Environ. Sci. Technol. **2000**, *34,* 4228−4235

Geostatistical Interpolation of Positively Skewed and Censored Data in a Dioxin-Contaminated Site

HIROTAKA SAITO AND
PIERRE GOOVAERTS*

Environmental and Water Resources Engineering, Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, Michigan 48109-2125

A correct delineation of hazardous areas in a contaminated site relies first on accurate predictions of pollutant concentrations, a task usually complicated by the presence of censored data (observations below the detection limit) and highly positively skewed histograms. This paper compares the prediction performances of four geostatistical algorithms (ordinary kriging, log-normal kriging, multi-Gaussian kriging, and indicator kriging) through the cross validation of a set of 600 dioxin concentrations. Despite its theoretical limitations, log-normal kriging consistently yields the best results (smallest prediction errors, least false positives, and lowest total costs). The cross validation has been repeated 100 times for a series of sampling intensities, which reduces the risk that these results simply reflect sampling fluctuations. Indicator kriging (IK), in the simplified implementation of median IK, produces good predictions except for a moderate bias caused by the underestimation of high dioxin concentrations. Ordinary kriging is the most affected by data sparsity, leading to a large proportion of remediation units wrongly declared contaminated when less than 100 observations were used. Last, decisions based on multi-Gaussian kriging estimates are the most costly and create a large proportion of false positives that cannot be reduced by the collection of additional samples.

Introduction

The quality of a site characterization and remediation plan relies on the collection of quality and representative data and their processing by technically sound statistical procedures. Geostatistics provides a set of statistical tools for incorporating the spatial and temporal coordinates of observations in data processing, and its increasing use in environmental applications bears its utility and success (*1*−*3*). In particular, a least-squares interpolation technique, known as kriging, is now commonly used for mapping pollutant concentrations in contaminated sites. However, the practitioner often gets confused in the face of the palette of kriging methods available. The choice of a kriging algorithm should be primarily guided by the characteristics of the data under study. Two common features of pollution data are (i) highly positively skewed histograms indicating the presence of a few very large concentrations (hot-spots) and (ii) presence of data below the detection limit (censored observations).

* Corresponding author phone: (734)936-0141; fax: (734)763-2275; e-mail: goovaert@engin.umich.edu.

Skewed data are encountered in many fields, such as in mining where geostatistics originated 40 years ago. To cope with the presence of a few high grades, mining geostatisticians developed log-normal kriging, which is kriging applied to log-normal transforms of data followed by a back-transform of final estimates (*4*). Log-normal kriging must however be used with caution because it is nonrobust against departures from the log-normal model and the back-transform is very sensitive to semivariogram fluctuations. Unlike the log-normal transform, normal-score transform allows one to symmetrize the distribution of data regardless of the shape of the sample histogram (*2*, *5*). The kriging of such normal scores is referred to as multi-Gaussian kriging. Although log-normal and normal-score transforms attenuate the impact of high values, they are not the most appropriate for censored data, which are better viewed as a different statistical population. Indicator kriging (*6*, *7*) offers a way to deal with classes or populations of both high values and values below the detection limit. The basic idea is to discretize the range of variation of the environmental attribute by a set of thresholds (e.g. deciles of sample histogram, detection limit, regulatory threshold) and to transform each observation into a vector of indicators of exceedence or not of each threshold. Kriging is then applied to the set of indicators and estimated values are assembled to form a conditional distribution of probability, the mean of which is used as an estimate of the pollutant concentration.

To our knowledge, the prediction performances of these four different kriging algorithms have never been compared for data displaying both a positively skewed distribution and censored observations. Such a performance comparison is the main objective of this paper, and it will be based on a large soil data set including 600 dioxin concentrations measured in a 1858 m^2 site. The impact of sampling intensity on prediction errors will be investigated by drawing random subsets of increasing size and using them to predict concentrations at the remaining locations (jackknife approach). In a recent study (*8*), a jackknife of the same kriging algorithms has been used for predicting heavy metal concentrations, and sampling fluctuation was found to be the dominant factor determining failure or success of all the methods. In this study, 100 random subsets will be considered for each sampling intensity, which will allow one to eliminate sampling interferences. The use of kriging estimates for classifying remediation units as safe or contaminated will also be investigated, and the economic cost of false positives (unnecessary remediation) will be compared to the cost of collecting additional observations, leading to an optimum (most cost-effective) number of observations for each algorithm.

Theory

Consider the problem of estimating the value of an attribute z (e.g., dioxin concentration) at an unsampled location $\mathbf{u}$, where $\mathbf{u}$ is a vector of spatial coordinates. The information available consists of values of the variable z at n locations $\mathbf{u}_\alpha$, $z(\mathbf{u}_\alpha)$, $\alpha = 1, 2, ..., n$. A proportion of these values is below the detection limit z_d, and such censored data are here reset to z_d in agreement with previous analysis of the data (*9*). This section reviews four geostatistical algorithms that build spatial estimates as linear combinations of surrounding z data or their transforms, and implementation issues are discussed.

