APPENDIX R
Geostatistics

R-1. Introduction. Geostatistics is a method for analyzing spatially correlated data. It is used to identify spatial patterns and to interpolate values at unsampled locations. Sampling and mapping in the earth sciences are complicated by spatial and temporal patterns. The structure and intensity of such patterns often cannot be reliably predicted with deterministic models of fate and transport or with classical statistical methods applied to sample observations. Geostatistics is a way of interpreting patterns from sample observations taking advantage of spatial correlation. In geosciences, spatial correlation arises when samples taken close to one another are more likely to have similar values than samples taken far apart Clark (1979).

R-1.1. Appendix O explains that covariance is a statistical measure of the association between two variables. If two variables are independent, the covariance is zero. For geostatistical analysis conducted on a regionalized variable, the auto-covariance between nearby samples is considered to be possibly not equal to zero. If the auto-covariance between two measurements taken close to each other is not zero, then the application of classical statistical methods may impart a substantial bias to the estimate.

R-1.2. Classical statistical methods rely on data being independent over distance or time. Hence, in many environmental problems, the use of classical statistics is not entirely accurate, because variables are frequently spatially controlled. Geostatistics recognizes the spatial correlation and provides methods for the following.

R-1.2.1. Calculating predictions (such as the concentration of a metal at a specific location in soil).

R-1.2.2. Quantifying the accuracy of the predictions.

R-1.2.3. Selecting optimal locations to sample given an opportunity to collect more data.

R-1.3. A geostatistician's main task is to predict a regionalized variable (e.g., hydraulic gradient or metal concentration in soil) from a set of measurements. More detailed treatment of geostatistical methods can be found in Cressie (1993) and Goovaerts (1997).

R-2. Semivariogram. The characteristic tool in geostatistics is the *semivariogram* to quantify and model the spatial correlation structure. A semivariogram is essentially a plot of the variance of groups of paired sample measurements as a function of the distance between samples. Typically, for the situation in which the variance depends only upon distance (and not direction), all possible sample pairs a fixed distance apart (h) are used to calculate a variance for h:

$$s^2(h) = \frac{\sum_{i,j}^{N_h} (x_i - x_j)^2}{N_h}$$

where x_i and x_j represent the value (i.e., concentration) at a pair of sample points i and j; the summation is over all possible pairs of points within a subgroup of the data that are a distance h apart (where $i < j$); and N_h denotes the total number of pairs that are h units part. For example, h is typically defined as increasing along constant intervals ($h = \{1d, 2d, 3d, \ldots\}$, where d is a distance interval such as 5 feet.) In practice, a window of allowable distances is used so that many points will be included in each calculation of $s^2(h)$. For example, a group of samples used to calculate s^2 (3 feet) may have inter-point distances that are between 2 feet to 4 feet apart, rather than exactly 3 feet apart. This window is defined using a tolerance δ for h, so that all points within $h \pm \delta$ of each other are grouped into the subset from which $s^2(h)$ is calculated. The user chooses this tolerance and other grouping parameters to define how the data will be grouped into subsets to calculate the $s^2(h)$ for each h. Different experimental variograms can be calculated for a given data set by varying the grouping parameters used to control the spatial geometry of the data subsets at each distance h.

R-2.1. With grouping parameters defined, computer software is used to do the intensive computations involved in calculating the variance $s^2(h)$ for different values of h. The quantity $\gamma(h) = (1/2)s^2(h)$ is plotted as a function of increasing distance, i.e., $1h$, $2h$, etc., and is referred to as the *experimental* or *empirical semivariogram*. Although the *variogram* is, by definition, twice the semivariogram, the terms variogram and semivariogram are often used interchangeably.

R-2.2. After experimental semivariograms are reviewed, a continuous mathematical curve, called a model semivariogram, is then fit to the experimental semivariogram. Examples of model semivariograms are displayed in Figure R-1. The model semivariogram (Figure R-1a) is assumed to characterize the relationship of how variance in neighborhoods increases as the neighborhoods get larger. This relationship must be estimated for each site application. In practice, 20 or more sample locations are necessary to construct a useful empirical semivariogram, and often geological site knowledge and statistical judgment are important considerations in estimating the model semivariogram.

