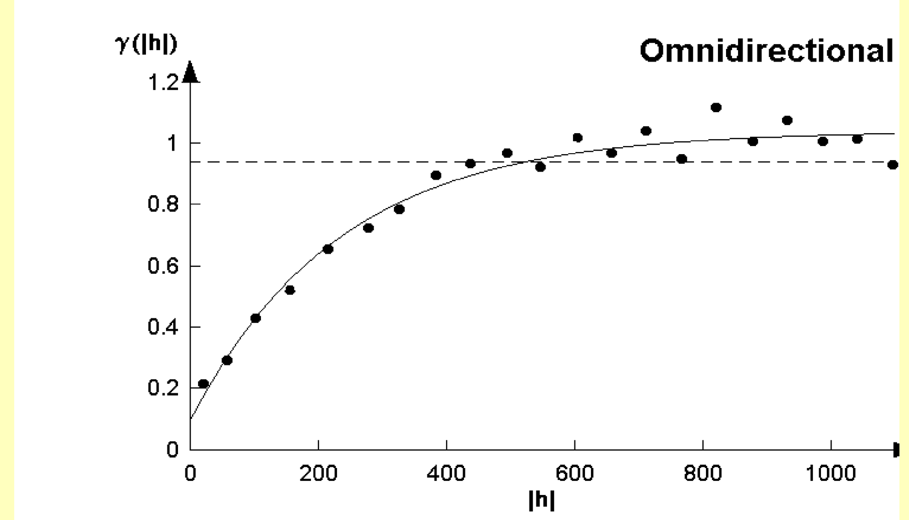



*Figure 5. The semivariogram for the data is shown above. The semivariogram provides a model for the relationship between distance and variance for a site. As shown in the figure, the concentration measured at two points located near to each other will tend to be similar (i.e., less variable); the variance will tend to increase with the distance between two points. The semivariogram is used in kriging to assign weights to the data that are used (in the weighted least squares algorithm) to estimate the concentration for areas that were not sampled.*

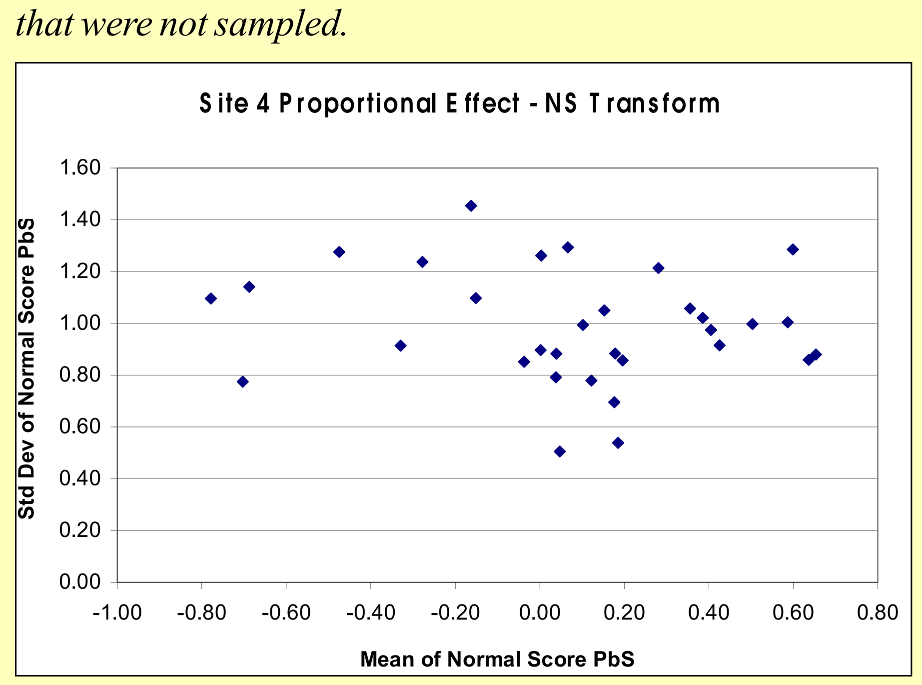



*Figure 6. The normal score transform of the data was successful in stabilizing the variance of the soil lead concentrations, as indicated by the lack of any apparent trend between the mean and the standard deviation.*