10.1021/es991450y CCC: $19.00
Published on Web 08/26/2000

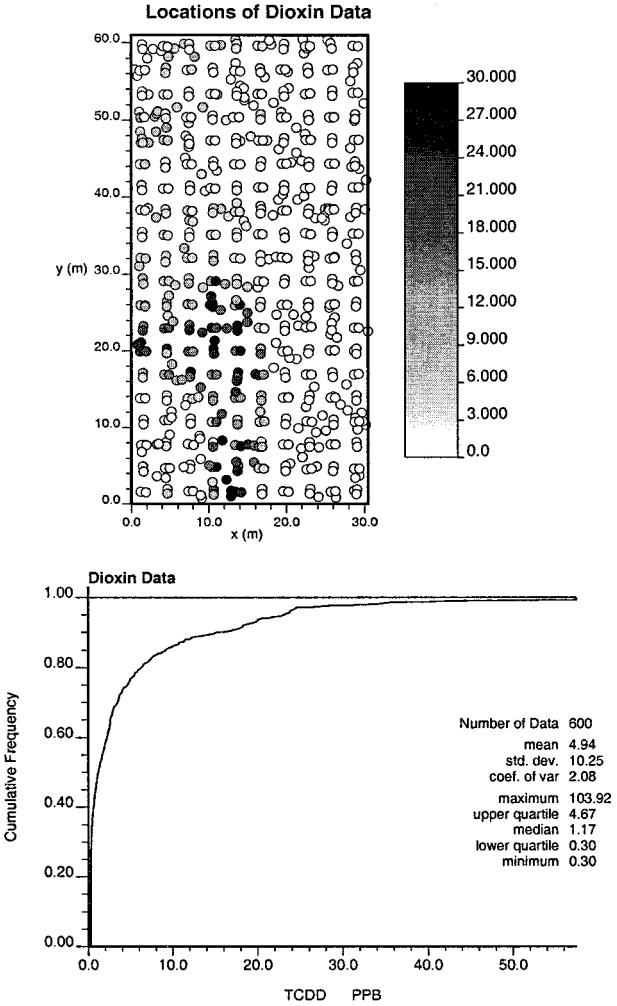

FIGURE 1. Location map of dioxin data at Piazza Road, MO, and cumulative distribution of dioxin concentration (μg/kg).

Ordinary Kriging (OK). The most common estimator is OK, which computes the unsampled value $z(\mathbf{u})$ as a linear combination of neighboring observations:

$$z^*_{\mathrm{OK}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\, z(\mathbf{u}_\alpha) \tag{1}$$

Typically only the observations closest to $\mathbf{u}$ being estimated are retained because they screen the influence of those farther away, that is $n(\mathbf{u}) \ll n$. In this paper, $n(\mathbf{u}) = 16$ for all locations and algorithms, and this choice is further discussed in the section on indicator kriging. The key issue is the determination of the weight $\lambda_\alpha(\mathbf{u})$ assigned to each observation.

OK weights are chosen so as to minimize the estimation or error variance $\sigma^2_{\mathrm{E}}(\mathbf{u}) = \mathrm{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$ under the constraint of unbiasedness of the estimator (eq 1). These weights are obtained by solving a system of linear equations, which is known as the "ordinary kriging system":

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u})\, \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu(\mathbf{u}) = \gamma(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1,\ldots,n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) = 1 \end{cases} \tag{2}$$

Unbiasedness of the estimator is ensured by constraining the weights to sum to one, which requires the definition of the Lagrange parameter $\mu(\mathbf{u})$. The only information required by the kriging system are semivariogram values for different

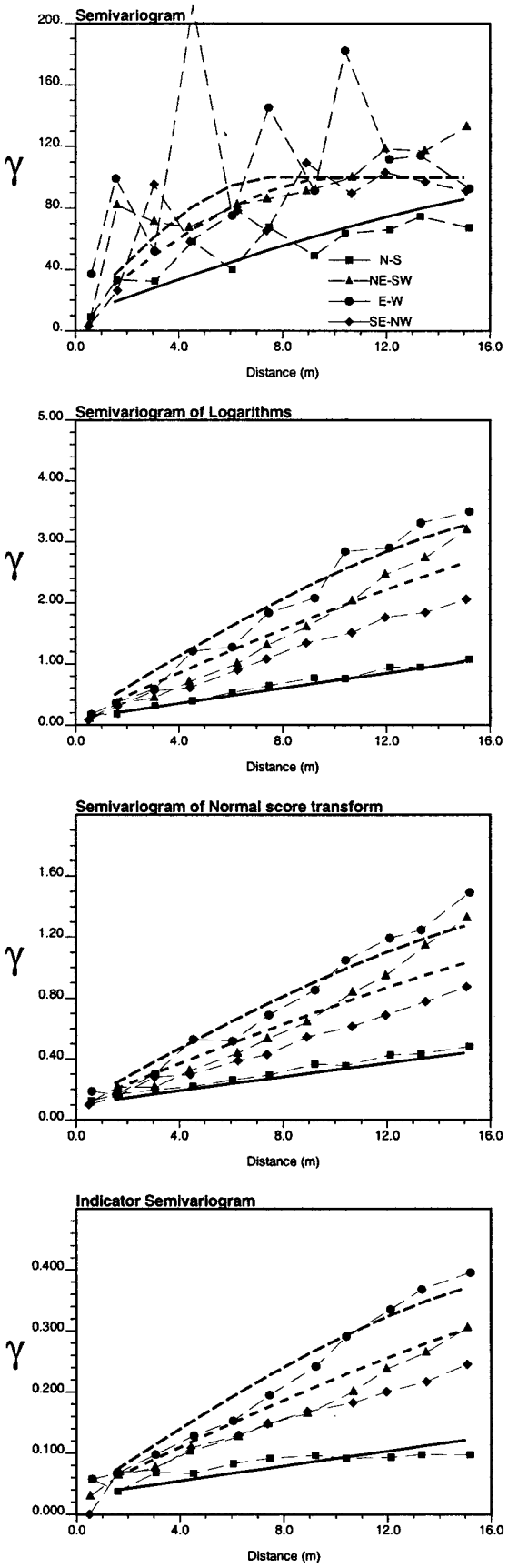

FIGURE 2. Experimental directional semivariograms of dioxin data and of various transforms with the geometric anisotropy models fitted.

lags, and these are readily derived from the semivariogram model fitted to experimental values:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2 \quad (3)$$

where $N(\mathbf{h})$ is the number of data pairs for a given distance.

In addition to an estimate for the unknown z value, OK provides an error variance that is computed as

$$\sigma^2_{\mathrm{OK}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\, \gamma(\mathbf{u}_\alpha - \mathbf{u}) - \mu(\mathbf{u}) \quad (4)$$

Log-Normal Kriging (LK). Although OK does not require the data to follow a normal distribution, prediction performances are usually better if the sample histogram does not display a strong skewness. Log-normal transform is a common way to symmetrize the distribution provided there is no zero data, and the kriging estimate is now computed as a linear combination of log-normal transforms of the observations:

$$y^*_{\mathrm{LK}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\, y(\mathbf{u}_\alpha) \quad \text{with } y(\mathbf{u}_\alpha) = \ln [z(\mathbf{u}_\alpha)] \quad (5)$$

The kriging weights and kriging variance are computed using expressions similar to eqs 1 and 4 except that the semivariogram is computed and modeled for the log-normal transforms of the data.

