A geostatistical approach for quantification of contaminant mass discharge uncertainty using multilevel sampler measurements

K. Betty Li,[1] Pierre Goovaerts,[2] and Linda M. Abriola[3]

[1] Contaminant mass discharge across a control plane downstream of a dense nonaqueous phase liquid (DNAPL) source zone has great potential to serve as a metric for the assessment of the effectiveness of source zone treatment technologies and for the development of risk-based source-plume remediation strategies. However, too often the uncertainty of mass discharge estimated in the field is not accounted for in the analysis. In this paper, a geostatistical approach is proposed to estimate mass discharge and to quantify its associated uncertainty using multilevel transect measurements of contaminant concentration (C) and hydraulic conductivity (K). The approach adapts the p-field simulation algorithm to propagate and upscale the uncertainty of mass discharge from the local uncertainty models of C and K. Application of this methodology to numerically simulated transects shows that, with a regular sampling pattern, geostatistics can provide an accurate model of uncertainty for the transects that are associated with low levels of source mass removal (i.e., transects that have a large percentage of contaminated area). For high levels of mass removal (i.e., transects with a few hot spots and large areas of near-zero concentration), a total sampling area equivalent to 6~7% of the transect is required to achieve accurate uncertainty modeling. A comparison of the results for different measurement supports indicates that samples taken with longer screen lengths may lead to less accurate models of mass discharge uncertainty. The quantification of mass discharge uncertainty, in the form of a probability distribution, will facilitate risk assessment associated with various remediation strategies.

Citation: Li, K. B., P. Goovaerts, and L. M. Abriola (2007), A geostatistical approach for quantification of contaminant mass discharge uncertainty using multilevel sampler measurements, *Water Resour. Res.*, *43*, W06436, doi:10.1029/2006WR005427.

1. Introduction

[2] The remediation of dense nonaqueous phase liquid (DNAPL) contamination in the subsurface raises significant technical and nontechnical challenges. When released into the subsurface, relatively large amounts of free phase DNAPL can be trapped within the pores of formations and form a "source zone" that continuously dissolves into the water, creating a persistent long-term plume down-gradient. The complexity of the source zone architecture [*Sale and McWhorter*, 2001] and the heterogeneity of the subsurface make restoration of the contaminated groundwater to drinking water standards or maximum contaminant levels (MCL) currently impractical within a reasonable time frame [*U.S. Environmental Protection Agency (EPA)*, 2004, 2003; *Stroo et al.*, 2003]. Conventional "pump and treat" technology is typically not effective to remedy DNAPL contamination [*Mackay and Cherry*, 1989; *U. S. National Research Council (NRC)*, 1994]. Over the past two decades, innovative remediation technologies [e.g.,

Pennell and Abriola, 1997; *Reddy et al.*, 1995; *Udell*, 1997; *Basel and Nelson*, 2000] have been demonstrated capable of removing considerable amounts of mass [*NRC*, 2004] from source zones within a relatively short time frame. However, even the small amounts of mass left in the subsurface continue to serve as long-term sources adversely impacting down-gradient water quality [*U. S. EPA*, 2003; *Sale and McWhorter*, 2001]. Under these conditions, the traditional metric, point measurement of contaminant concentration, is insufficient to evaluate of the effectiveness of partial mass removal [*Interstate Technology and Regulatory Council (ITRC)*, 2004; *Stroo et al.*, 2003]. Another potential performance metric, the percentage of the source mass removed (mass removal efficiency), may not be well correlated to the impact of the source zone on the down-gradient plume and is not easily predictable a priori [*Lemke et al.*, 2004; *Lemke and Abriola*, 2003; *Rao et al.*, 2002; *Sale and McWhorter*, 2001].

[3] Mass discharge, defined as the contaminant mass per unit time ([M/T]) migrating across a hypothetical control plane orthogonal to the mean groundwater flow, has been proposed as a more intuitive metric [*ITRC*, 2004; *Soga et al.*, 2004; *U. S. EPA*, 2003; *Stroo et al.*, 2003; *Rao et al.*, 2002; *Freeze and McWhorter*, 1997; *Feenstra et al.*, 1996] because it is closely related both to the source mass removal efficiency and to the down-gradient plume. Since numerical modeling studies have yielded different conclusions relating

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan, USA.
[2]BioMedware, Inc., Ann Arbor, Michigan, USA.
[3]Department of Civil and Environmental Engineering, Tufts University, Medford, Massachusetts, USA.

to mass discharge reduction behavior after partial source mass removal, it has been difficult to define an acceptable remediation level in terms of mass removal efficiency. Thus available field information on mass discharge pre- and postremediation has become extremely valuable. Currently, there are three major mass discharge measurement technologies used in the field: the Multilevel Sampling method (MLS), the Integral Pumping Test (IPT) method [*Schwarz et al.*, 1998; *Holder et al.*, 1998; *Bockelmann et al.*, 2001; *Bockelmann et al.*, 2003], and the Passive Flux Meter (PFM) method [*Hatfield et al.*, 2002, 2004]. Unfortunately, regardless of the measurement technology applied, mass discharge measured in the field is subject to great uncertainty [*U. S. EPA*, 2004; *ITRC*, 2004; *Soga et al.*, 2004; *Stroo et al.*, 2003; *Einarson and Mackay*, 2001; *Wilson et al.*, 2000].

[4] Contaminant mass discharge across a control plane is typically computed as $\sum_{i}^{n} C_i q_i A_i$ [e.g., *Borden et al.*, 1997; *Einarson and Mackay*, 2001], where q_i is the flow rate at location i; C_i is the corresponding contaminant concentration, and A_i is the weight for $C_i q_i$, which usually takes the value of the area corresponding to $C_i q_i$. Because C_i and q_i are local measurements (C_i is inversely calculated in the IPT method, and $C_i q_i$ is directly measured in the PFM method), to obtain the mass discharge across the entire control plane, these measurements must be interpolated and aggregated. In previous studies, researchers have used a number of interpolation methods and aggregation sequences [*Kübert and Finkel*, 2006]. In many studies, C_i is interpolated, while q_i is assumed to be uniform [e.g., *Semprini et al.*, 1995; *Borden et al.*, 1997; *King et al.*, 1999; *Kao and Wang*, 2001]. When assuming a uniform q_i, some studies aggregate C_i along monitoring wells first and then among wells [e.g., *Kao and Wang*, 2001]. The Thiessen Polygon method is the most common interpolation technique [e.g., *Borden et al.*, 1997; *Einarson and Mackay*, 2001; *Kao and Wang*, 2001], where the weight A_i takes the value of the influence area (in the shape of a polygon) of each observation. *Semprini et al.* [1995] used a more advanced nonlinear geostatistical estimator to interpolate C_i with the assumption of a uniform flow field. Regardless of the method applied, the calculated mass discharge is always subject to uncertainty arising from interpolation and aggregation processes, even if the measurement errors are negligible.

