

Accounting for measurement error in uncertainty modeling and decision-making using indicator kriging and p -field simulation: application to a dioxin contaminated site

Hiroataka Saito[†] and Pierre Goovaerts^{*‡}

Environmental Water Resources Engineering, Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, MI 48109-2125, U.S.A.

SUMMARY

In many environmental studies spatial variability is viewed as the only source of uncertainty while measurement errors tend to be ignored. This article presents an indicator kriging-based approach to account for measurement errors in the modeling of uncertainty prevailing at unsampled locations. Probability field simulation is then used to assess the probability that the average pollutant concentration within remediation units exceeds a regulatory threshold, and probability maps are used to identify hazardous units that need to be remediated. This approach is applied to two types of dioxin data (composite and single spoon samples) with different measurement errors which were collected at the Piazza Road dioxin site, an EPA Superfund site located in Missouri. A validation study shows that the proportion of contaminated soil cores provides a reasonable probability threshold to identify hazardous remediation units. When a lower probability threshold is chosen, the total remediation costs are unreasonably high while false negatives are unacceptably frequent for a higher probability threshold. The choice of this threshold becomes critical as the sampling density decreases. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: measurement error; soft indicator kriging; p -field simulation; misclassification; remediation; dioxin

1. INTRODUCTION

Delineation of contaminated areas usually proceeds in two steps: (i) pollutant concentrations are mapped using interpolation techniques such as kriging, and (ii) a decision rule is applied to the estimated concentrations; for example, the sites where the estimated concentration exceeds a regulatory threshold are classified as hazardous. Most practitioners are aware that such estimates are uncertain and that such uncertainty must be accounted for in the decision making process. Geostatistical modeling of uncertainty has thus received increasing attention in the last few years (Goovaerts, 2001), and techniques fall into two main classes: non-linear kriging and stochastic simulation.

*Correspondence to: Pierre Goovaerts, Environmental Water Resources Engineering, Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, MI 48109-2125, U.S.A.

[†]E-mail: hiroataka@engin.umich.edu

[‡]E-mail: goovaert@engin.umich.edu

Among non-linear kriging techniques, indicator kriging (Journel, 1983) is one of the most commonly used. The approach models the uncertainty of unknown concentrations by a conditional cumulative distribution function (ccdf), which provides the probability that any regulatory threshold value is not exceeded at an unvisited location. The technique is referred to as non-parametric because neither a particular shape nor analytical expression is assumed for the ccdf. Instead, the ccdf value (i.e. cumulative probability) is determined for a series of threshold values discretizing the range of attribute values. For each threshold, the ccdf value is estimated by indicator kriging. The indicator coding is generally hard in that indicators are either zero or one, which means that there is no uncertainty about whether the threshold is exceeded or not at sampled locations. Measurement errors may, however, be non-negligible, in particular when cheap recording devices are used, and this source of uncertainty should be considered in addition to the uncertainty arisen from interpolation in space.

Another challenge is that the model of uncertainty provided by indicator kriging applies only to the measurement support (size of soil cores), which is usually much smaller than the size of remediation units and can be assimilated to a point with respect to the site area. Point ccdfs thus need to be aggregated or 'upscaled' to derive the probability that the average concentration over the remediation unit does not exceed particular thresholds (block ccdf). To account for this change of support, several authors (Journel, 1992; Englund and Heravi, 1994; Goovaerts, 1997a; Kyriakidis, 1997) have used stochastic simulation. The idea is to simulate the spatial distribution of pollutant concentrations across the site. The simulated block value is then computed as the arithmetic average of simulated point values within the remediation unit. Many realizations are generated and the block ccdf is numerically approximated by the empirical distribution of simulated block values.

Uncertainty assessment is not a goal *per se*, but it is a preliminary step for decision making. A common question, then, is at which level of risk or probability threshold should we decide to clean a polluted area (Goovaerts, 1997b). When the probability is very high or very low, the risk-based decision is quite straightforward. Decision making is much more difficult for units with intermediate probabilities, say in the interval [0.3, 0.7]. Therefore, there is an inherent risk that a hazardous area may be declared safe (false negative) or that a safe area may be classified as contaminated (false positive). Such misclassifications should be carefully examined and minimized (Goovaerts *et al.*, 1997), especially false negative rates because of potentially harmful consequences for human health.

The main objective of this article is to present a geostatistical methodology for incorporation of measurement errors and change of support in uncertainty modeling. A combination of soft indicator kriging and *p*-field simulation is used to numerically estimate block ccdfs through the generation of multiple block values. The technique is illustrated with a large soil data set that includes 600 dioxin concentrations measured in a 1858.60 m² (20 000 ft²) site. The impact of the number of observations (sampling intensity) and probability threshold on the site characterization and remediation decisions is investigated by repeating the analysis for a series of probability values and random subsets of original data of various sizes.

2. THEORY

2.1. Soft indicator kriging

Consider the problem of modeling the uncertainty of the value of an attribute z (e.g. a dioxin concentration) at an unsampled location \mathbf{u} , where \mathbf{u} is a vector of spatial coordinates. The information

available consists of z -values at n locations $\mathbf{u}_\alpha, z(\mathbf{u}_\alpha), \alpha = 1, 2, \dots, n$. A fraction of these values is below the detection limit z_d , and such censored data are here reset to half the limit z_d in accordance with previous studies of the data (Zirsky *et al.*, 1985). The uncertainty at \mathbf{u} is modeled by the conditional cumulative distribution function (ccdf) of the random variable (RV) $Z(\mathbf{u})$:

$$F(\mathbf{u}; z|(n)) = \text{Prob}\{Z(\mathbf{u}) \leq z|(n)\} \quad (1)$$

The function $F(\cdot)$ gives the probability that the unknown is no greater than any threshold z . Ccdfs are modeled using a non-parametric (indicator) approach which estimates the probability (1) for a series of K threshold values z_k discretizing the range of variation of z :

$$F(\mathbf{u}; z_k|(n)) = \text{Prob}\{Z(\mathbf{u}) \leq z_k|(n)\} \quad k = 1, \dots, K$$