The delicate step in LK is the back-transform of the estimated value $y^*(\mathbf{u})$ to retrieve the estimate of the original variable z at $\mathbf{u}$. The unbiased back-transform of the ordinary LK estimate (*10*) is

$$z^*_{\mathrm{LK}}(\mathbf{u}) = \exp[y^*_{\mathrm{LK}}(\mathbf{u}) + \sigma^2_{\mathrm{LK}}(\mathbf{u})/2 + \mu(\mathbf{u})] \quad (6)$$

Because of the exponentiation, final results are very sensitive to the values of the kriging variance and Lagrangian parameter, which both rely on the somewhat arbitrary modeling of the semivariogram. Note that in the presence of a large proportion of censored data the log-normal assumption becomes inappropriate since the transform does not eliminate the high-frequency class of low values displayed by the sample histogram.

Multi-Gaussian Kriging (MG). Normal-score transform is a graphical transform that allows one to normalize any distribution, regardless of its shape (*2*, *5*). It can be seen as a correspondence table between equal p quantiles z_p and y_p of the z cdf $F(z)$ (cumulative histogram) and the standard Gaussian cdf $G(y)$. In practice, the normal-score transform proceeds in three steps:

(i) The n original data $z(\mathbf{u}_\alpha)$ are first ranked in ascending order:

$$[z(\mathbf{u}_{\alpha'})]^{(1)} \le \ldots \le [z(\mathbf{u}_\alpha)]^{(k)} \le \ldots \le [z(\mathbf{u}_{\alpha''})]^{(n)} \quad (7)$$

where the superscript k is the rank of datum $z(\mathbf{u}_\alpha)$ among all n data. Since the normal-score transform must be monotonic, ties in z values must be broken, which may be a problem in the presence of a large proportion of censored data. In this paper, such untying or despiking has been done randomly as implemented in GSLIB software (*5*). An alternative is to despike according to the local averages of the data surrounding each tied value (*11*): tied values in high-valued areas would then get larger ranks than those in low-valued areas.

(ii) The sample cumulative frequency of the datum $z(\mathbf{u}_\alpha)$ with rank k is then computed as $p^*_k = k/n - 0.5/n$.

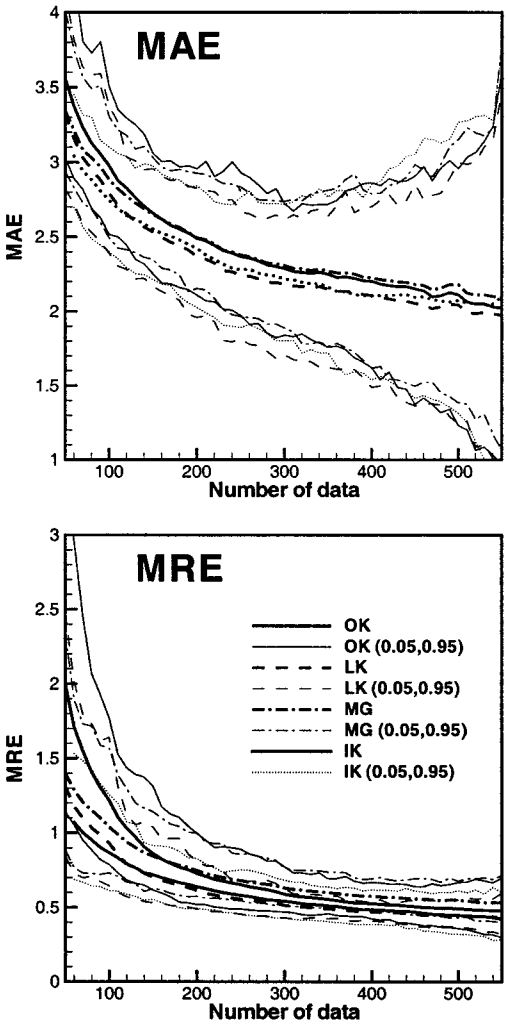

FIGURE 3. Impact of the sample size on the mean absolute error (MAE) and the mean relative error (MRE) of prediction produced by different kriging algorithms. The 5th and 95th percentiles of the error distribution are also plotted.

(iii) The normal-score transform of the z datum with rank k is matched to the p^*_k quantile of the standard normal cdf:

$$y(\mathbf{u}_\alpha) = G^{-1}[F(z(\mathbf{u}_\alpha))] = G^{-1}(p^*_k) \quad (8)$$

Once the normal-score transform has been performed, the normal score at $\mathbf{u}$, $y(\mathbf{u})$, is estimated as a linear combination of surrounding normal-score data (recall eq 1). The kriging weights and kriging variance are computed using expressions similar to eqs 1 and 4 except that the semivariogram (eq 3) is computed and modeled for the normal-score transforms of the data.

Like LK, a delicate operation is the back-transform of the normal-score estimate $y^*_{\mathrm{OK}}(\mathbf{u})$: simply applying the inverse of the normal-score transform, eq 8, to the estimated y value leads to a biased estimation of z (underestimation for the present case study). In this paper, we propose the following procedure:

(i) Model the probability distribution of Y at location $\mathbf{u}$ as Gaussian with for mean and variance the OK estimate $y^*_{\mathrm{OK}}(\mathbf{u})$ and SK variance $\sigma^2_{\mathrm{SK}}(\mathbf{u})$ computed from the normal-score data.

(ii) Discretize that distribution using 100 quantiles $y_p(\mathbf{u})$ corresponding to probability $p = k/100 - 0.5/100$ with $k = 1, 2, \ldots, 100$.