[5] Although the uncertainty of field-estimated mass discharge is beginning to be widely acknowledged [e.g., *Jarsjo et al.*, 2005; *Zeru and Schafer*, 2005; *Hatfield et al.*, 2004], there have been few attempts to quantify mass discharge uncertainty. Only one case study [*Wilson et al.*, 2000] reported a few statistics (e.g., mean, 95% confidence interval) of variables (e.g., hydraulic conductivity, concentration) for the mass discharge estimation. This study was based, however, on a classical statistical approach which may not be appropriate when the data are correlated in space. In addition, the classical way of expressing uncertainty, using mean and variance, fully characterizes the uncertainty space of a random variable only for particular distribution functions (e.g., Gaussian assumption).

[6] This paper adopts a geostatistical stochastic simulation algorithm (*p*-field simulation) to propagate the uncertainty of mass discharge in a nonparametric way; that is, the

probability distribution of the mass discharge, not just a few statistical moments, is provided. Joint conditional simulation of hydraulic conductivity (K) and contaminant concentration (C) allows the generation of multiple, equally probable realizations of local mass flux ($[M/T \cdot L^2]$), which can then be upscaled to provide the probability distribution of mass discharge ($[M/T]$). As intermediate results, realizations of mass flux can also be used in decision-making processes, such as mapping the probability that the local mass flux exceeds a given threshold. The inputs of this approach are measurements of C and K from multilevel samplers. These measurements do not need to be collocated and the measurement supports (volume that one sample represents) can be different for the two attributes. C and K measurements are, thus, processed separately in this approach, and the results of the geostatistical analysis are then combined and aggregated [*Heuvelink and Pebesma*, 1999]. Direct measurements of mass flux from flux meter observations could be processed in a similar way (equivalent to calculating $C_i q_i$ first, followed by an aggregation) to quantify the uncertainty of mass discharge estimates. This paper focuses on the MLS method, however, because we recognize that the spatial patterns of C and K may be quite different, and hence, the uncertainty analysis may benefit from the spatial information of both C and K.

[7] The approach presented in this paper is essentially a Monte Carlo simulation analysis of the uncertainty resulting from both subsurface geologic and source zone saturation variability [*Beck*, 1987]. The paper first presents the methodology that combines a series of specific geostatistical methods (e.g., indicator kriging, *p*-field simulation) for the uncertainty analysis of mass discharge. Then key factors (e.g., sampling density, screen length, etc.) that control the magnitude of the uncertainty are explored on the basis of the application of the approach to 48 numerically generated transects. The paper ends with an examination of current practice in mass discharge sampling/monitoring in light of these findings.

2. A Geostatistical Approach for Uncertainty Analysis of Mass Discharge

[8] In the geostatistical approach, C and K are modeled as spatial random variables, and the mass discharge across the control plane is estimated as the following linear combination of point estimates of mass flux:

$$M = \sum_{i=1}^{N} -C_i \times K_i \times \text{gradient} \qquad (1)$$

[9] Here C_i and K_i are the interpolated concentration and hydraulic conductivity values at location i, respectively. N is the number of nodes on the sectionalized control plane. The longitudinal gradient is considered a constant across the control plane because it generally varies little on a plane that is orthogonal to the groundwater flow. Note that this method estimates a time snapshot of mass discharge and its uncertainty; therefore, it must be applied either at a steady concentration distribution or on concentration measurements

taken at particular times (e.g., before and after remediation) to illustrate the evolution of mass discharge.

2.1. Propagation of Uncertainty

[10] The computation of mass discharge according to equation (1) relies on the interpolation of C and K measurements at N unsampled locations on a control plane. Each estimate is necessarily uncertain and the associated error (thus, the uncertainty) will impact the accuracy of the mass discharge estimate. The uncertainty of mass discharge can be assessed by propagating the uncertainty attached to the estimated C and K values through equation (1) (Monte Carlo simulation). In this paper, we propose to use a simulation-based approach [*Goovaerts*, 2001] to model the uncertainty relating to the spatial distribution of C and K values within the control plane and to propagate it through equation (1) to obtain the mass discharge uncertainty. The basic idea is to generate multiple, equally probable realizations of local mass flux at each location i on the control plane: $\{M_i^{(l)}, i = 1,\ldots,N; l = 1,\ldots,L\}$. Each realization of simulated mass flux values can then be aggregated over the control plane, yielding a set of L simulated mass discharge values that empirically define the probability distribution of mass discharge. The probability distribution fully characterizes the uncertainty space because it provides the probability that the true value of the mass discharge is no greater than any given mass discharge threshold. Local mass flux values, $M_i^{(l)}$, are computed as the product of collocated K and C values, $K_i^{(l)}$ and $C_i^{(l)}$, that need to be simulated jointly in order to account for their potential correlation.

2.2. Joint Simulation of K and C Fields

[11] Methods to simulate random fields jointly can be broadly classified into two main categories: the spectrum-based approach and the covariance-based approach. Spectrum-based approaches jointly generate cross-correlated random fields with the Discrete Fourier Transform (DFT) algorithm or the Fast Fourier Transform (FFT) algorithm [*Robin et al.*, 1993; *Gutjahr et al.*, 1994; *Ruan and McLaughlin*, 1998]. The common limitation of these methods is that the generated random fields are Gaussian, because of the derivation processes [*Gutjahr et al.*, 1994]. In addition, a transform function between variables (e.g., based upon a physical relationship) is needed to obtain the cross spectrum function, and multivariate normality is usually assumed [*Gutjahr et al.*, 1994; *Robin et al.*, 1993]. Theoretically, covariance-based approaches can jointly simulate two or more variables by directly simulating a vectorial random function using sequential simulation algorithms such as Sequential Gaussian Simulation (SGS) [*Gómez-Hernández and Journel*, 1993; *Verly*, 1993] or Sequential Indicator Simulation (SIS) [*Gómez-Hernández and Srivastava*, 1990; *Goovaerts*, 1997]. In these methods, the difficulty resides in the inference and modeling of the cross covariance matrix. Not only might there not be enough data for the reliable inference of each covariance function, but all the direct and cross covariance functions must be fit jointly to ensure that the linear model of coregionalization (LMC) is permissible [*Goovaerts*, 1997]. The implementation becomes even more difficult when Gaussian-related algorithms are not suitable. The indicator-based algorithm requires the joint modeling of a (cross) covariance matrix that not only includes each variable, but also all the

thresholds for each variable, a tedious and impractical task. Note that for both spectrum-based and covariance-based approaches, an alternative to the direct simulation of correlated variables using a cross covariance matrix or cross spectrum function is to separately simulate a set of independent factors [i.e., factors obtained through principal component analysis (PCA) or minimum/maximum auto-correlation factors (MAF)] from which the original variables can be reconstituted [*Luster*, 1985; *Desbarats and Dimitrakopoulos*, 2004].