The indicator approach requires a prior coding of each observation $z(\mathbf{u}_\alpha)$ into a series of K values indicating whether or not the threshold z_k is exceeded. The error attached to a measurement at \mathbf{u}_α is accounted for by replacing the single-valued datum $z(\mathbf{u}_\alpha)$ by a normal distribution with a standard deviation $s(\mathbf{u}_\alpha) = CV \cdot z'(\mathbf{u}_\alpha)$, where CV is the coefficient of variation of the analytical error (Journel, 1986). To incorporate any bias in the laboratory measurement the mean of the distribution is shifted by an amount that is proportional to the observed value; that is, the new mean $z'(\mathbf{u}_\alpha)$ is $z(\mathbf{u}_\alpha) \cdot (1 + \text{bias})$. Thus, the K indicators are given by

$$j(\mathbf{u}_\alpha; z_k) = \text{Prob}\{Z(\mathbf{u}_\alpha) \leq z_k|(n)\} = G\{z_k - z'(\mathbf{u}_\alpha)/s(\mathbf{u}_\alpha)\} \quad (2)$$

where $G(\cdot)$ is the standard normal cdf. The indicators $j(\mathbf{u}_\alpha; z_k)$ have values between 0 and 1 (soft indicators), which expresses the uncertainty about whether thresholds are exceeded at sampled locations.

The ccdf value $F(\mathbf{u}; z_k|(n))$ at the unvisited \mathbf{u} is estimated by a linear combination of soft indicator transforms of neighboring z -data:

$$[F(\mathbf{u}; z_k|(n))]^* = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}; z_k) j(\mathbf{u}_\alpha; z_k)$$

where the weights $\lambda_\alpha(\mathbf{u}; z_k)$ are solutions of a system of $(n(\mathbf{u}) + 1)$ linear equations for each threshold z_k (ordinary indicator kriging, Goovaerts, 1997a). Ccdf values are then interpolated within classes of thresholds $[z_k, z_k + 1]$ and extrapolated beyond extreme thresholds z_1 and z_k . Several options are available. Besides the subjective choice of an analytical interpolation model (e.g. linear, hyperbolic), the shape of the ccdf within a threshold class or beyond extreme thresholds can be assumed similar to the shape of the sample cdf, that is the marginal distribution of data. The technique, known as 'linear interpolation between tabulated bounds', is described in detail in Goovaerts (1997a, pp. 327–328).

2.2. Probability field simulation

Indicator kriging provides a model of uncertainty over a support that corresponds to the size of the soil core, which is usually much smaller than that of the remediation units (RU). A probabilistic classification of RUs requires the derivation of the block ccdf $F_V(\mathbf{u}; z|(n))$ that models the uncertainty

about the average z -value over the RU or block V centered at \mathbf{u} :

$$F_V(\mathbf{u}; z|(n)) = \text{Prob}\{Z_V(\mathbf{u}) \leq z|(n)\} \quad (3)$$

Different techniques are available for the derivation of the function (3) (e.g. see Goovaerts, 2001): (i) multi-Gaussian block kriging which models the block cdf using a normal distribution with mean and variance estimated by simple block kriging of normal score transforms of the data; (ii) volume-variance correction (VVC) of the point cdf (1) which involves a decrease of the cdf variance and a gradual symmetrization of the block cdf as the block size increases. In this article we used a more versatile—yet computationally intensive—alternative provided by stochastic simulation. The block cdf is numerically approximated by the empirical cumulative distribution of many simulated block values (Journel and Huijbregts, 1978, p. 511):

$$F_V(\mathbf{u}; z|(n)) \approx \frac{1}{L} \sum_{l=1}^L i_V^{(l)}(\mathbf{u}; z)$$

where $i_V^{(l)}(\mathbf{u}; z) = 1$ if $z_V^{(l)}(\mathbf{u}) \leq z$, and zero otherwise. Each simulated block value $z_V^{(l)}(\mathbf{u})$ is obtained by averaging a set of z -values simulated at the J points \mathbf{u}'_j discretizing the block $V(\mathbf{u})$:

$$z_V^{(l)}(\mathbf{u}) = \frac{1}{J} \sum_{j=1}^J z^{(l)}(\mathbf{u}'_j) \quad l = 1, \dots, L$$

Attractive features of the simulation approach include: (i) non-linear averaging functions (e.g. geometric or harmonic mean) can be considered; (ii) no assumption is made regarding the impact of the change of support on the shape and variance of the block cdf; (iii) once the grid of point values has been simulated cdfs for various block sizes and shapes can be derived at little computational cost.

The practical implementation of this approach relies on the fast generation of many simulated fields. Among the available geostatistical simulation techniques, p -field simulation (Srivastava, 1992; Froidevaux, 1993) is one of the most straightforward and less CPU demanding. The basic idea is to sample the set of cdfs using a set of autocorrelated probability values $\{p^{(l)}(\mathbf{u}'_j), j = 1, \dots, N\}$, known as probability or p -field, so that the histogram and semivariogram of the resulting set of simulated values $\{z^{(l)}(\mathbf{u}'_j), j = 1, \dots, N\}$ are close to target ones (i.e. sample histogram and model fitted to sample semivariogram):

$$z^{(l)}(\mathbf{u}'_j) = F^{-1}(\mathbf{u}'_j; p^{(l)}(\mathbf{u}'_j)|(n)) \quad j = 1, \dots, N$$