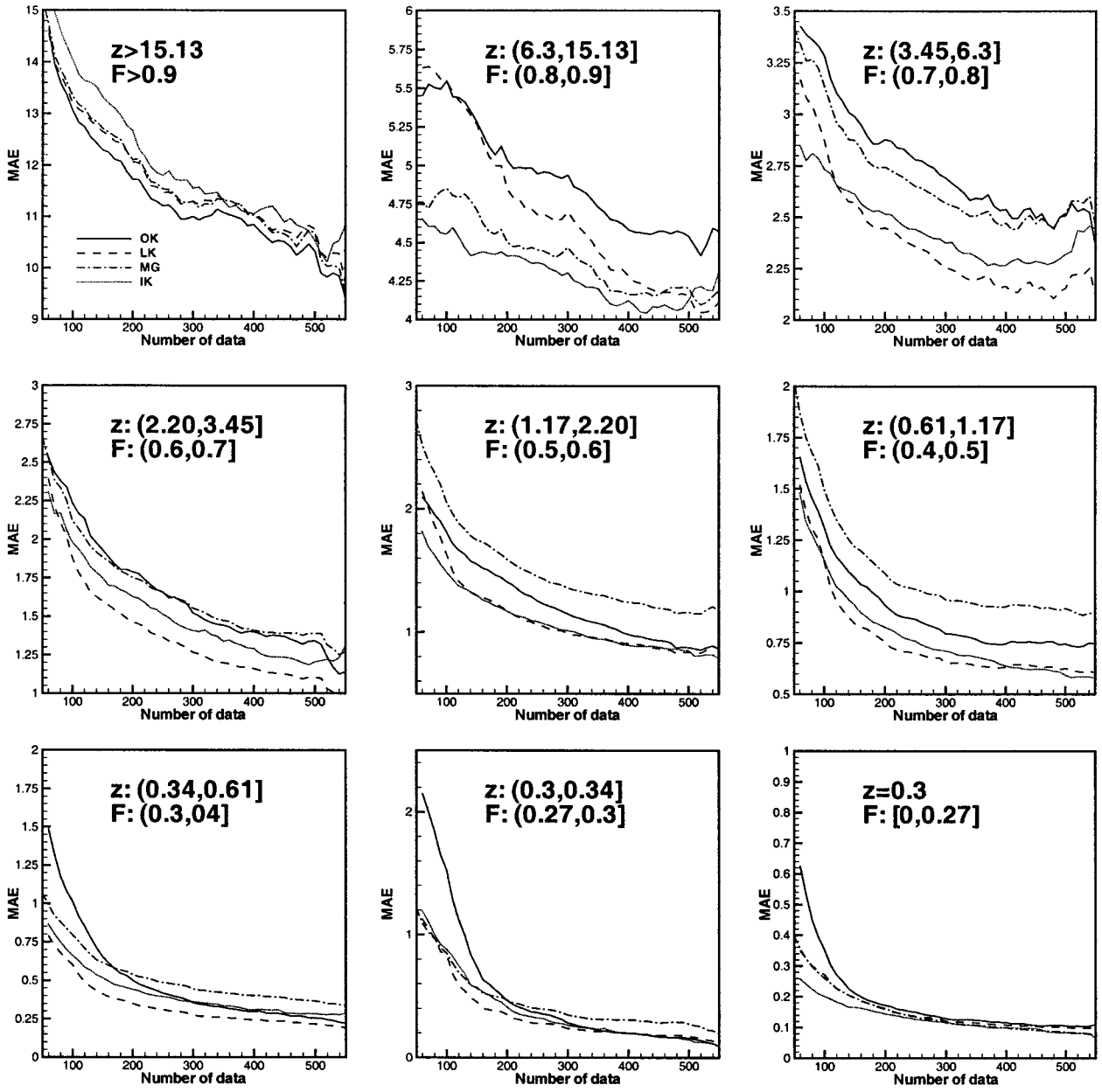

FIGURE 4. Impact of the sample size and class of dioxin concentrations on the mean absolute error (MAE) of prediction.

(iii) Compute the corresponding quantiles of the local probability distribution (ccdf) of Z, $F(\mathbf{u}; z|(n)) = \text{Prob}\{Z(\mathbf{u}) \leq z|(n)\}$, using a back-transform of $y_p(\mathbf{u})$:

$$z_p(\mathbf{u}) = F^{-1}[G(y_p(\mathbf{u}))] \quad (9)$$

(iv) Identify the kriging estimate $z^*_{\text{MG}}(\mathbf{u})$ to the mean of the ccdf $F(\mathbf{u}; z|(n))$ (E-type estimate); in this case

$$z^*_{\text{MG}}(\mathbf{u}) = \frac{1}{100}\sum_{k=1}^{100} z_p(\mathbf{u}) \quad \text{with } p = k/100 - 0.5/100 \quad (10)$$

An implicit assumption of MK is that the multi-point cdf of the random function $Z(\mathbf{u})$ is Gaussian. Unfortunately, the normality of the one-point cdf (histogram), which is achieved by the normal-score transform, is a necessary but not sufficient condition to ensure the normality of the multi-point cdf (*2*). Although graphical procedures exist for checking the appropriateness of the normality assumption for the two-point cdf (*5*), there is no formal statistical test. Furthermore, to be complete one should also check the normality of the three-point, four-point, ..., N-point cumulative distribution functions, which is unfeasible in practice. For all these reasons, the MG approach is usually adopted with little regard to the underlying assumptions (*12*).