[12] The approach proposed herein is covariance-based. It alleviates the cross covariance modeling effort using the method introduced by *Almeida and Journel* [1994] whereby a hierarchy of variables is defined and then each variable is simulated in turn. This method is built on the concept of collocated cokriging and the Markov-type approximation: the collocated higher ranking datum Z_1 (**u**) prevails over (screens) the influence of any more distant data Z_1 (**u** + **h**) on the currently simulated data Z_2(**u**) (lower ranking). Here **u** is the location vector, and **h** is the displacement. This hypothesis greatly simplifies the modeling step at the expense of no control on the cross covariance at $|\mathbf{h}| > 0$. Another limitation is the underlying multiGaussian assumption [*Almeida and Journel*, 1994; *Grimmet and Stirzaker*, 2001]. Other cross covariance modeling methods are available, such as least squares techniques [*Goulard and Voltz*, 1992], FFT approaches [*Yao and Journel*, 1998], and modified MAF techniques [*Vargas-Guzman*, 2004]; see references for the discussion of their advantages and limitations.

2.3. The p-Field Simulation Approach

[13] In this paper the probability-field simulation, commonly referred to as 'p-field' simulation, is adopted to jointly simulate contaminant concentration and hydraulic conductivity. The p-field simulation is a popular algorithm introduced by *Srivastava* [1992] and further documented by *Froidevaux* [1993]. *Journel* [1995] showed that theoretically, the univariate and bivariate statistics of the p-field simulated values are correct in the absence of any local conditioning data (i.e., sample data). The effectiveness of this algorithm has been demonstrated in many field applications [e.g., *Goovaerts*, 2002; *Saito and Goovaerts*, 2002; *Painter*, 2001; *Yao and Chopra*, 2000; *Mao and Journel*, 1999; *Dungan*, 1998]. Unlike sequential simulation algorithms, the p-field simulation algorithm separates the task of constructing conditional cumulative distribution functions (ccdfs) from the task of covariance reproduction by using autocorrelated probability values (p values) to sample the preconstructed ccdfs. Thus literally any algorithm can be used to construct ccdfs, and the spatial structure of the target variable is imparted from the spatial structure of the set of p values, known as the p field. The simulation proceeds as follows: (1) build ccdf model at each location using any appropriate algorithm (e.g., indicator or multi-Gaussian kriging); (2) generate unconditional realizations of autocorrelated probability values (p field) with a uniform marginal distribution and covariance structure of the uniform transform of the original variable; (3) generate conditional realizations of the original variable by sampling the ccdf models using the corresponding simulated probability at each location. In this paper, ccdfs were constructed using the indicator-based approach, and the Markov-type hypothesis was used to jointly simulate
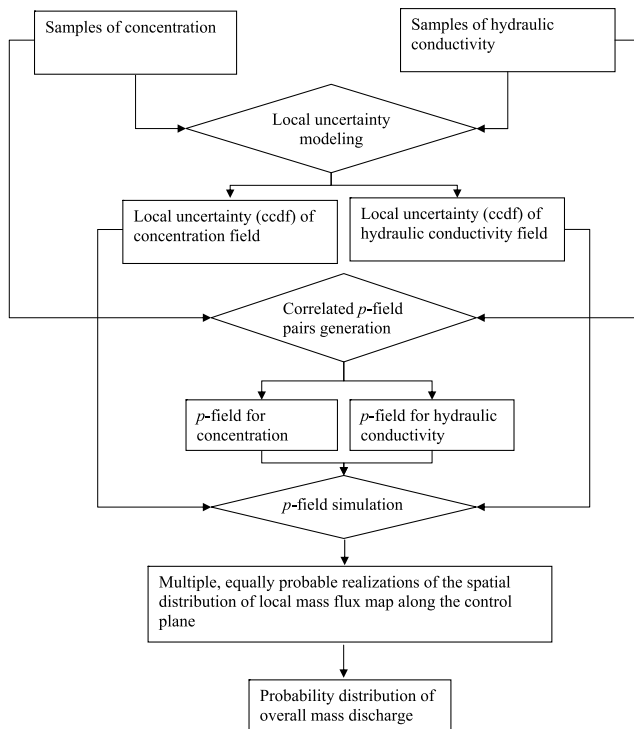
Figure 1. Flowchart of the geostatistical approach for uncertainty analysis of mass discharge using multilevel transect data.

"p-field pairs" for C and K, where the p values are derived from normal scores generated using sequential Gaussian simulation [Goovaerts, 1997].

2.4. The Approach Flowchart

[14] As shown in Figure 1, the proposed approach comprises four major steps: (1) model the local uncertainty (ccdfs) of C and K using indicator kriging; (2) generate unconditional, correlated p-field pairs and use them to sample the ccdfs of C and K; (3) propagate the uncertainty of local mass flux using the sampled C and K values (Monte Carlo simulation) and generate realizations of the spatial distribution of local mass flux within the control plane; and (4) upscale the local mass flux realizations to generate the probability distribution of the mass discharge across the entire control plane.

[15] Local uncertainty modeling of C and K is achieved by indicator kriging, which does not assume any type or shape for the probability distribution of a random variable (nonparametric approach). The ccdf is modeled through a series of thresholds that partitions the range of variation of the physical attribute [Goovaerts, 1997]. The joint simulation algorithm proposed by Almeida and Journel [1994] is adopted to generate correlated p-field pairs. In order to avoid artificially defining a hierarchy between C and K, the first p field (for either C or K) is generated through SGS unconditionally. This p field serves as soft information to cosimulate the second p field using collocated cokriging and the Markov-type approximation. The multiple, equally probable realizations of p-field pairs are used to sample the ccdfs of C and K to generate realizations of C, K.

Multiple realizations of C and K are then incorporated into equation (1) to simulate local mass flux. Aggregation of one local mass flux realization results in one mass discharge realization (a single value). The ensemble of simulated mass discharges provides the probability distribution of the overall mass discharge across the entire control plane.

3. The Synthetic Data Sets

[16] In the field, it is impossible to measure the "true value" of the mass discharge across a control plane. However, the true value should be known to evaluate the quality of the modeled mass discharge uncertainty. Therefore numerically simulated transects were constructed on the basis of a real field site [Abriola et al., 2005]. These transects served as the reference fields that provide exclusive information to test the approach.