R-2.3. Figures R-1b and R-1c illustrate two model forms that have a *sill*, or maximum variance. A sill is the upper limit of any semivariogram model that levels off at large distances. In physical terms, the sill is the variance of concentrations at the site that are at a large enough distance from each other to be statistically independent. The distance at which spatial correlation becomes insignificant is called the *range*. Sample points separated by this distance or more are considered statistically independent and can be analyzed using a classic statistical approach. Another feature of a semivariogram illustrated in Figure R-1c is the *nugget*. In a model having a nugget, $\gamma(h)$ does not approach zero as h approaches zero but rather a positive value that is gen-

erally attributed to such things as measurement error for a single observation or small-scale variability.

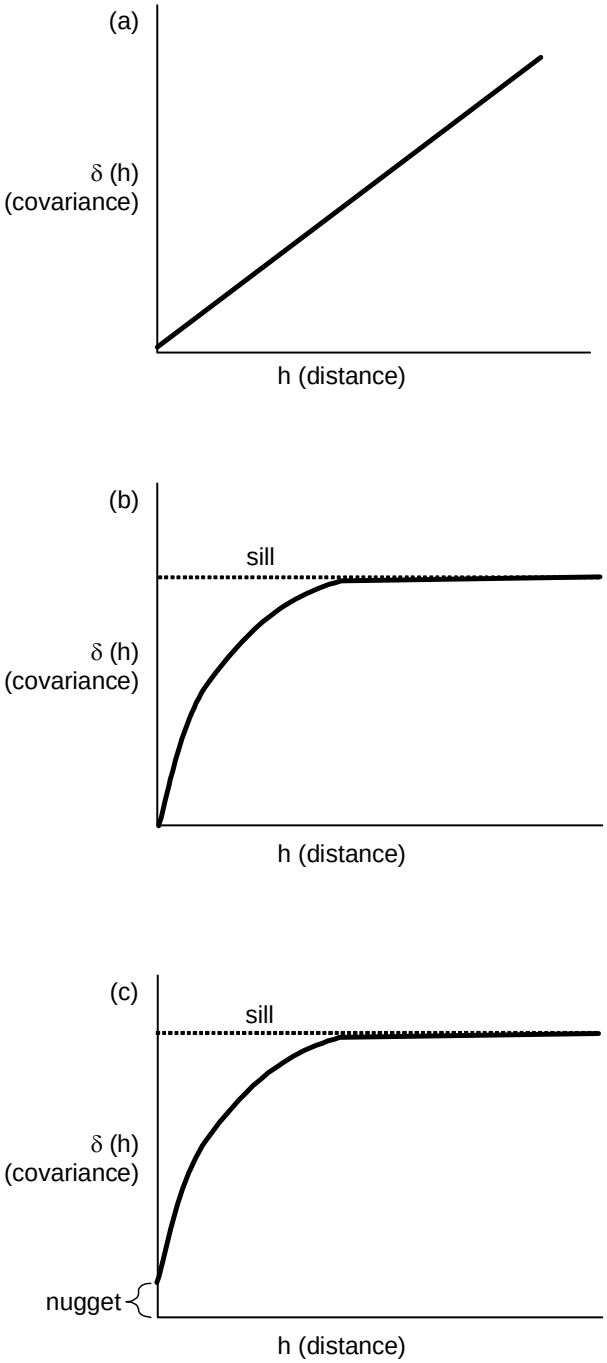


Figure R-1. Types of semivariograms.

R-3. Kriging. The geostatistical interpolation method of *kriging* uses the concepts and model established in a semivariogram data evaluation to develop both an unbiased estimate of the expected value at any specified location, as well as the uncertainty associated with this estimate. Typically, estimates are derived along a regularly spaced grid. With relatively dense grids, the estimate at each grid point is also the estimate of the mean value within the block centered on the grid point.