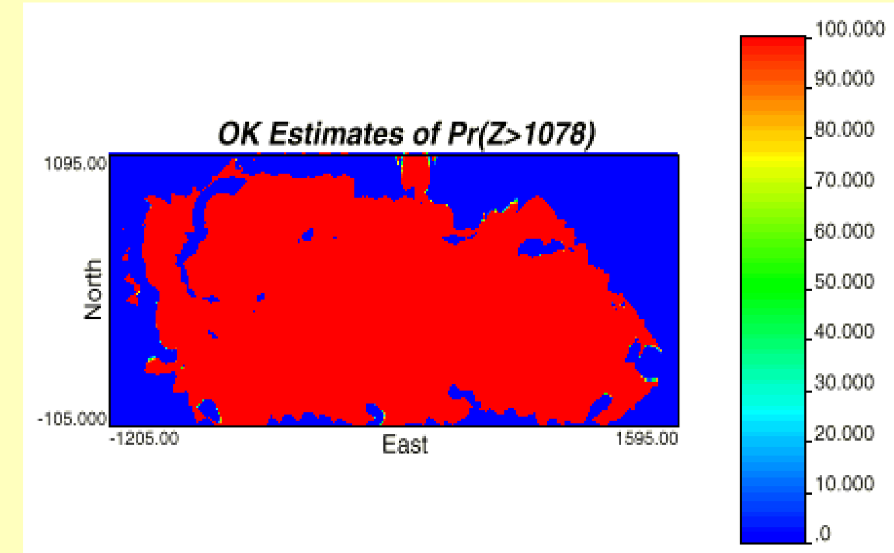
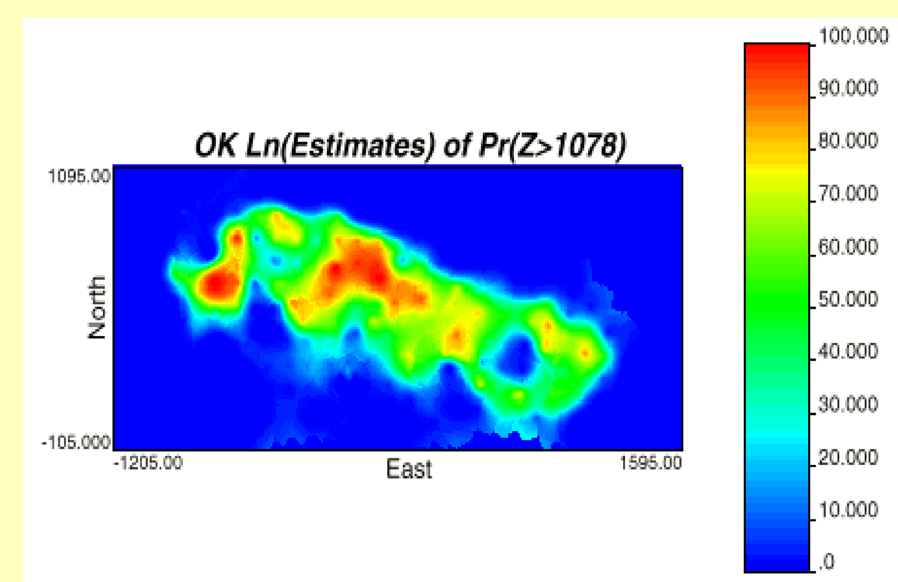



*Figure 7. Using the untransformed data with the ordinary kriging (OK) algorithm produces very high estimates of the uncertainty in the estimated soil lead concentrations. Indeed, this map indicates the entire site exceeds the threshold concentration of 1,078 ppm. This map appears to be unrealistic given the information provided by the site data (Figure 1). For example, measured concentrations in the southeast portion of the site are consistently at or below 500 ppm yet the uncertainty distributions produced by the OK algorithm indicate the concentration in this area of the site is almost certainly (i.e., close to 100% probability) greater than the 1,078 ppm threshold.*