[17] The geostatistical approach presented here was tested on 48 reference fields, which were provided by Lawrence Lemke [Lemke et al., 2004] and John Christ [Christ et al., 2005, 2006]. The generation process was designed to produce physically realistic plume transect data associated with heterogeneous source zones. Field sample data were obtained from the Bachman road site, MI [Abriola et al., 2005], where the perchloroethylene (PCE)-contaminated aquifer is composed of relatively homogeneous glacial outwash sand. The Kozeny-Carman hydraulic conductivity estimates varied from 1 to 48 m/d, with a lognormal variance of 0.29. Representative statistics such as histograms and semivariograms were calculated to construct three-dimensional realizations of the formation, where all of the realizations were conditional to the sample data and the calculated statistics [Lemke et al., 2004]. The source zones generated from two-dimensional simulations of hypothetical PCE infiltration and entrapment were analyzed to select representative three-dimensional formation realizations, which were then used to generate three-dimensional source zones using the multiphase simulator University of Texas Chemical Compositional Simulator (UTCHEM) [Lemke et al., 2004; Christ et al., 2005]. After this step, 16 representative source zones were used as initial inputs for a modified version of MT3D [Parker and Park, 2004] to simulate mass recovery (dissolution) processes and the evolution of the downstream plume [Christ et al., 2006]. The three-dimensional domain for this mass recovery simulation was 7.9248 m long (x) by 7.9248 m wide (y) by 9.7536 m deep (z). The initial spill was at the center of the top "x-y" plane ("x" is the groundwater flow direction). The resolution of the hydraulic conductivity field was 0.3048 m by 0.3048 m by 0.3048 m, which is based upon the measurement support of the field data. The resolution for the concentration field was 0.3048m (x) by 0.3048m (y) by 0.0762m (z). For more details regarding the synthetic data set, please refer to the studies by Lemke et al. [2004] and Christ et al. [2005, 2006].

[18] The downstream boundary "y-z" plane of the three-dimensional plume realizations was selected as the location for the "control plane". This plane was located 3.962m downstream of the hypothetical spill. The concentration distributions on the control plane changed with time as the mass recovery was progressing. To obtain representative concentration distributions, the following temporal events in
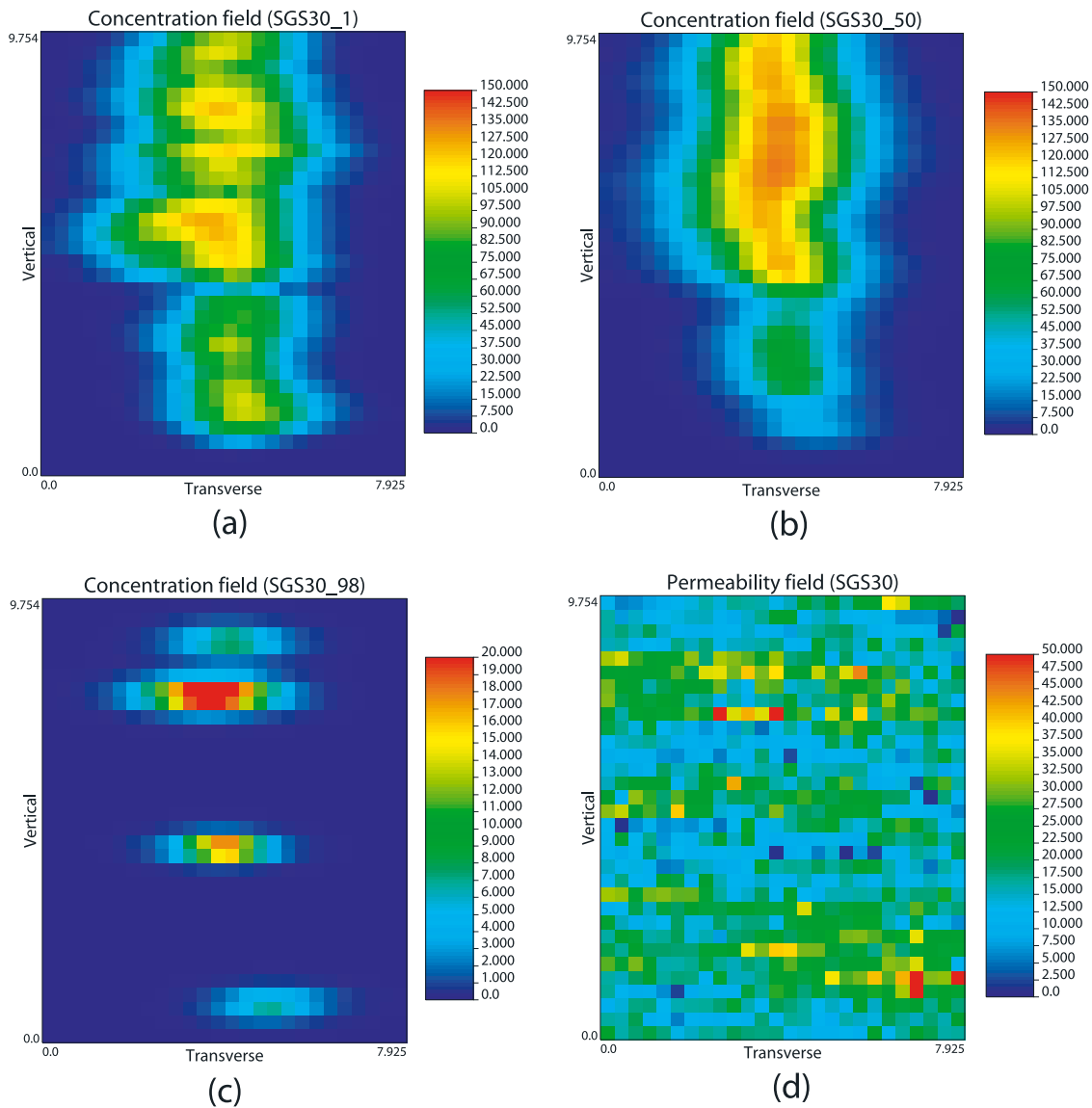
Figure 2. Illustration of the reference fields. (a) Concentration field corresponds to 1% of source mass removal (mg/l); (b) concentration field corresponds to 50% of source mass removal (mg/l); (c) concentration field corresponds to 98% source mass removal (mg/l); (d) hydraulic conductivity field (m/d).

the mass recovery process were selected for each of the 16 realizations: the times at which 1, 50, and 98% of the source mass removal had occurred. As a result, there are 16 reference hydraulic conductivity fields and 48 reference concentration fields.

[19] An example of the planar cross sections is illustrated in Figure 2. Note that the 1 and 50% mass removal cases share similar spatial patterns for concentration. The highly concentrated area is spatially continuous and occupies a relatively large area on the transects, which is similar to the pretreatment situation. In contrast, the spatial distribution of concentrations for the 98% mass removal case is much more heterogeneous, with scattered small hot spots and large near-zero concentration areas. This type of spatial pattern is more likely to be encountered in the field immediately

downstream of a source zone and can be considered as the distribution characteristic of posttreatment.