The probability field is generated as a non-conditional realization of a random function $P(\mathbf{u})$ with a uniform marginal cdf, and the covariance $C_P(\mathbf{h})$ is assumed equal to the covariance $C_X(\mathbf{h})$ of the uniform transform of the original variable Z :

$$\begin{aligned} \text{Prob}\{P(\mathbf{u}) \leq p\} &= p \quad \forall p \in [0, 1] \\ C_P(\mathbf{h}) &= C_X(\mathbf{h}) = E\{X(\mathbf{u}) \cdot X(\mathbf{u} + \mathbf{h})\} - [E\{X(\mathbf{u})\}]^2 \end{aligned}$$

where $X(\mathbf{u}) = F(Z(\mathbf{u})) \in [0, 1]$ is the uniform transform of $Z(\mathbf{u})$. The p -field does not need to be conditional since at any datum location \mathbf{u}_α the cdf is a unit-step function centered on the datum value

$z(\mathbf{u}_\alpha)$. Thus, whatever the simulated p -field value $p^{(l)}(\mathbf{u}_\alpha)$ at that location,

$$F^{-1}(\mathbf{u}_\alpha; p^{(l)}(\mathbf{u}_\alpha)|(n)) = z^{(l)}(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha) \quad \forall p^{(l)}(\mathbf{u}_\alpha) \in [0, 1]$$

Multiple realizations over large grids can be generated reasonably fast because cdfs need to be modeled only once, and only multiple non-conditional p -fields need to be generated. Until recently the main shortcoming was the lack of a theoretical basis for the generation of p -fields. Journel (1995), however, proved that, under conditions of ergodicity and on average over a large number L of realizations, the z -histogram and the covariance of the z -uniform scores are reproduced by the set of simulated z -values.

2.3. Selection of remediation units

Whether block V is safe or hazardous can be determined by application of different criteria (Goovaerts, 1997b) to the block cdf $F_V(\mathbf{u}; z|(n))$. A straightforward approach consists of classifying as contaminated all remediation units where the probability of exceeding the regulatory threshold z_c is greater than a critical probability threshold p_c :

$$\text{RU } \mathbf{u} \text{ is hazardous if } \text{Prob}\{Z_V(\mathbf{u}) > z_c\} = 1 - F_V(\mathbf{u}; z_c|(n)) > p_c \quad (4)$$

A first option for selecting p_c consists of negotiating with regulatory authorities to define the meaning of an 'acceptably low' probability for the site under study; for example, both Johnson (1996) and Buxton *et al.* (1997) reported a negotiated probability threshold of 0.2. Garcia and Froidevaux (1997) used two probability thresholds to delineate safe areas ($p_c \leq 0.2$), hazardous areas ($p_c > 0.8$), and uncertain or unclassified areas ($0.2 < p_c \leq 0.8$) where further investigation should be conducted. Using a jackknife approach, Goovaerts (1997a) found that the proportion of wrongly classified locations was minimum for a probability threshold close to the marginal probability of contamination, which designates as contaminated all places where the local risk of contamination exceeds the average risk of contamination over the study area. A similar approach is used here to investigate how the choice of p_c influences the proportion of RUs that are misclassified, that is wrongly declared contaminated (false positives) or wrongly declared safe (false negatives).

3. MATERIAL AND METHODS

The study area is a U.S. EPA Superfund site called Piazza Road located in Rosati, Missouri. In 1971, dioxin was introduced along with a waste oil serving as a dust suppressant; later the site was found to be contaminated (Ryti *et al.*, 1992; Ryti, 1993; Englund and Heravi, 1994; Abt *et al.*, 1999). Policy in EPA Region 7 has set the acceptable risk due to residual dioxin at less than one additional cancer in one million population, which corresponds to a concentration of $1 \mu\text{g}/\text{kg}$ in a residential exposure scenario. In the original study (Ryti, 1993) the site, which is $30.48 \times 60.96 \text{ m}$, was stratified into four exposure units (EU) taken as average residential lots of 464.52 m^2 ($15.24 \times 30.48 \text{ m}$). The remedial strategy was to subdivide each EU into smaller remediation units (RU) so that only hot spots of dioxin would be removed, until the average dioxin concentration within an EU is less than $1 \mu\text{g}/\text{kg}$. If the true concentration after remediation exceeds the action level of $1 \mu\text{g}/\text{kg}$, a false negative would occur, and it cannot happen in more than 5% of EUs to meet EPA Region 7 requirements.

Ryti's study found that the most cost-effective remediation scheme was achieved by stratifying the site into 50 RUs of size $10\text{ m} \times 6.10\text{ m}$. For illustration purposes, in this article exposure units will be identified to remediation units. Thus, the action level of $1\text{ }\mu\text{g/kg}$ as well as the maximum rate of false negatives (5%) will be applied to each of the 50 RUs.

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 four EUs and a composite sample was formed from nine tablespoon samples collected within a 0.30 m sampling frame (Ryti, 1993). The sole isomer of dioxin at the site is 2, 3, 7, 8 tetrachloro-dibenzo-para-dioxin. The laboratory technical procedure (HRGC/MS/MS or HRGC/LRMS) has a high analytical precision (4.5%), a low bias (-2.6%), and a detection limit of $0.3\text{ }\mu\text{g/kg}$. Readings below the detection limit are reset to half the limit, $0.15\text{ }\mu\text{g/kg}$, in accordance with previous studies (Zirschky *et al.*, 1985). For the purpose of this study, measurement errors include laboratory (analytical) errors and errors resulting from incomplete mixing of the field samples. The coefficient of variation (CV) for sampling errors ranges from 11% for composite samples to 15% for single spoon samples which were not homogenized, and these two rates were used for the soft indicator coding (2) of the 200 and 400 concentration data, respectively. Figure 1 (top graphs) illustrates the use of the