*Figures 8. Using the log transformed data, the OK algorithm indicates that some areas of the site are probably less than the threshold of 1,078 ppm. The area targeted for remediation is a function of the confidence level that one selects for the remediation decision. By convention, the 95% confidence level (i.e., 95% probability that the concentration is less than 1,078 ppm) is often selected, which in this case would result in a remediation of all areas contained within the light blue contour.*

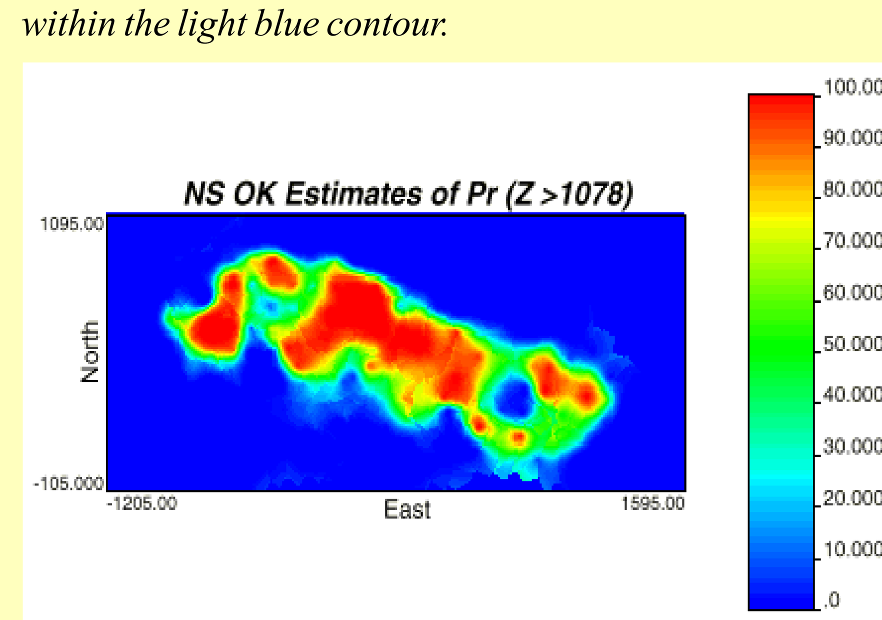



*Figure 9. The normal score transform of the data with the OK algorithm indicates the area that requires remediation (i.e., exceeds 5% probability of being greater than 1,078 ppm) is slightly less than is indicated in Figure 8 (OK with logtransformed data) and much less than is indicated in Figure 7 (OK with untransformed data).*