R-3.1. The kriging estimate is a weighted mean of the neighboring samples, where each weight reflects the amount of unique (non-redundant) information contained about the location to be estimated that is in a given sample. The assignment of weights to each neighboring sample is based on the model semivariogram, and includes consideration of the inter-point distance between the sample and the location to be estimated, as well as the inter-point distances between this sample and its neighboring samples. Neighboring samples, if close together, are spatially correlated and, therefore, contain redundant (non-independent) information about the location to be estimated. Kriged estimates are more accurate than an un-weighted arithmetic mean; that is, they are unbiased (the bias from clustered samples is removed), and they have a lower variance.

R-3.2. Some types of kriging that may be encountered are *ordinary kriging*, *indicator* or *probability kriging*, and *block kriging*. Ordinary kriging is used to predict the value of some variable at a specific location. In block kriging, the technique allows the prediction of a variable mean within a block or area.

R-3.3. The required assumptions for kriging are that the sample to be estimated lie within the neighborhood for which the model semivariogram has been estimated, that there be adequate empirical evidence (sample data) or scientific support (e.g., source history) for the appropriateness of the model semivariogram, and that the neighborhood be homogeneous, with no distinct trends in the data values. For kriging, a trend is a deterministic gradient that can be modeled (such as an exponential decrease in deposition with distance from a point release). Such trends should be characterized and then subtracted from the regionalized variable being modeled. Kriging can then be run on the residuals to account for local patchiness and clustered sample data. Alternatively a release or plume of contamination can often be divided into strata in which the conditions are approximately homogenous (e.g., geological strata, differing source areas). The blocks of each neighborhood are then kriged using their corresponding semivariogram.

R-3.4. Any estimation procedure has an associated estimation variance. The special property of kriging is that it selects the set of weights that minimizes the estimation variance and produces the best linear unbiased estimator.

R-3.5. The assessment of uncertainty in geostatistics is highly quantitative; interpolated concentrations are estimated on the basis of an underlying model of correlation and variability. As such, the estimates themselves are directly linked to estimates of uncertainty. A predicted value may be expressed as a quantity plus or minus some quantity representing the uncertainty

($X \pm \varepsilon$), or the predicted value may be associated with a probability (X, $p = 0.9$). Specific methods of estimating the uncertainty are beyond the scope of this document; they are usually calculated using computer software.

R-3.6. Geostatistics can be used to evaluate and manage the uncertainty associated with remedial activities for a study area. Even with ample site characterization data (borings or wells), the boundaries of the treatment zone are imperfectly defined. Geostatistics allows us to evaluate the risk that the size, and, therefore, cost, of the remediation may be larger or smaller than expected. First, the site is characterized and adequate data are collected. Second, the data are transformed by assigning a value of 1 or 0 (indicator values), depending on whether the value is above or below, respectively, a given cleanup value or other criterion. Third, the transformed data are used to construct a variogram. Fourth, the variogram is modeled as previously described. This model is then used to perform kriging with the indicator values. The kriging estimates reflect a probability that the concentration at the points of estimation exceed the cleanup value or other standard. These kriging estimates can be contoured to define areas or volumes of material that have a certain likelihood of exceeding some cleanup value. The contour value is essentially the probability of exceedance. Last, the size of the area defined by different probabilities of exceedance can be determined and, using a unit cost or similar approach, a cost-versus-risk curve can be developed.

R-3.7. This can be used in programming money for the project, as a basis for negotiating cleanup levels with regulators, or to help determine if the cost and time of additional characterization work will be offset by less risk during construction. Alternatively, rather than transforming the data to ones and zeros, the actual values can be kriged, and the kriging variances can be used to determine prediction intervals for each estimated value. In the vicinity of the point estimate, these prediction intervals can be used to define the spread of potential values expected within a given probability. This assumes the data are normally distributed or have been transformed to be normally distributed.

R-4. Software for Geostatistics. There are a number of software applications to assist in geostatistical calculations. Two older applications developed by the U.S. Environmental Protection Agency (EPA) are GeoPack and Geo-EAS (EPA 600/4-88/033).

R-4.1. GeoPack conducts analysis of variability for one or more random functions. GeoPack includes basic statistics, such as mean, median, variance, standard deviation, skew, and kurtosis. The package also does regressions, distribution testing, and percentile calculations. Sample semivariograms, cross-semivariograms, or semivariograms for combined random functions for a two-dimensional, spatially dependent random function can also be determined. GeoPack includes ordinary kriging and co-kriging estimators in two dimensions, along with their associated estimation variance and the conditional probability that the value is greater than a user-specified cutoff level. Graphical tools include linear or logarithmic line plots, contour plots, and block (pixel) diagrams.