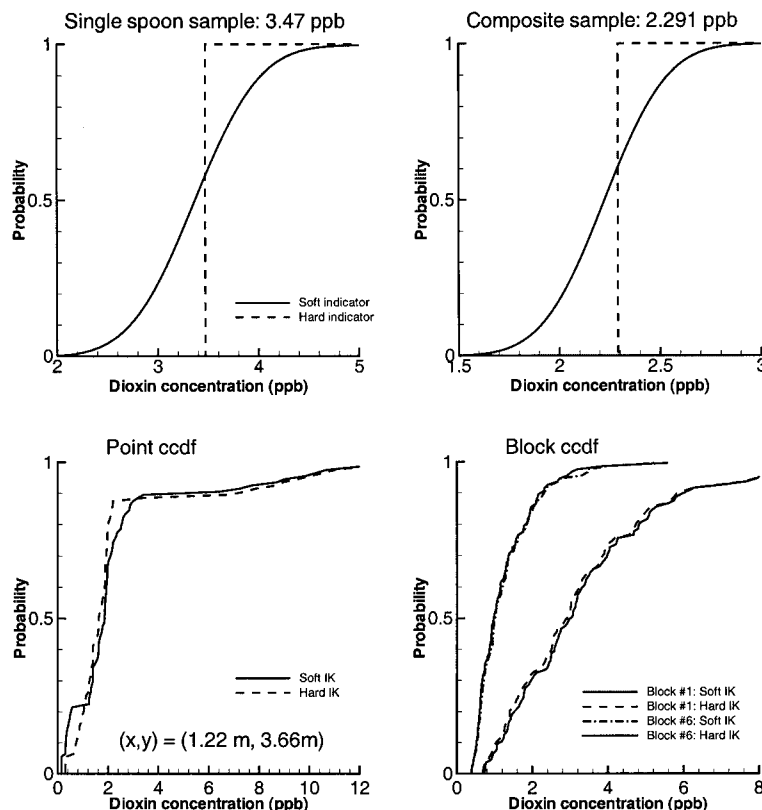


Figure 1. Examples of hard and soft indicator coding for both single spoon samples and composite samples. Point cdf (located in the block #1) generated using soft indicator kriging and two block cdfs that are numerically approximated by the empirical distributions of 100 simulated block values. Dashed lines depict results obtained when measurement errors are ignored

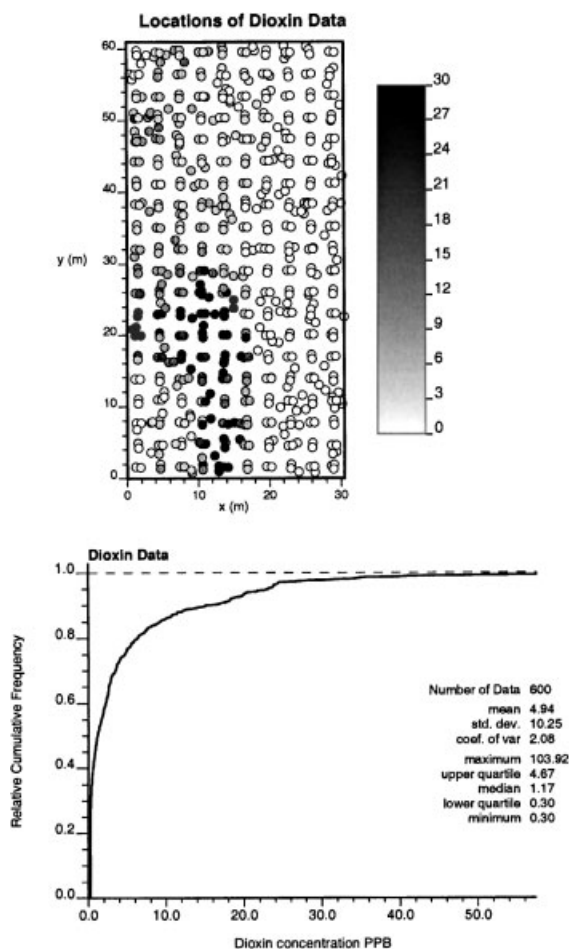


Figure 2. Dioxin data at Piazza Road, MO, and cumulative distribution of dioxin concentrations ($\mu\text{g}/\text{kg}$)

coefficient of variation to derive a probability distribution for composite and single spoon samples. When measurement errors are ignored, the distribution is a step function depicted by the dashed line in Figure 1.

Figure 2 shows the locations and cumulative distribution of 600 dioxin concentrations. The distribution is highly positively skewed (median = $1.17 \mu\text{g}/\text{kg}$ < mean = $4.94 \mu\text{g}/\text{kg}$). Of the observations 27 per cent are below the detection limit, while the threshold of $1 \mu\text{g}/\text{kg}$ is exceeded at 53 per cent of the locations. High concentrations are located mainly in the left lower area of the site due to the accumulation of dioxin by water erosion (Ryti, 1993). A previous study (Saito and Goovaerts, 2000) showed larger spatial connectivity along the N-S direction which is consistent with the direction of the stream channel located in this pilot.

Conditional distributions (ccdf) were derived using ordinary soft indicator kriging and eight threshold values corresponding to the detection limit and the seven deciles of the sample distribution that are larger than $0.3 \mu\text{g}/\text{kg}$. The resolution of the discrete ccdf was increased by a linear