Indicator Kriging (IK). Although the normal-score transform makes the sample histogram perfectly symmetric, it is not well suited to censored data since it requires a necessarily subjective ordering of all these equally valued observations. IK (*6*) is an alternative to the use of MG to infer the ccdf $F(\mathbf{u}; z|(n))$ and the corresponding E-type estimate:

$$z^*_{\text{IK}}(\mathbf{u}) = \frac{1}{100}\sum_{k=1}^{100} z_p(\mathbf{u}) \quad \text{with } p = k/100 - 0.5/100 \quad (11)$$

Estimates (eqs) 10 and 11 differ in the way the ccdf is modeled and so in how the series of quantiles $z_p(\mathbf{u})$ are computed. Instead of assuming that the ccdf is Gaussian and fully characterized by two parameters (parametric approach), the indicator approach estimates the conditional

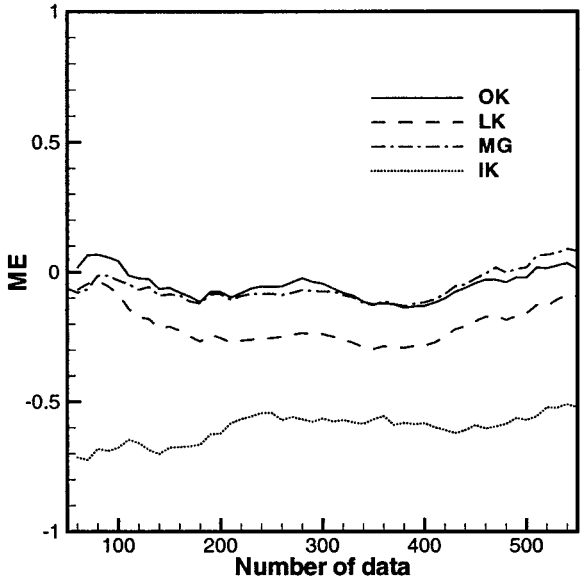

FIGURE 5. Impact of the sample size and type of kriging algorithm on the prediction bias.

probabilities for a series of threshold z_k discretizing the range of variation of z:

$$[F(\mathbf{u}; z_k|(n))]^* = [\text{Prob}\{Z(\mathbf{u}) \leq z_k|(n)\}]^* = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}; z_k)\, i(\mathbf{u}_\alpha; z_k) \quad (12)$$

These probabilities are built as linear combinations of surrounding indicator transforms of the data $z(\mathbf{u}_\alpha)$:

$$i(\mathbf{u}_\alpha; z_k) = \begin{cases} 1 & \text{if } z(\mathbf{u}_\alpha) \leq z_k \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

Kriging weights $\lambda_\alpha(\mathbf{u}; z_k)$ are computed using a system similar to eq 2 except that the semivariogram (eq 3) is computed and modeled for the indicator transforms of the data.

Once conditional probabilities have been estimated and corrected for potential order relation deviations (*5*), the set of K probabilities must be interpolated within each class (z_k, z_{k+1}), equivalent to $z_k < z \leq z_{k+1}$, and extrapolated beyond the smallest and the largest thresholds to build a continuous model for the conditional cdf. In this paper, the resolution of the discrete ccdf has been increased by performing a linear interpolation between tabulated bounds provided by the sample histogram of dioxin concentrations (*5*). Implementation of the IK approach is further discussed in refs *2* and *12*.

The indicator approach appears much more demanding than the multi-Gaussian approach both in terms of semivariogram modeling and computer requirements since K indicator semivariograms must be modeled and K kriging systems must be solved. To allow a fair comparison between IK and other algorithms, median IK has been implemented, which amounts to using the same semivariogram model (typically the one corresponding to the median threshold value $z_M = F^{-1}(0.5)$) for all threshold values. Not only is a single semivariogram model required but also a single kriging system needs to be solved at each location $\mathbf{u}$ since the kriging weights are the same for all threshold values.

Rather than using the same set of thresholds z_k over the study area, the thresholds for IK at $\mathbf{u}$ were identified with the $n(\mathbf{u}) = 16$ closest data values $z(\mathbf{u}_\alpha)$, which is enough for a reasonable discretization of the ccdf (*13*). Such tailoring of thresholds to the local information available allows one to reduce order relation problems that are due to the lack of data in some classes (z_k, z_{k+1}) (*2*). Moreover, it leads to a better resolution of the discrete ccdf by selecting low thresholds in the low-valued parts of the study area and high thresholds in the high-valued parts. The detection limit z_d was used as first threshold when censored data were among the 16 closest observations, which allows a more objective coding of the information available, i.e., no need to assign an "arbitrary" value to censored data. In other words, censored data provide only information on whether the concentration is below or above z_d and not on the particular value within the interval $[0, z_d]$. In this sense, IK is less arbitrary than the three other techniques (OK, LK, and MG) where data below the detection limit were set to z_d.

Despite the existence of other nonparametric algorithms (e.g., probability kriging, rank-order kriging), only IK has been considered in this paper for the following reasons: (a) it is the most commonly used algorithm, and unlike other nonparametric techniques it is implemented in public domain software (*5*); (b) it does not require the tedious modeling of cross-semivariograms involved in probability kriging; and (c) it does not require a prior ranking of all observations, which is a problem in the presence of a large proportion of censored data (recall discussion on normal-score transform).

Materials and Methods

Study site. The study area is an EPA Superfund site called Piazza Road located in Missouri. Dioxin was applied in 1971 with a waste oil as a dust suppressant, and the site was found to be contaminated (*9*, *14*, *15*). Policy in the U.S. EPA Region 7 has set the acceptable risk due to residual dioxin at less than one additional cancer in one million people, which corresponded to a concentration of 1 μg/kg under a residential exposure scenario (*9*). Following a previous study (*9*), the exposure unit (EU) was assumed to be an average residential lot of 464.52 m^2 (15.24 $\times$ 30.48 m).

The pilot study site is 30.48 $\times$ 60.96 m and consists of 4 EUs (Figure 1). Two sets of samples were collected: (i) 200 pairs of tablespoon soil samples were taken 0.30 m on either side of the nodes of a 3.05 m square grid, and (ii) 50 locations were randomly selected in each of the 4 EUs, and a composite sample was formed from 9 tablespoon samples collected within a 0.30-m sampling frame (*9*). The sole isomer of dioxin at the site is 2,3,7,8 tetrachlorodibenzo-*p*-dioxin The laboratory technical procedures (HRGC/MS/MS or HRGC/LRMS) had a high precision (4.5%), a small bias (−2.6%), and a detection limit of 0.3 μg/kg: readings below that limit are reported as 0.3 μg/kg.