R-4.2. Geo-EAS was also developed by the EPA and is a collection of interactive software tools for doing two-dimensional geostatistical analyses of spatially distributed data. Programs are provided for data management, data transformations, univariate statistics, semivariogram analysis, cross-validation, kriging, contour mapping, post plots, and line-and-scatter graphs. The application is DOS-based.

R-4.3. A publicly available package of geostatistical software that is more comprehensive than these EPA packages is GSLIB, available at http://www.gslib.com. The DOS-executable freeware may be downloaded from this site. Alternatively, the software source code and a supporting textbook may also be purchased at the site for a nominal fee.

R-4.5. Commercial software for Windows, Sun, or Macintosh systems include WinGSLIB, Environmental Visualization System, and Groundwater Modeling System (GMS), which is currently available to all USACE, U.S. Department of Defense, EPA, and U.S. Department of Energy personnel.

R-5. Case Study: Geostatistical Analysis of Remediation by In Situ Ozonation.

R-5.1. *Introduction*. An application of geostatistics to environmental remediation will be explored in this case study. Three-dimensional kriging was used to support the Remedial Investigation/Feasibility Study, Remedial Action Plan, Confirmation Sampling, and Remedial Action Report (site closure) for a former manufactured gas plant (MGP) located in Long Beach, California. The former MGP operated from approximately 1901 to 1913 and produced gas from coal and crude oil feedstocks. The project was conducted pursuant to an agreement with the California Environmental Protection Agency Department of Toxic Substances Control under their Expedited Remedial Action Program. In-situ ozonation was used to lower levels of polycyclic aromatic hydrocarbons (PAHs) to meet the selected risk-based cleanup levels for this site. The kriging results played an important role in several estimation and decision processes, including:

R-5.1.1. Contouring the original distribution of PAH.

R-5.1.2. Defining the footprint and depths for the treatment zone.

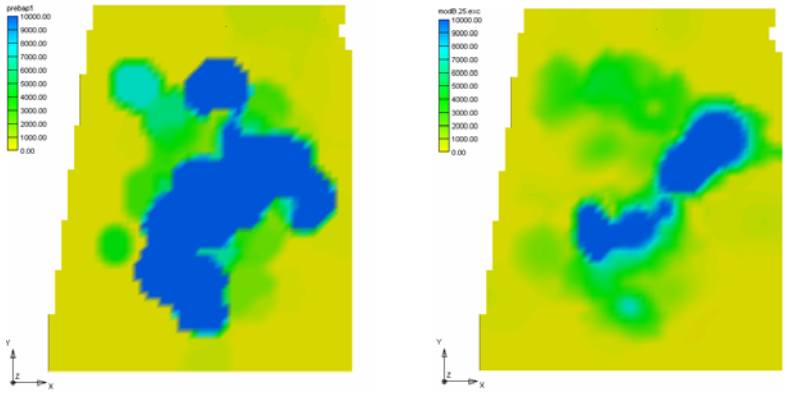
R-5.1.3. Supporting decisions regarding placement for the ozone-injection well system.

R-5.1.4. Selecting quarterly monitoring locations for soil samples during the treatment process as well as for post-treatment confirmation samples.

R-5.1.5. Contouring the final post-treatment distribution.

R-5.1.6. Estimating the site-wide exposure concentration used for risk assessment and site closure.

R-5.2. *Post-Treatment Modeling*. Quarterly monitoring results indicated substantial reductions in PAH levels early in the treatment, which began in 1998. However, within 2 years, monitoring indicated that the reductions had reached an asymptote, reflecting the diminishing return of continued ozonation and the recalcitrant nature of the residual PAHs. In 2000, confirmation samples were taken from random locations within the defined treatment zone. Kriging was then used to model the post-treatment spatial distribution of PAHs and compare it to the pre-treatment distribution (Figure R-2). Kriging uncertainty was estimated and used to determine whether cleanup goals had been met.



a. Pre-ozonation. **b. Post-ozonation.**

Figure R-2. Comparison of krige-interpolated benzo-(*a*)pyrene concentrations before and after treatment by in situ ozonation.