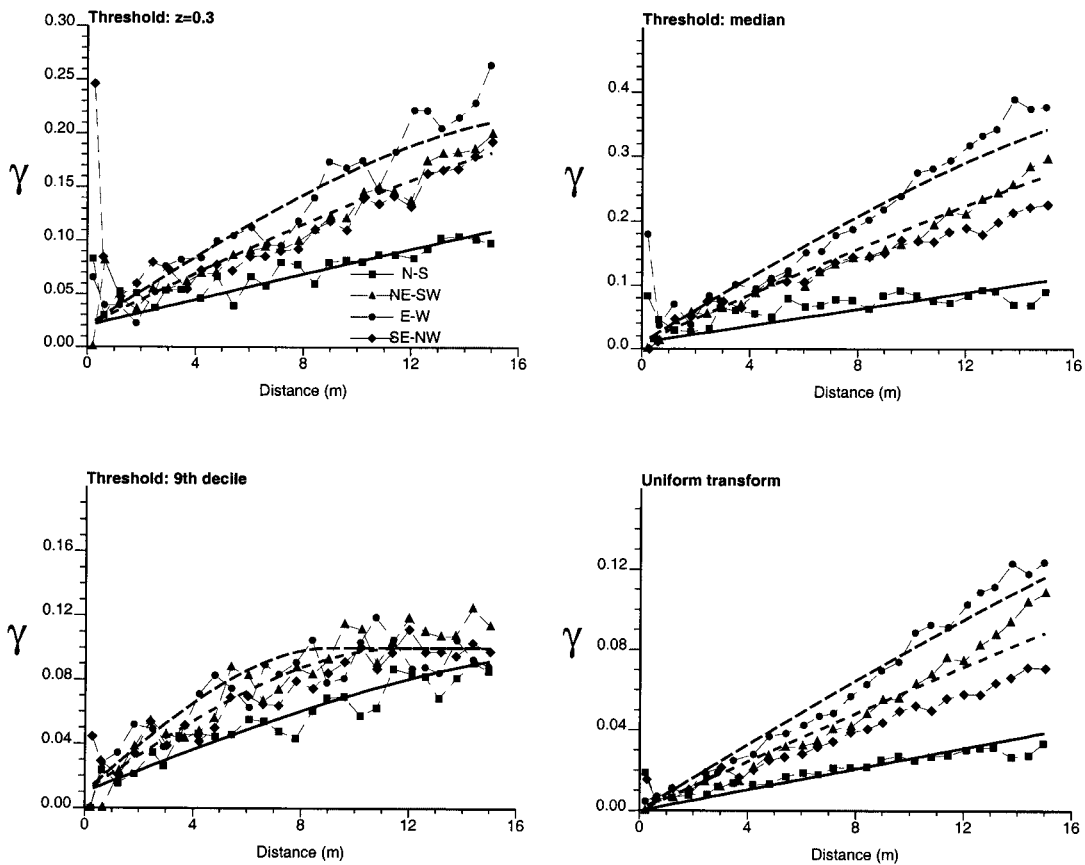


Figure 3. Experimental soft indicator semivariograms (thresholds: $0.3\mu\text{g}/\text{kg}$, median, and 9th decile of the sample histogram) and the semivariogram of uniform transforms, with the geometric anisotropy models fitted

interpolation between tabulated bounds provided by the sample histogram (Deutsch and Journel, 1998). Intuitively, accounting for measurement errors through soft indicator coding should lead to wider cdfs (more uncertainty) than a classical hard coding when extrapolated in space, which is illustrated in Figure 1 (left bottom graph). Experimental soft indicator semivariograms and visually fitted models are displayed in Figure 3, indicating a better spatial connectivity (larger range) of low concentrations. All models consist of two components, a nugget effect and an anisotropic spherical model with a larger range along the E-W direction. One hundred probability fields were generated at the nodes of a 1.22-m spacing grid using sequential Gaussian simulation (Deutsch and Journel, 1998) and the semivariogram of uniform transforms shown in Figure 3 (bottom right). These p -fields were used to sample the sets of cdfs, yielding 100 realizations of the spatial distribution of dioxin concentration values. Figure 4 (top graphs) shows the first realization of simulated punctual values and the block values that are computed as the arithmetical averages of the 25 values falling within each RU. The procedure is repeated 100 times, yielding a distribution of 100 block values whose mean and variance are mapped in Figure 4 (middle graphs). The probability map at the bottom of the figure represents the proportion of simulated block values that exceed $1\mu\text{g}/\text{kg}$. This map is then used for