4. Approach Demonstration and Evaluation

4.1. Parameters for the Geostatistical Approach

[20] The approach requires some user-defined parameters that are standard for geostatistical methods implemented in the software library Geostatistical Software Library (GSLIB) [*Deutsch and Journel*, 1998]. These include the maximum number of data for kriging, the number of thresholds for indicator kriging, the interpolation/extrapolation methods for the enhancement of the resolution of ccdfs, and the number of realizations for the *p*-field simulation. See the works by *Deutsch and Journel* [1998] and *Goovaerts* [1997]

for a general discussion. Preliminary results showed that the mass discharge estimation was not sensitive to the maximum number of data used in kriging when sample data are regularly spaced. This parameter was set to 24 for all of the simulations presented herein. The thresholds for indicator kriging were selected as the nine deciles of the sample histogram in order to achieve a reasonable discretization of the range of variation. If an empty class occurred (i.e., existence of identical deciles), the number of thresholds was reduced by 1 and all of the thresholds were redefined automatically.

[21] For all of the simulations in this section, the resolution of the ccdfs was enhanced by interpolating between thresholds according to the shape of the marginal distribution (cumulative distribution function, cdf of the sample data), which recovered some information that was lost during the indicator transformation process [*Deutsch and Journel*, 1998]. This method was also used in the extrapolation of ccdfs with the allowed maximum set to the maximum sampled value. The number of realizations for the *p*-field simulation determines whether the uncertainty space is fully explored. *Goovaerts* [1999] found that 20 is the minimum number to achieve a stable assessment of the extent of the space of uncertainty for major simulation algorithms. In this study, preliminary simulations confirmed that 20~25 realizations allowed one to capture the extent of the uncertainty space. The number of realizations was conservatively set to 100 for the results presented herein.

4.2. Sampling Designs

[22] The reference fields in this demonstration adopted the resolution of the hydraulic conductivity field and used both 0.3048 m (1 ft) and 0.6096 m (2 ft) as the vertical resolutions for the concentration field to explore the influence of the screen length. Finer vertical resolution for the concentration field was not used because the concentration varies little within the 0.3048-m intervals. The same resolution (support) was used for the hydraulic conductivity field to be consistent with the measurement support of the field data used to construct the synthetic permeability fields; *K* values were measured in the lab by homogenizing 0.3048-m segments of soil cores [*Lemke et al.*, 2004]. Although the same measurement support for *C* and *K* was assumed in the present case study, the proposed methodology can be applied in presence of different supports since *C* and *K* values are first interpolated separately.

[23] "Multilevel samplers" were positioned in the reference fields according to a regular pattern, consistent with vertical well installation. Without prior information, a regular grid is the most appropriate sampling pattern to cover the whole area and to prevent preferential sampling. The circles in Figures 3a and 3b indicate the locations of sample measurements. A series of regular sampling events with different sampling densities was applied on each reference field. Sampling density is defined as the ratio of the sampled area over the whole area of the control plane, which in the demonstration is calculated as $\frac{N \times S}{77.2953 \text{m}^2}$, where N is the number of samples, S is the measurement support, and 77.2953 m^2 is the area of the reference field. Although sampling density is a convenient metric, note that the horizontal resolution of the reference field is 0.3048 m, which may be too large to mimic the horizontal sampling scale in the field (i.e., well diameter). This factor may limit

direct comparison with field-sampling densities. The number of samples is therefore also reported herein.

4.3. Accuracy and Precision of Mass Flux Uncertainty Models

[24] The output probability distributions for the 48 reference fields were analyzed to explore the impact of the sampling density on their accuracy and precision. Figure 3c illustrates the output probability distribution of mass discharge. The dot below the *x* axis depicts the true value of the mass discharge. The sides of the "box" represent the 25 and 75% quantiles, respectively (the range in between is the 50% confidence interval). The line represents the 95% confidence interval bounded by the 2.5 and 97.5% quantiles. The probability distribution is accurate to some degree if the true value is contained within the fixed probability interval. For a symmetric distribution, the ideal situation is when the mean of the distribution coincides with the true value. The spread of the probability distribution illustrates another assessment criterion: precision. The sharper the curve, the more precise is the distribution.

[25] The accuracy plot (e.g., Figure 4a) is a tool that can be used to assess the accuracy and precision of models of uncertainty. It was originally proposed by Deutsch [*Deutsch*, 1997] to evaluate local uncertainty models (ccdfs) and plots the frequencies (computed from ccdfs) for which local true values fall within a series of symmetric *p*-probability intervals (PI) bounded by the $(1 - p)/2$ and $(1 + p)/2$ quantiles. An uncertainty model is accurate if the frequency of true values falling in the *p*-PI exceeds *p* for all *p*-PIs. The precision of an accurate uncertainty model is measured by the closeness of the computed frequency to *p*. In this paper, the accuracy plot was adapted to evaluate the uncertainty space of the mass discharge modeled by the proposed approach. Given a certain sampling strategy (number of samples, measurement support, sampling locations), the probability distributions derived from different reference fields were analyzed to compute the frequency of occurrence of the true values within a series of symmetric *p*-PIs. The accuracy plot then depicts the calculated frequencies as a function of the expected probabilities (*p*). For example, to create one curve (representing one sampling density) on the accuracy plot in Figure 4a, 16 reference fields were used, and 16 true values of mass discharge were calculated for each reference field. A series of symmetric *p*-PIs with *p* varying from 2 to 100% (50 *p*-PIs in total) was calculated from the 16 output probability distributions. Theoretically, the frequency with which the *p*-PI includes the true value should be close to the expected probability *p*; that is, all of the points should fall along the 45° line. If all of the points fall above the 45° line, the observed frequency is higher than the expected probability, indicating that the model of uncertainty is accurate but is not as precise as expected.