R-5.3. *Reporting*. The reporting of the kriging analysis was included in the Remedial Action Report as an appendix with an organization and level of detail consistent with guidelines given in *Standard Guide for the Contents of Geostatistical Site Investigation Report* (ASTM D5549-94e1). To enhance the practical value of this case study, the following parts of the ASTM outline are used below: software, data sources, exploratory analysis (and conceptualization), spatial continuity analysis, estimation, and uncertainty.

R-5.4. *Software*. The analysis was conducted using the three-dimensional kriging utilities of the GMS software mentioned in Paragraph R-4.

R-5.5. *Data Sources*. The variable of interest was *benzo(*a*)pyrene equivalents* (a weighted sum of carcinogenic PAHs in each sample). The first kriging analysis (pre-remediation) was conducted on sample data dispersed over approximately 75 soil borings primarily from the remedial investigation (RI) program that was completed in 1997. In contrast, the post-remediation

kriging was conducted on a composite data set that consisted of 1997 RI samples that were outside the treatment area, together with the latest samples available for the treatment area taken in 2002. Thus, the post-remediation data set reflected the assumption that soil concentrations in the untreated areas were stable over time (reasonable for the PAHs involved) and, therefore, well represented by older data, while soil concentrations in the treatment zone were expected to change over time so that older samples were not included in the kriging analysis.

R-5.6. *Exploratory Analysis and Site Conceptualization*. The recommended Exploratory Analysis section in the ASTM guidelines is expanded here to be a conceptual discussion, deemed important for all sites, that considers all relevant qualitative and quantitative information about the site. The integration of these different types of information is crucial for explicitly identifying a conceptual model of the contamination distribution that will guide a number of assumptions and decisions throughout the analysis. Beyond the analytical sample results, such information includes topography, stratigraphy, observations made in boring logs, site history, and other qualitative and semi-quantitative information. For the MGP site, all examples from the above list were applied in some way during formulation of the geostatistical analysis. The following description of some of the qualitative information about the site is included before the transition into the exploratory data analysis.

R-5.6.1. Well-established site history provided engineering process information, as well as maps of potential source structures, that could be used to compare with the posted analytical results. An additional factor at the site is that its current condition includes an engineered soil levee along the Los Angeles River as well as soil fill set around large concrete supports for a bridge and on-ramp built across the site in 1953–1963 (subsequent to decommissioning of the MGP). Thus, the topography is quite varied and includes imported soil brought in to cover large parts of the site. Topography, native or fill, definitely influenced soil volumes and, therefore, had to be incorporated explicitly into the kriging estimation. Furthermore, the three-dimensional visualization of the topography and sample data (Figure R-3) indicated that spatial correlation occurred along a relatively level elevation rather than following the highs and lows of the present surface topography. (An approximate two-fold vertical exaggeration is used to aid the visualization of data points within a boring.) This is consistent with the expected pattern produced by an originally flat plant site. Because the subsequent mixing and earth movement are somewhat uncertain, the large volume of soil covering the former plant was sampled, along with the native soil, as part of the RI and was included in the site-wide model and calculations.

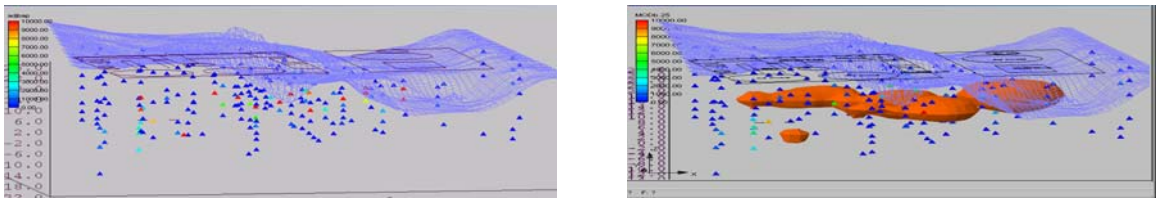


Figure R-3. Surface topography with benzo(*a*)pyrene concentrations and a kriged isovolume.