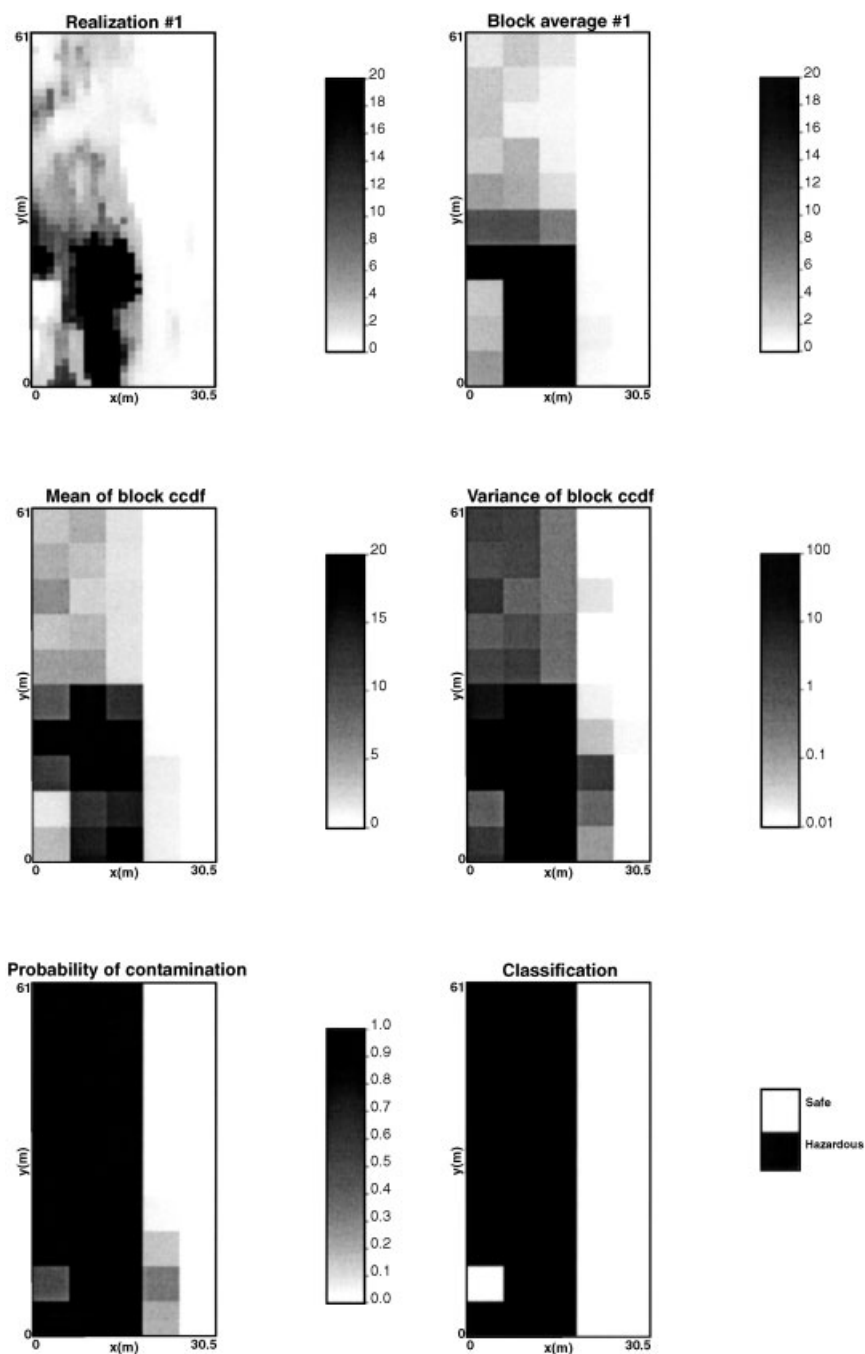


Figure 4. First realization of the spatial distribution of dioxin concentrations generated using p -field simulation conditionally to 300 dioxin data, and the resulting map of block averaged concentrations. One hundred realizations were generated and the mean, variance and proportion of simulated block values that exceed the regulatory threshold of $1 \mu\text{g}/\text{kg}$ are mapped. A block is classified as hazardous if the probability is larger than 0.65

delineation of RUs that are hazardous on the basis of criterion (4). In this example, the probability threshold p_c is set to 0.65 (see explanations in following section).

To investigate the impact of the choice of the probability threshold and sampling intensity on remediation decision, the following procedure was used:

1. Select a random subset of N data values with $50 \leq N \leq 550$ (prediction set). For each sample size, 100 different random subsets were selected to account for sampling fluctuations.
2. For each random subset and sample size:
 - (a) Model the ccdf of the nodes of a 1.22-m spacing grid using soft indicator kriging.
 - (b) Derive the distribution of block simulated values for each remediation unit using p -field simulation, and classify each unit as safe or hazardous on the basis of a series of probability thresholds ranging from 0.05 to 0.95.
 - (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). 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.
 - (d) Compute the sum of sampling costs (\$156 per sample) and remediation costs. In accordance with a previous investigation (Ryti, 1993), remediation costs are modeled as a function of only the number of units to be remediated (\$7280 per RU); this cost covers the removal, storage and incineration of the top 10.16 cm of the entire unit. Health costs which result from false negatives are unknown and not considered here.

4. RESULTS AND DISCUSSION

Figure 5 (top left graph) shows, on average over 100 random subsets, the proportion of remediation units where the probability of exceeding the regulatory dioxin concentration is larger than three widely different thresholds: 0.05, 0.65 and 0.95. For the same probability thresholds, the average proportions of false negatives and false positives are plotted at the bottom of Figure 5. The lack of overlap of 95 per cent confidence intervals depicted by thin lines indicates that the three probability thresholds lead to significantly different results. At the 0.05 probability threshold there are no false negatives regardless of the sampling intensity, while false positives are frequent. Consequently, remediation action rates are high (true rate = 58 per cent) as well as total costs. Minimum cost for the probability threshold of 0.05 is achieved for 110 observations, which is much larger than the 60 obtained by indicator kriging (Saito and Goovaerts, 2000). For the probability threshold of 0.95, the rate of false negatives is above 5 per cent, which is the EPA acceptable rate for residential areas (Ryti, 1993). This choice which leaves the site hazardous to human health is thus unacceptable despite smaller total costs. An intermediate threshold, such as $p_c = 0.65$, leads to a combination of consistent remediation action rates and low misclassification rates.

In an attempt to identify the optimal probability threshold, the procedure was repeated for a series of p_c values and on average over all sampling intensities; see Figure 6 (solid line). As expected, the remediation action rate steadily decreases as the probability threshold increases, and quickly drops beyond 0.9. The misclassification rate, which is computed as the sum of the rates of false negatives and false positives, reaches a minimum value of 1 per cent at $p_c = 0.65$. This value is reasonably close to the marginal probability of contamination (proportion of observations that exceed $1 \mu\text{g}/\text{kg}$), which is 0.53. The discrepancy can be explained by the change of support: the probabilistic classification