[26] Because the spatial distribution of concentration varies remarkably among the reference fields with 1, 50, and 98% source mass removal (Figure 2), the approach was assessed separately on three sets of reference fields (each set has 16 reference fields). For the screen length of 0.3048 m, regular sampling with sampling densities varying from 1 to 25% (number of samples from 9 to 208) was applied on each reference field. For each set of reference fields, accuracy plots for all of the sampling densities used were
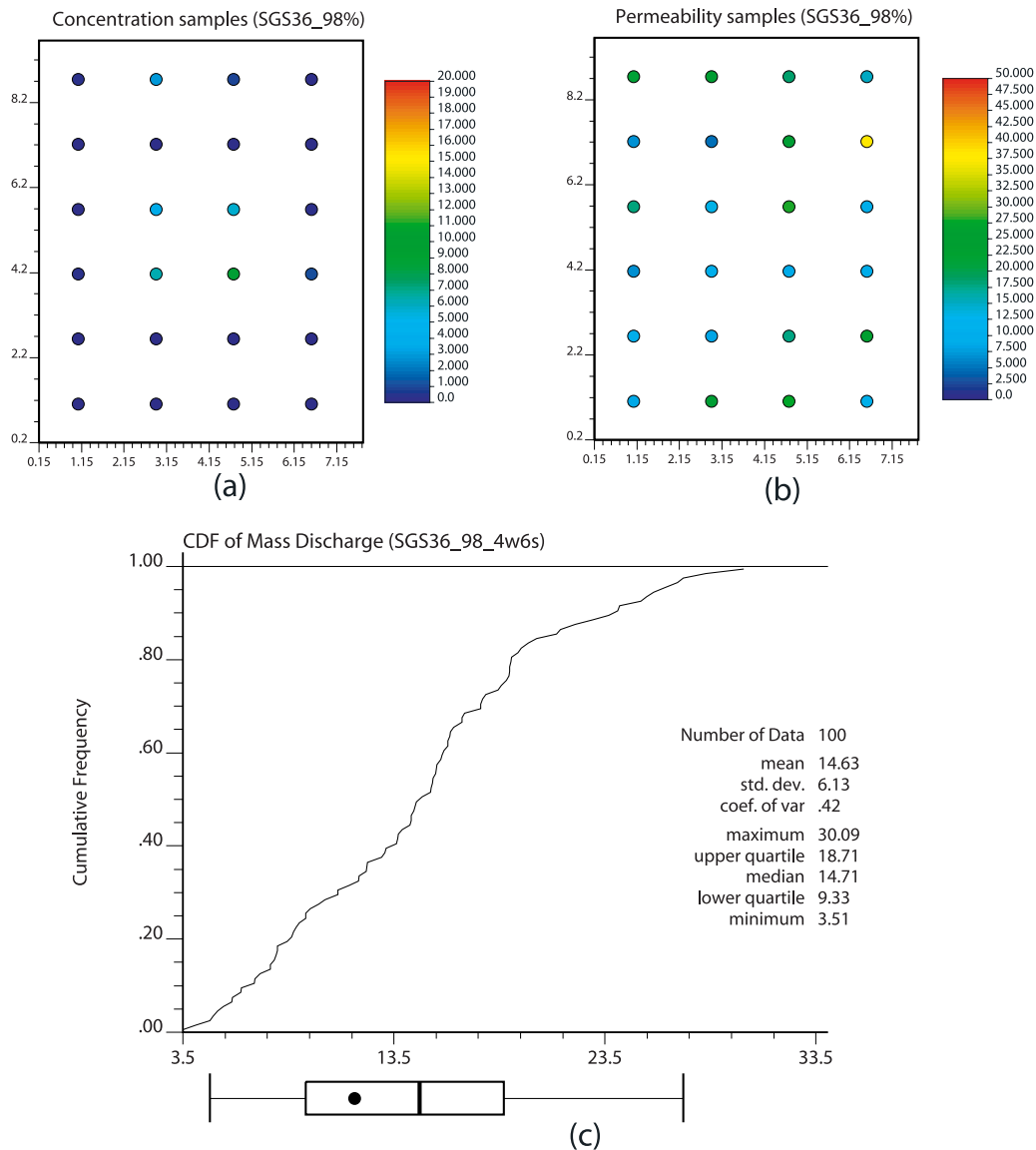
Figure 3. An example of the inputs and output of the proposed approach. (a) Concentration samples (mg/l); (b) hydraulic conductivity samples (m/d); (c) the output probability distribution of mass discharge.

grouped into one figure. Different symbols represent different sampling densities. Figure 4 shows the grouped accuracy plots for the 1, 50, and 98% mass removal cases, which leads to the following findings:

[27] 1. For concentration distributions that display large continuous regions of high concentration (corresponding to low levels of mass removal (Figures 2a and 2b), the approach provides an accurate model of uncertainty for all of the sampling densities used in this study. All of the points fall above the 45° line, indicating that the frequency of capturing the true value is higher than the theoretical (i.e., expected) probability value, which means that the uncertainty model is accurate but conservative (The reasons for this behavior are discussed further in sections 4.5 and 4.6).

[28] 2. For concentration distributions that consist of scattered hot spots and large areas of near-zero concentration [the case for high levels of mass removal (Figure 2c)], a sampling density of 6∼7% (corresponding to 50∼60 samples in an area of 77.3 m^2) is required to obtain an accurate

model of uncertainty, according to the accuracy plots. This sampling density appears unrealistic for field applications. For example, *King et al.* [1999] used 21 points (six wells) in an area of 50 m^2. Assuming the same measurement support of 0.0929 m^2, the sampling density is 3.9%. *Kao and Wang* [2001] had 20 points (five wells) in an area of 283.8 m^2. The screen length is 0.6 m. If it is assumed that the horizontal scale of each sample is 0.3048 m as in this work, their sampling density is 1.3%. Einarson and Mackay used 49 points (seven wells) in an area of 154.29 m^2 with unknown measurement support. Again, assuming 0.0929 m^2 measurement support, the sampling density is 3.0%.