R-5.6.2. The additional exploration of the analytical data, in the form of a histogram and descriptive statistics, indicated a high degree of skewness with suggestions of a composite of two different populations: one with low concentrations (i.e., "background") that were found in the outlying areas, and the other with moderate-to-relatively high concentrations that still presumably reflected some varying amount of impact from the historical contamination (even after treatment). Samples in the outlying areas were more sparse than in the central area, but still provided ample evidence to confirm the central positioning of the impacted soil in and around former MGP structures. Therefore, this potentially distinct population of low values was considered important to keep in the data set so that it would help define the outward extent of the residual contamination.

R-5.6.3. Including the low concentrations together with the more central data had the following implications.

R-5.6.3.1. *Site Mean*. The site-wide estimate of exposure concentration would reflect a site mean that included, in accordance with the defined site boundaries, both background and impacted volumes of soil. The site-wide mean to be calculated based on the kriging analysis would have contributions from both parts of the site in a manner that was "volume-weighted." Given that any future redevelopment of the site would require the removal of the bridge support structures and intensive mixing of soil across the entire site, this site-wide mean was considered a realistic assumption for the conservative residential risk scenario.

R-5.6.3.2. *Spatial Pattern or Lateral Extent*. The lower concentrations confirm site historical information regarding the "edges" of the impacted zone. Given this confirmation, the sparse outlying data can and should be supplemented with "soft data" to fill in areas of low data density and create a well-controlled boundary condition for the edges of the site. Such soft data, termed the "extended data set," were added to the kriging for the estimation phase conducted after development of the variogram.

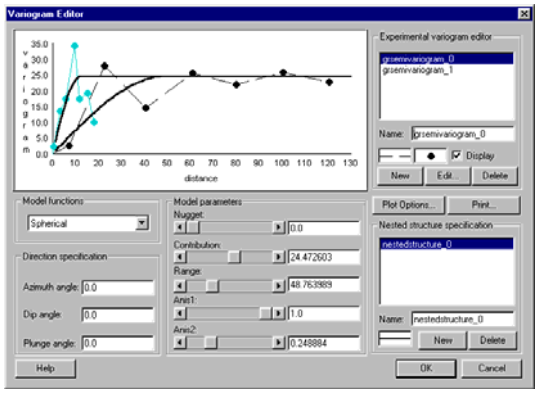
R-5.7. *Spatial Continuity Analysis*. It is reasonable to estimate soil concentrations across the site based on the underlying kriging assumption of spatial continuity. The fate and transport processes, involved in both the contamination and the ozone dispersion and effect, are presumably spatially continuous on some scale. Although soil structure and sample concentrations are

notoriously variable or even discrete on a small scale, the resolution requirements implied by both risk assessment and remediation allow a broader focus. The view from risk assessment is one of exposure accumulated over time and space (a spatial average), and the view from remediation might be described as akin to the scoop size of a backhoe or some other scale useful for feasibility and cost estimation of the specific treatment. This larger scale of variability is more forgiving in the sense that an interpolation of the mean concentration in a cubic-yard block of soil has a lower uncertainty than an interpolation of any particular shovel-full of soil in the same block.

R-5.7.1. Therefore, the search for evidence and range of spatial continuity need not be a matter of finely tuned research for many sites although the level of rigor must be consistent with the site conceptualization. For example, one might argue that the outlying data, if they are truly background, may be a different population altogether than the central data, with different continuity ranges to be found by analyzing the two sets separately. On the other hand, there is no bright line around the site to delineate these two populations spatially (at least "a priori," before the spatial analysis was done). More realistically, there is likely to be a gradient of soil impacted by some level of contamination and also some level of remediation, such that the net impact, or probability of impact, on the soil decreases with distance from the central area and individual injection wells. The spatial range of correlation that is defined for the kriging variogram should ideally be appropriate for this transition zone, as well as for the obvious central or outlying areas of the site. In other words, practicality points to the simplest assumptions that will "work."