Figure 1 shows the location map of 600 dioxin data and their cumulative distribution. The distribution is highly positively skewed (median = 1.17 μg/kg < mean = 4.94 μg/kg). A total of 27% of the observations are below the detection limit, while the threshold of 1 μg/kg is exceeded at 53% of locations. High concentrations are located mainly in the left lower part of the site because of the accumulation of dioxin by water erosion (*9*). Although the large proportion of censored data makes the application of LK nonoptimal, the technique was considered because of its common use in practice. Regarding the multi-Gaussian assumption, a graphical check of the two-point normality indicates that the multi-Gaussian model will lead to an overestimation of the spatial connectivity of high concentrations.

Figure 2 shows the experimental semivariograms for the 600 dioxin data and the three types of transforms (log-normal, normal-score, and indicator). Four directions were considered: N−S, NE−SW, E−W, and NW−SE with an angular tolerance of 22.5°. The fluctuations displayed by the semivariogram of nontransformed data are caused by the presence of high values, the influence of which is reduced by the various transforms yielding smoother curves. The semivariograms

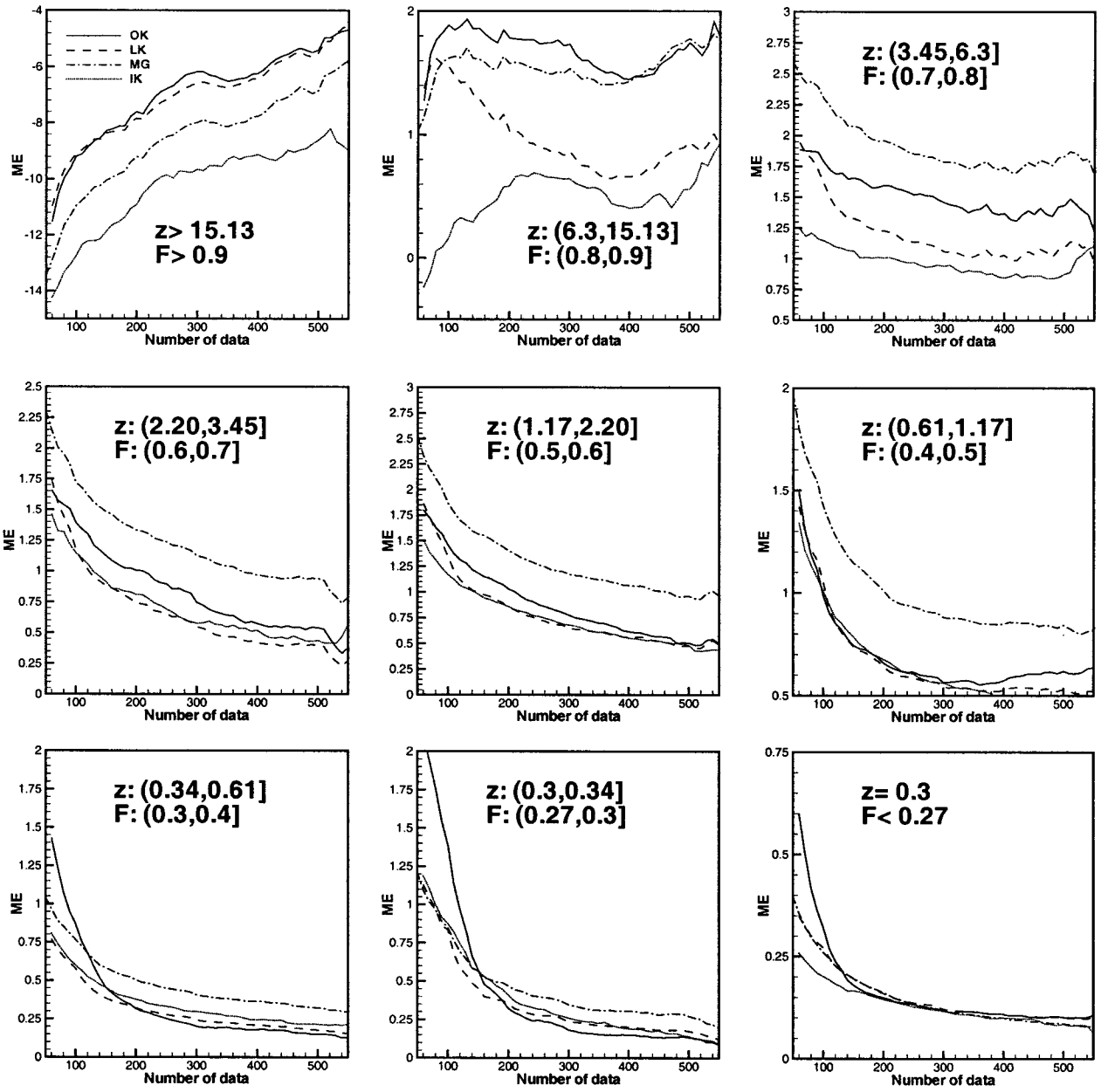

FIGURE 6. Impact of the sample size and class of dioxin concentrations on the mean error (ME) of prediction.

of transforms have similar shape with the larger variability observed along the E−W direction. Geometric anisotropy models were fitted visually to experimental curves following a common practice in the geostatistics literature. To quote (*16*), "Outside a fairly limited class of models (e.g., linear model with nugget effect) variograms are non linear in some of their parameters, typically ranges and anisotropies. Sample variograms also exhibit fluctuations. For these reasons fitting a variogram model is generally done by a geostatistician rather than entirely automatically".

Prediction Performances. The prediction performances of the different techniques with respect to sampling density was investigated using the following procedure:

(1) Select a random subset of N data with $50 \leq N \leq 550$ (prediction set). For each sampling intensity, 100 different random subsets were selected to account for sampling fluctuations.