[29] According to the definition of the accuracy plot, whether or not the uncertainty model is accurate is not only influenced by the closeness of the median of the probability distribution to the true value but also by the width of the p-PIs. Therefore to further elucidate the performance of the proposed approach, the accuracy and precision of the uncertainty models were assessed respectively by the fol-
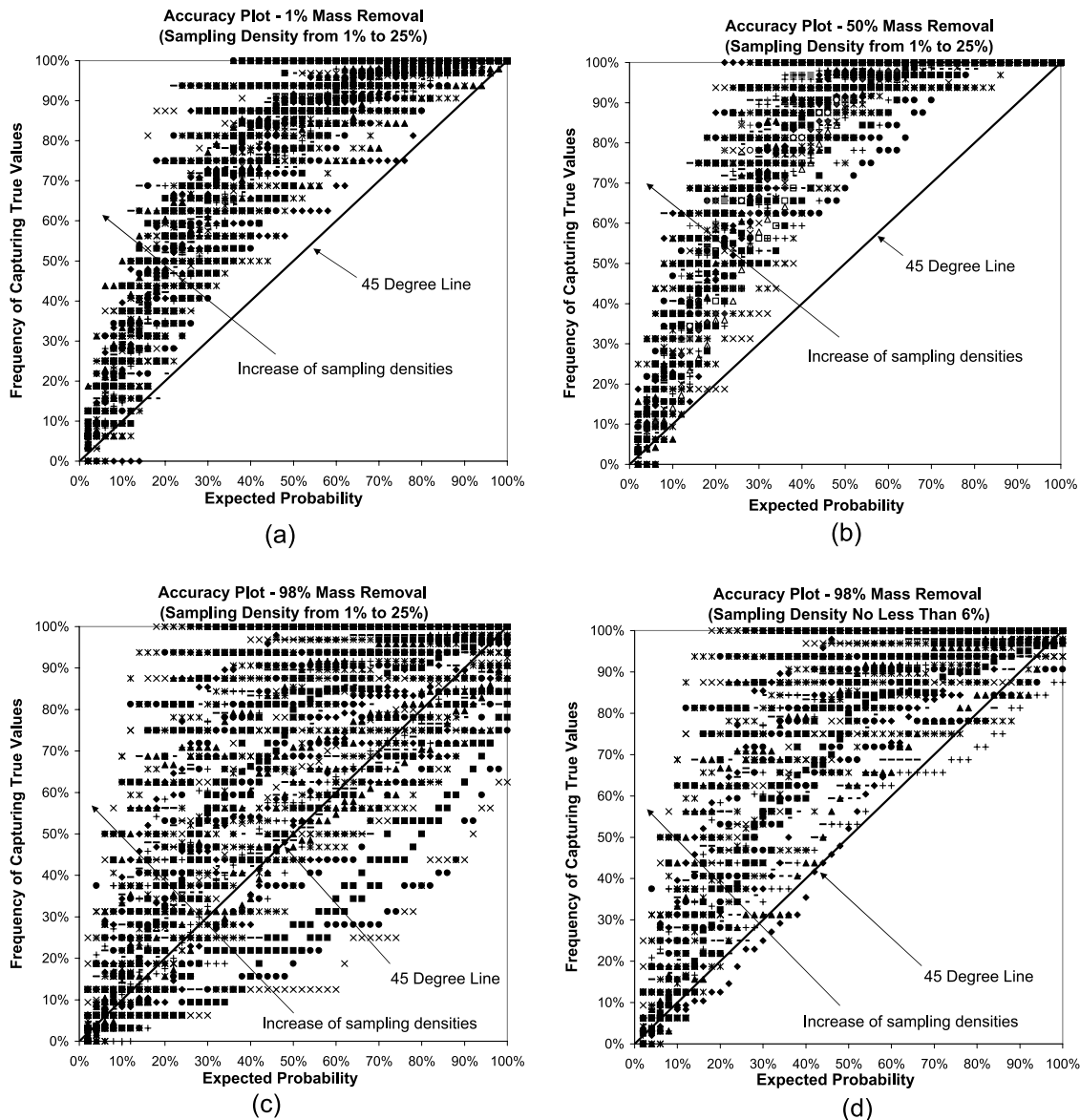
Figure 4. Grouped accuracy plots for the samples with a screen length of 0.3048 m (different symbols represent different sampling densities). (a) 1% mass removal cases (sampling density from 1 to 25%); (b) 50% mass removal cases (sampling density from 1 to 25%); (c) 98% mass removal cases (sampling density from 1 to 25%); (d) 98% mass removal cases with sampling density higher than 6%.

lowing two statistics: (1) magnitude of the normalized mean absolute error of prediction (MAEP), which is defined as the absolute value of the difference between the mean of the probability distribution and the true value, normalized by the true value (Figure 5), and (2) the width of the 95% confidence interval that is calculated as the difference between the 97.5% quantile and the 2.5% quantile (Figure 6). Figure 5 reveals that if the mean of the probability distribution is used as the mass discharge estimate, the absolute error of prediction does not exceed 20% of the true value at most sampling densities for the concentration distributions corresponding to 1 and 50% mass removal. Beyond a sampling density of about 4~5%, this error is below 10% of the true value. For these cases, Figure 5 also suggests that the MAEP is much less sensitive to the increase in the sampling density compared to the 98% mass removal case. This is likely because the
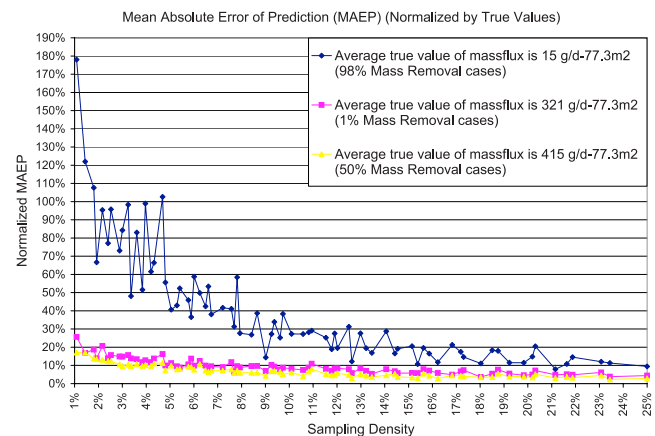

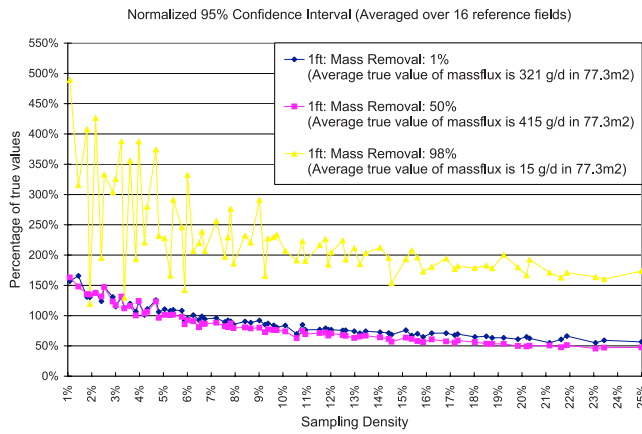**Figure 5.** Normalized mean absolute error of prediction, averaged over 16 reference fields.

Figure 6. The 95% confidence interval of the probability distributions averaged over 16 reference fields.

concentration distribution is more continuous, and its spatial structure is easier to recognize. For the 98% mass removal case, the MAEP improves quickly with more samples initially, and the improvement slows down at about a sampling density of 6∼7%, which corresponds to a prediction error of about 50% of the true value. This behavior is attributed to a slow improvement in the recognition of spatial structure (see section 4.5 for more detail). Figure 6 shows that the 95% confidence intervals are quite large at low sampling densities, and their width decrease (precision increases) slowly as the sampling density increases, indicating the need for reporting the uncertainty of mass discharge, in addition to the "best estimate". Note that the probability distribution of mass discharge is highly skewed (Figure 3c); therefore the uncertainty cannot be quantified simply by confidence intervals which are typically used for symmetric distributions. In addition to insufficient sampling density, the large confidence interval may also be caused by the regular sampling design, where a large portion of samples are placed in the near-zero concentration areas, which does not tend to contribute significantly to the mass discharge and therefore does not help to refine the mass discharge estimate. The impact of the sampling design will be investigated in another paper.

4.4. Impact of the Measurement Support

[30] All of the results above are for a screen length of 0.3048 m (1 ft). To explore the impact of screen length on the mass discharge uncertainty, a 0.6096-m (2-ft) vertical sampling scale was applied for the same 48 reference fields. Thus for the same sampling densities, the number of samples was reduced to half of the amount collected in the corresponding 1-ft sample case. The normalized MAEP and the 95% confidence interval were used to compare the accuracy and precision of the probability distributions modeled from the two different measurement supports. The comparison of the normalized MAEP is shown for equal sampling density in Figure 7.