R-5.7.2. Spatial continuity was investigated on the entire post-remediation data set using a *general relative* variogram, which automatically adjusts for the *proportional effect* commonly found in contaminant concentration data and lognormal tending data in general. The variances calculated for a relative variogram were modified by dividing the group variances by the square of the local mean, which can be calculated in several ways. This improved the structure of the experimental variogram and, specifically for the case study data, allowed the modeler to observe lower relative variances (stronger correlations) at inter-point distances of about 5 to 10 feet (in the laterally direction), moderate variances at about 30 to 40 feet, and highest variances reaching a plateau at about 50 to 60 feet (see Figure R-4a).

a. Directional horizontal variograms.



b. Horizontal and vertical variograms.

Figure R-4. Two sets of experimental variograms used to define the krige variogram model.

R-5.7.3. Horizontal anisotropy was reviewed by limiting vertical and angular grouping parameters (depends on software package) to create different directional "horizontal variograms." No horizontal anisotropy was present. However, the comparison between the horizontal variograms and vertical variograms, created by limiting horizontal grouping parameters, indicated that the vertical range of spatial correlation was approximately one-fourth that of the horizontal range. A spherical model variogram with vertical anisotropy was selected, with a horizontal range of 49 feet and a vertical range of 12.4 feet (Figure R-4b).

R-5.8. *Estimation*. A concentration estimate was developed for every cell center of a three-dimensional grid with cell dimensions 6 by 6 by 3 feet. Cell dimensions were chosen to be consistent with earlier modeling work, but also were considered to provide an adequate balance between the resolution needs of risk and remediation resolution and the increased run-time and overall unwieldiness of denser grids. The three-dimensional contour map could be compared to the pre-remediation maps in plan view by layers, or by cross sections or rendered iso-volumes in GMS. The site-wide mean was then simply a matter of calculating the arithmetic average of all cell mid-points that were defined as "soil" (as opposed to "above ground").

R-5.9. *Uncertainty*. Kriging standard error estimates are automatically produced for each cell at the time the concentration estimate is assigned. They reflect uncertainty in a particular cell estimate and cannot be used directly to estimate uncertainty for the site-wide mean, which is the standard error term required for a 95% upper confidence limit (95% UCL), i.e., the exposure concentration. The standard error for the site-wide mean was conservatively estimated by using the kriging error resulting when the variogram model was run on a new grid consisting of one large three-dimensional cell encompassing the entire site. The intuitive definition of this error term is that it represents the uncertainty implied by using the available 233 *spatially correlated*

sample points to estimate the mean concentration of the entire block of soil containing the 233 correlated samples. As the site boundaries and especially the topography result in an irregularly shaped zone of soil within this large rectangular block, the "block type" of uncertainty results in an overestimate for the actual soil subzone within the block. This is because the block uncertainty reflects large regions of "air" that are not properly distinguished from soil, and these regions have no sample data and are relatively far from the nearest sample datum. This method was conservative but was considered reasonable for use in the risk assessment. Detailed discussions of the many uncertainty approaches for kriging can be found in Meyers (1997), which focuses on environmental contamination, and other geostatistical texts (e.g., Goovaerts, 1997). Thus, the 95% UCL on the mean was calculated as

$$95\%\ UCL = UE + 1.645\ KSE$$

where

UE = unbiased estimate of the mean (obtained from the high resolution model)
KSE = kriging standard error (conservative estimate)
1.645 = the 95^{th} percentile of the standard normal distribution

R-6. Conclusion.

R-6.1. Although several conservative analysis assumptions were built into the model and uncertainty formulation, the site-wide volume-weighted exposure concentration (95% UCL) was reduced by 37 to 58% compared to that calculated from the most commonly used non-spatial formulas identified in numerous risk assessment guidances (e.g., *t*-based, Land, bootstrap). The reduction in the exposure concentration came from the more rigorous use of spatial correlation and soil volume when kriging rather than the classical assumption that all sample points were identically distributed, i.e., without spatial correlation. Thus, the lower kriged exposure concentration was important in determining the attainment of risk-based cleanup goals.

R-6.2. The kriged model contours of the post-treatment spatial distribution allowed the visual comparison of the estimated pre- and post-remediation distributions, and were instrumental in concluding the effectiveness of in situ ozonation for this site.