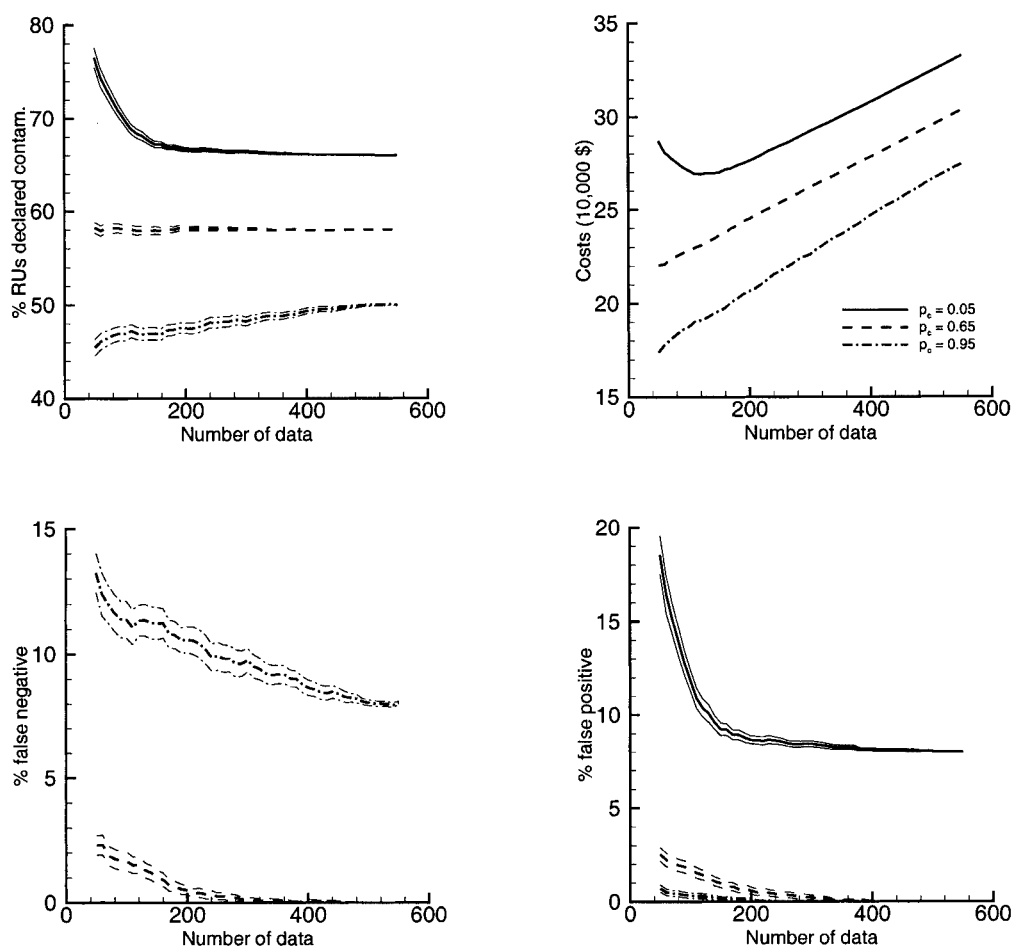


Figure 5. Impact of the sampling intensity on the remediation action rate, the total cost (sampling + remediation), and the rates of false negatives and false positives. Three different probability thresholds are considered: $p_c = 0.05, 0.65$ and 0.95 . Thin lines represent the limits of the 95% confidence interval inferred from 100 random subsets for each sampling intensity

applies to remediation units, while the sample statistics are computed from soil cores. Thus, in absence of further information, the marginal probability of contamination could be used as the threshold. Figure 6 (bottom graphs) shows the impact of the probability threshold for each type of misclassification error. As p_c increases, the proportion of RUs classified as hazardous decreases, which lowers the risk of wrongly classifying RUs as hazardous (false positives). A similar reasoning can be made for the proportion of RUs classified as safe and the rate of false negatives. Note that for most probability thresholds the rate of false negatives is below the acceptable level (5 per cent) set by EPA.

Besides average values, Figure 6 (dashed lines) displays results obtained for three different sampling intensities (50, 300 and 550 observations). As in Figure 5 these curves represent results averaged over 100 random subsets of the same size to attenuate sampling fluctuations. Increasing the number of data lowers the misclassification rates, in particular from 50 to 300 data. When data are sparse, the impact of the choice of p_c is more pronounced.