(2) For each random subset and sampling intensity,

(a) Predict the dioxin concentration at the $600 - N$ remaining locations (validation set) using each of the four algorithms. Regardless of the sampling intensity, the exhaustive semivariogram models of Figure 2 have been used to reduce computational burden and to avoid the use of backbox (fully automatic) semivariogram modeling procedures. Since each technique requires the modeling of a single semivariogram, no algorithm is clearly favored, and the results of the comparative study are not expected to be influenced by this decision.

(b) Compute the mean error (bias ME), the mean absolute error (MAE), and the mean relative error (MRE) of prediction as:

$$\mathrm{ME} = \frac{1}{600 - N} \sum_{j=1}^{600-N} z^*(\mathbf{u}_j) - z(\mathbf{u}_j) \qquad (14)$$

$$\mathrm{MAE} = \frac{1}{600 - N} \sum_{j=1}^{600-N} |z^*(\mathbf{u}_j) - z(\mathbf{u}_j)| \qquad (15)$$

$$\mathrm{MRE} = \frac{1}{600 - N} \sum_{j=1}^{600-N} \frac{|z^*(\mathbf{u}_j) - z(\mathbf{u}_j)|}{z(\mathbf{u}_j)} \qquad (16)$$

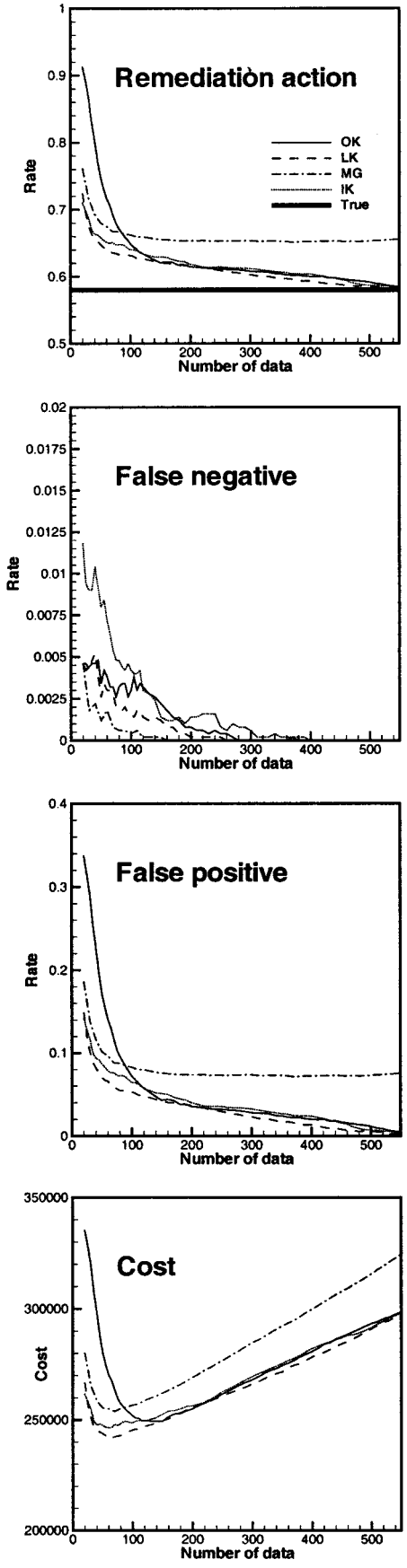

FIGURE 7. Impact of the sample size and type of kriging algorithm on the remediation action rate, the rates of false negative and false positive, and the total cost (sampling + remediation).

where $z^*(\mathbf{u}_j)$ and $z(\mathbf{u}_j)$ are the estimated and true dioxin concentrations at $\mathbf{u}_j$, respectively.

Cleanup Decision. Prediction of dioxin concentrations is not a goal per se, but it is a preliminary step toward the delineation of areas where remediation measures should be taken. The impact of the different interpolation algorithms on remediation decision and costs was investigated using the following procedure:

(1) Select a random subset of N data with $50 \leq N \leq 550$ (prediction set). For each sampling intensity, 100 different random subsets were selected to account for sampling fluctuations.

(2) For each random subset and sampling intensity,

(a) Predict the dioxin concentrations at the nodes of a 1.22-m spacing grid.

(b) Divide the site into 50 remediation units (RUs) of 6.10 m × 6.10 m and decide to remediate a RU if the average estimated concentration exceeds 1 μg/kg. The actual RU concentration is taken as a weighted average of all observations within that unit; paired observations received half the weights of randomly located composite samples.

(c) Compute the proportion of RUs to be remediated (remediation action rate) and the proportion of RUs that are wrongly declared contaminated (false positive) or safe (false negative).

(d) Compute the sum of sampling cost ($156/sample) and remediation cost. In agreement with a prior analysis of the data (9), remediation costs are modeled as a function of only the number of units to be remediated ($7280/RU) and cover the removal, storage, and incineration of the top 10.16 cm of the entire unit. Health costs that result from false negatives have not been considered because of the arbitrariness of their modeling.

Results and Discussion

Prediction Performances. Figure 3 shows, for each kriging algorithm, the average MAE and MRE statistics as a function of the sampling intensity. Besides the mean, the 5th and 95th percentiles of the distribution of 100 values are plotted to illustrate the impact of sampling fluctuations. As expected, increasing the size of the prediction set (sampling intensity) leads to smaller averaged MAE but larger sampling fluctuations because of the smaller overlap of little validation sets. Largest absolute errors are produced by MG and OK, while LK and IK yield smaller errors. Differences are however small both in terms of absolute concentrations and with respect to sampling fluctuations.

Because the distribution of dioxin concentrations is highly positively skewed (recall Figure 1), some validation sets include very large values that can have a great impact on the computation of the MAE. To attenuate the impact of these extreme values, the mean relative errors (MRE) have been computed according to eq 16 and plotted in Figure 3 (bottom graph). The rescaling enhances differences between algorithms, but the ranking remains the same with the best results for IK and LK. Differences between algorithms increase as data become sparser, and so their processing is increasingly important. Collecting more than 200 observations decreases only slightly the prediction error.