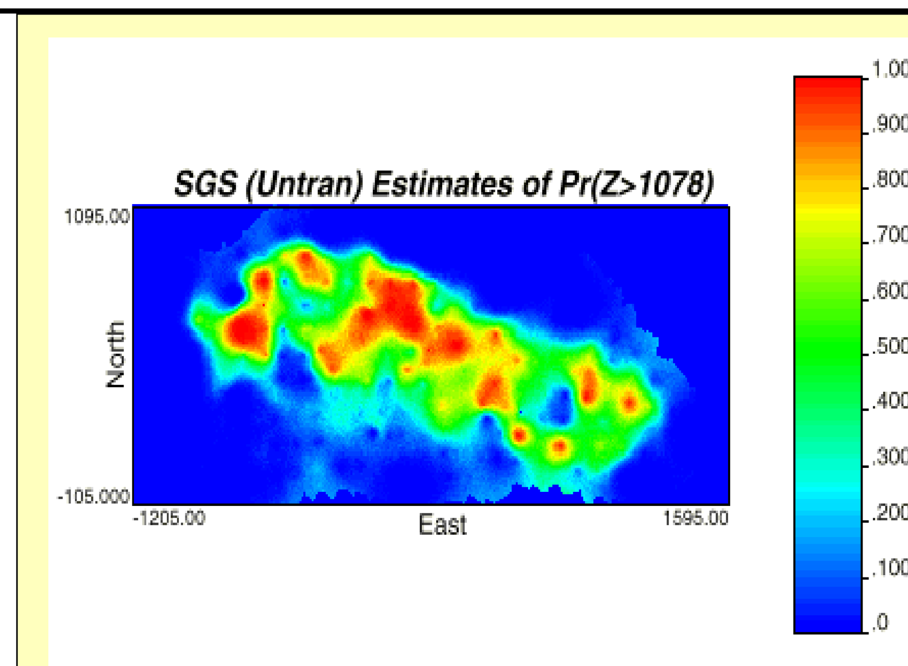



*Figure 10. The Sequential Gaussian Simulation technique was used with the untransformed soil lead data to generate this map. Compare to Figure 7, which was also prepared with untrans-formed data. This map indicates a smaller area to remediate.*

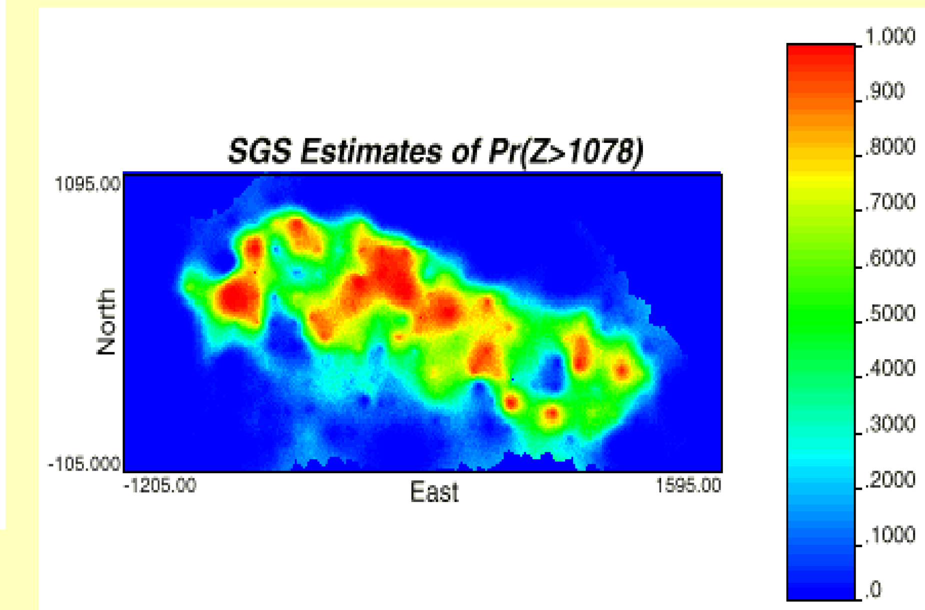



*Figure 11. The Sequential Gaussian Simulation technique was used with the normal score transformed soil lead data to generate this map. Compare to Figure 10, and note that this algorithm does not appear to be sensitive to the heteroscedasticity of the data.*

## Results

The normal scale transform stabilized the variance of the data (Figure 6).

The transformed data appears to produce maps of uncertainty (i.e., probability of exceeding the threshold concentration) that are more reasonable when compared to the measured site data when the ordinary kriging algorithm is used (Figures 1, 8, 9).

The maps prepared from the untransformed and normal score transform using the SGS algorithm differ very little, indicating the SGS algorithm is not sensitive to the non-constant variance of the data.

## Future Research

We hope to compare the performance of simulation and kriging algorithms using randomly generated subsets of large samples. We will also assess the power of post remediation sampling plans using geostatistics. We will integrate Value of Information into sampling design and post remediation verification.

We would ultimately like to develop tools for on-site, real time decision analysis that use geostatistics.

## Reference

Goovaerts, P. *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press, 1997.