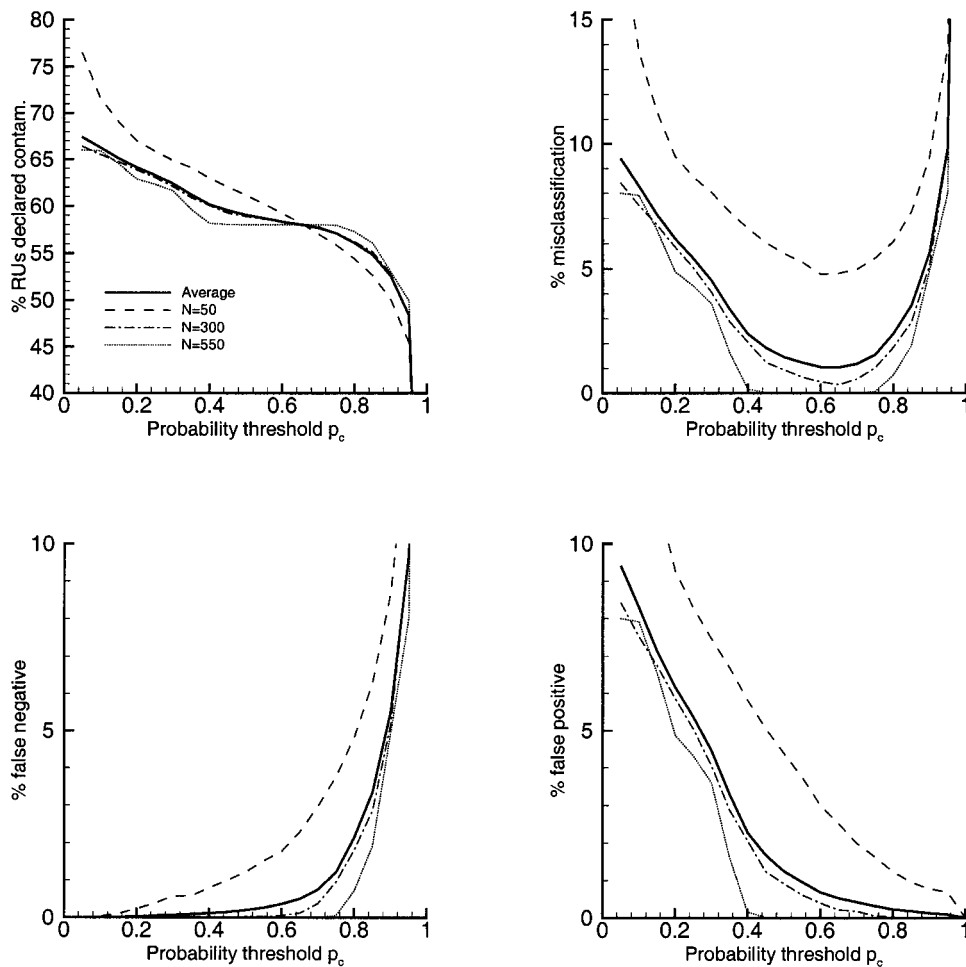


Figure 6. Impact of the probability threshold on the remediation action rate and on the proportions of wrongly classified remediation units (false negatives and false positives). Dashed lines depict results obtained for three different sampling intensities (50, 300 and 550 observations), while the solid line represents the average over all sample sizes

5. CONCLUSIONS

Geostatistics allows the incorporation of measurement errors in site characterization through a soft indicator coding of measurements at each sampled location. Indicator kriging is then used to interpolate these local probability distributions in space, leading to wider probability distributions at unvisited locations. Once the uncertainty arising from both measurement and spatial interpolation errors has been modeled, p -field simulation leads to a fast aggregation of punctual cdfs to derive the probability that the average concentration over a remediation unit does not exceed any particular threshold. The last step is the use of that probabilistic assessment in decision-making, such as the delineation of hazardous areas. The case study confirms that the choice of a probability threshold should be site-dependent and that the marginal probability of contamination is a reasonable choice if

the observations are representative of the site, both in terms of the uniform spatial repartition of samples and of the measurement support. In the presence of data clustering or preferential sampling, declustering procedures (Deutsch and Journel, 1998) could be used to infer a histogram that is more representative of the distribution of concentrations within the entire site. More research should be devoted to the incorporation of the size of remediation units in the choice of a probability threshold.

REFERENCES

- Abt M, Welch WJ, Sacks J. 1999. Design and analysis for modeling and predicting spatial contamination. *Mathematical Geology* **31**: 1–22.
- Buxton BE, Wells DE, Diluise G. 1997. Comparison of three kriging methods for making soil remediation decisions. In *Geostatistics Wollongong '96*, Baafi EY, Schofield NA (eds). Kluwer Academic Publishers: Dordrecht; 984–995.
- Deutsch CV, Journel AG. 1998. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press: New York.
- Englund EJ, Heravi N. 1994. Conditional simulation: practical application for sampling design optimization. *Environ. and Ecol. Stat.* **1**: 247–263.
- Froidevaux R. 1993. Probability field simulation. In *Geostatistics Troia '92*, Vol. 1, Soares A (ed.). Kluwer Academic Publisher: Dordrecht; 73–84.
- Garcia M, Froidevaux R. 1997. Application of geostatistics to 3-D modeling of contaminated sites: a case-study. In *geoENVI—Geostatistics for Environmental Applications*, Soares A, et al. (eds). Kluwer Academic Publishers: Dordrecht; 309–325.
- Goovaerts P. 1997a. *Geostatistics for Natural Resources Evaluation*. Oxford University Press: New York.
- Goovaerts P. 1997b. Accounting for local uncertainty in environmental decision-making processes. In *Geostatistics Wollongong '96*, Baafi EY, Schofield NA (eds). Kluwer Academic Publishers: Dordrecht; 929–940.
- Goovaerts P. 2001. Geostatistical modeling of uncertainty in soil science. *Geoderma* **103**: 3–26.
- Goovaerts P, Webster R, Dubois JP. 1997. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environmental and Ecological Statistics* **4**: 31–48.
- Johnson RL. 1996. A Bayesian/geostatistical approach to the design of adaptive sampling programs. In *Geostatistics for Environmental and Geotechnical Applications*, Rouhani S, et al. (eds). ASTM STP 1283; 102–116.
- Journel AG. 1983. Non-parametric estimation of spatial distributions. *Mathematical Geology* **15**: 445–468.
- Journel AG. 1986. Constrained interpolation and qualitative information. *Mathematical Geology* **18**: 269–286.
- Journel AG. 1992. Geostatistics: roadblocks and challenges. In *Geostatistics Tróia '92*, Soares A (ed.). Kluwer Academic Publishers: Dordrecht; 213–224.
- Journel AG. 1995. Probability fields: another look and a proof. In *Report 8*, Stanford Center for Reservoir Forecasting, Stanford, CA.
- Journel AG, Huijbregts CJ. 1978. *Mining Geostatistics*. Academic Press: New York.
- Kyriakidis PC. 1997. Selecting panels for remediation in contaminated soils via stochastic imaging. In *Geostatistics Wollongong '96*, Baafi EY, Schofield NA (eds). Kluwer Academic Publishers: Dordrecht; 973–983.
- Ryti RT. 1993. Superfund soil cleanup: developing the Piazza Road remediation design. *J. Air Waste Management Association* **43**: 197–202.
- Ryti RT, Neptune D, Groskinsky B. 1992. Superfund soil cleanup. *Environmental Testing & Analysis* **1**: 26–31.
- Saito H, Goovaerts P. 2000. Geostatistical interpolation of positively skewed and censored data in a dioxin contaminated site. *Environmental Science & Technology* **34**: 4228–4235.
- Srivastava RM. 1992. Reservoir characterization with probability field simulation. In *SPE Annual Conference and Exhibition, Washington, DC*, number 24753; Society of Petroleum Engineers: Washington, DC; 927–938.
- Zirschky J, Keary GP, Gilbert RO, Middlebrooks EJ. 1985. Spatial estimation of hazardous waste site data. *Journal of Environmental Engineering* **111**: 777–789.