MAE has also been computed per classes of true values (conditional errors) to investigate whether the four algorithms perform similarly over the range of concentrations. Figure 4 shows nine conditional MAE curves computed for the seven upper decile classes, the class (from 0.3 to 0.34 μg/kg) and the class including the 27% measurements below the detection limit. Interestingly, the best predictor of the 10% highest concentrations (above 15.13 μg/kg) is OK while IK yields the worst predictions. IK performs best for the next class of concentrations (from 6.3 to 15.13 μg/kg) and for the class that includes all the data below the detection limit. For the

six other classes, LK produces the smallest prediction errors, with IK being a close second.

The prediction unbiasedness is investigated by computing the ME over the full range of concentrations or per classes. Figure 5 shows that ordinary and MG are globally unbiased while the two other algorithms, in particular IK, underestimate concentrations. The bias of ordinary LK was reported by Roth (*17*). The underestimation of IK is caused mainly by the use of surrounding observations as threshold values: such local choice of thresholds entails a zero probability of being smaller than the minimum neighboring observation or of being larger than the maximum neighboring observation. Excluding the minimum and maximum values from the set of thresholds greatly reduces the bias, but it creates unacceptably large MAE, which reflects a balance of large overestimations by large underestimations.

The conditional ME curves of Figure 6 indicate that the IK estimates are systematically the smallest with a large underestimation of the highest observations and a slight overestimation of concentrations below 6.3 μg/kg. Also, the overall unbiasedness of MG is achieved at the expense of a large underestimation of the highest concentrations while medium concentrations are strongly overestimated. Differences between algorithms tend to attenuate for smaller concentrations, except when data are sparse.

In summary, OK and LK are more suitable than other algorithms for the prediction of high values, while IK and LK yield better results for other classes.

Cleanup Decisions. The impact of the various prediction errors on cleanup decisions and on expenses related to sampling and incorrect classification of RUs are now investigated.

Figure 7 (top graph) shows, for each algorithm and sampling intensity, the average proportion of remediation units where the estimated dioxin concentration exceeds 1 μg/kg. The actual remediation rate is 0.58 and is depicted by the thick horizontal line. The LK and IK rates are the closest to the true value and greatly improve over OK when data are sparse. MG leads to larger action rates because of its tendency to overestimate medium concentrations (recall Figure 6). Worse still, MG results do not get better as more data become available.

The proportions of false negatives and false positives are plotted at the middle of Figure 7. Regardless the number of data and the kriging algorithm, the rate of false negatives is below 5%, which is the EPA acceptable rate for residential areas (*9*). The proportion of RUs that are wrongly declared contaminated (false positive) is much larger. Like the remediation action rate, the worst decisions are based on OK estimates for small sampling intensity and on MG estimates when the number of observations exceeds 100. Best results are obtained for LK, with IK being a close second.

While the prediction errors, hence the RU misclassification errors, usually decrease as more data are collected, sampling costs increase and may become prohibitive in practice. Figure 7 (bottom graph) shows the impact of taking additional samples on the sum of sampling costs and remediation costs. At the beginning, increasing the number of observations reduces the rate of false positives and the resulting remediation costs. Beyond 100 observations, additional sampling costs overwhelm the small benefit of lesser false positives, hence the total cost increases. Again, LK and IK outperform MG systematically and OK for small sampling intensities. Even though health costs are unknown, it is clear than an increase in these costs will penalize IK because IK yields a slightly larger proportion of false negatives (recall Figure 7, middle graph). In this case study, the impact of health costs would however be small because of the few false negatives relative to false positives. The number of observations that minimizes the total cost is much larger for OK (140) than for other algorithms (LK, 65; MG, 70; IK, 60).

Literature Cited

(1) Journel, A. G. *Geostatistics for the Environmental Sciences*; EPA project CR 811893; Technical report; U.S. EPA, EMS Lab: Las Vegas, 1987.
(2) Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, 1997.
(3) Myers, J. C. *Geostatistical Error Management: quantifying uncertainty for environmental sampling and mapping*; Van Nostrand Reinhold: New York, 1997.
(4) Journel, A. G.; Huijbregts, C. J. *Mining Geostatistics*; Academic Press: New York, 1978.
(5) Deutsch, C. V.; Journel, A. G. *GSLIB: Geostatistical Software Library and User Guide*; Oxford University Press: New York, 1998.
(6) Journel, A. G. *Math. Geol.* **1983**, *15*, 445−468.
(7) Journel, A. G.; Arik, A. In *Computer Applications in the Mineral Industry*; Fytas, et al., Eds.; Balkema: Rotterdam, 1988; pp 161−171.
(8) Papritz, A.; Dubois, J.-P. In *geoENV II−Geostatistics for Environmental Applications*; Gómez-Hernández, J., et al., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999; pp 429−440.
(9) Ryti, R. T. *J. Air Waste Manage. Assoc.* **1993**, *43*, 197−202.
(10) Journel, A. G. *Math. Geol.* **1980**, *12*, 285−303.
(11) Verly, G. In *Proceedings of the 19th International APCOM Symposium*; Society of Mining Engineers: Littleton, CO, 1986; pp 283−298.
(12) Goovaerts, P. *Geoderma* **1999**, *89*, 1−45.
(13) Deutsch, C. V.; Lewis, R. W. In *Proceedings of the 23rd International APCOM Symposium*, Tucson, AZ, 1992; pp 133−148.
(14) Abt, M.; Welch, W. J.; Sacks, J. *Math. Geol.* **1999**, *31*, 1−22.
(15) Englund, E. J.; Heravi, N. *Environ. Ecol. Stat.* **1994**, *1*, 247−263.
(16) Chilès, J.-P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*; John Wiley & Sons: New York, 1999.
(17) Roth, C. *Math. Geol.* **1998**, *30*, 999−1009.

Received for review December 30, 1999. Revised manuscript received July 13, 2000. Accepted July 18, 2000.

ES991450Y