[31] The comparison reveals that using a 2-ft sampling interval does not lead to more accurate results. For the same number of samples, there is no discernible difference in accuracy for the 1-ft and 2-ft screen samples (not shown). For the same sampling density (same sampling volume), Figure 7 reveals that the 2-ft screen samples generate less

accurate probability distributions for the mass discharge than the 1-ft screen samples. The reduction in the accuracy is especially remarkable for the concentration distribution associated with 98% source mass removal (Figure 2c).

[32] When comparing the 95% confidence intervals for the samples with 1 ft and 2 ft screens at the same sampling density, the probability distributions of 2 ft screen samples were slightly less precise. In terms of sample numbers, there was little difference in precision between the two measurement supports.

[33] This comparison suggests that increasing the measurement support (sampling volume) does not improve the quality of the probability distribution of mass discharge. This may be explained through the change in the sample statistics. Indeed, increasing the measurement support results in changes in the statistics of sample data such as the histogram (spread, asymmetry, and standard deviation) [*Isaaks and Srivastava*, 1989] and semivariogram (nugget, range, and sill) [*Journel and Huijbregts*, 1978]. Typically, larger measurement support reduces the spread of samples values (smoothing effect) while the histogram becomes more symmetric. The elimination of the spatial variance within the sampling volume causes a reduction of variance, which changes the structure of the experimental semivariogram. Most simulation algorithms, including *p*-field simulation, aim to reproduce sample histogram and semivariogram; thus, the changes in sample statistics directly affect the generated realizations.

[34] The magnitude of the impact of increasing measurement support depends upon the spatial structure of the data set. Spatially uncorrelated data are affected the most. As the spatial continuity increases, the statistics become more robust [*Isaaks and Srivastava*, 1989]. Note that two data sets with the same semivariogram may exhibit different degrees of robustness to measurement support as it is increased. The data set with larger spatial disorder (that is, extreme values are poorly connected) is affected more readily. Thus the optimal screen length will be site specific, related to the spatial pattern of the target attribute. It may be determined by analyzing the degree of heterogeneity (e.g., the variance), the correlation length, and the degree of spatial disorder (i.e., entropy, see *Journel and Deutsch*
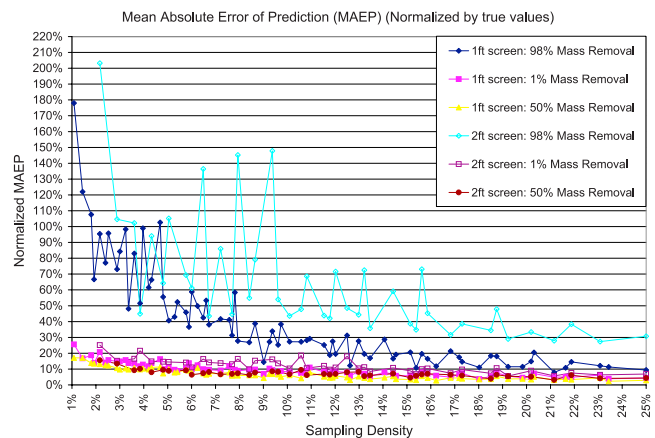

Figure 7. The comparison of the mean absolute error of prediction expressed in sampling densities for the 0.3048-m (1-ft) screen and the 0.6096-m (2-ft) screen samples.
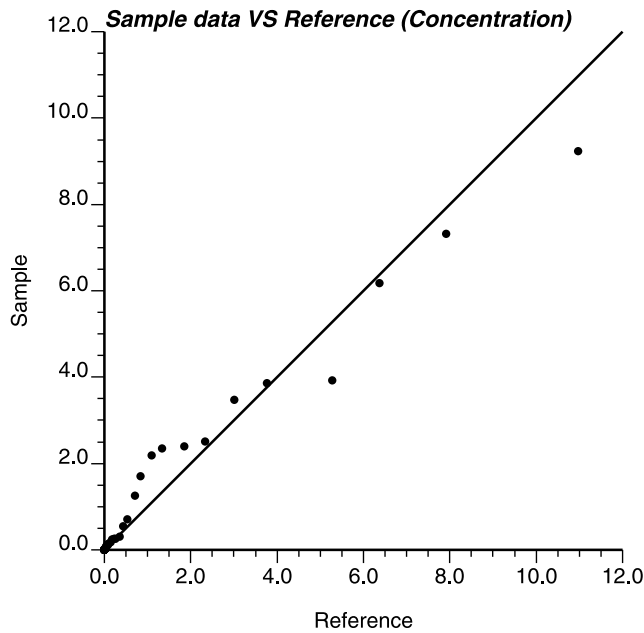
Figure 8. The Q-Q plot, comparing the histograms of 50 samples and the reference histogram (98% mass removal case).

[1993]). However, this topic is beyond the scope of this paper. The proposed approach cannot automatically choose an appropriate measurement support.

4.5. Impact of Sample Representativeness

[35] The quality of the output probability distributions is affected not only by the stochastic simulation algorithm but also by the characteristics of the data, such as their histogram and semivariogram model. Stochastic simulation aims to reproduce the univariate (histogram) and bivariate statistics (semivariogram) derived from the observations. When field data are sparse and not representative of the control plane properties, any algorithm will likely fail to provide accurate and precise models of uncertainty. The earlier results were obtained using relatively homogeneous reference hydraulic conductivity fields. Thus for this scenario, the problem of sample representativeness is, to a large extent, caused by the concentration field. In this section, the histograms and semivariograms of the concentration samples (1-ft screen) are analyzed to explore this problem.

[36] The difference between two histograms can be illustrated by a Q-Q plot [Deutsch and Journel, 1998], which can be viewed as a scatterplot of the series of p quantiles of the two data sets (Figure 8). If the points fall along the 45° line, the two histograms are similar. The Q-Q plot in Figure 8 compares the histogram of a 50-data sample set to the reference histogram (50 p quantiles). The plot indicates that this sample histogram (98% mass removal case) is not representative. For small cumulative frequencies, the sample cdf has larger quantiles, and when the probability gets higher, the corresponding quantiles for the sample become lower than the reference. This means that both the low and high values are undersampled (sample histogram is more symmetric than the reference).

[37] The average histogram deviation can be quantified by calculating the averaged difference of p quantiles be-

tween the sample histogram and the reference histogram [Goovaerts, 1999]:

$$D_h = \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} \left| q_k^{\text{sample},n} - q_k^{\text{ref},n} \right|$$

[38] Here N is the number of reference fields, K is the number of quantiles used for the comparison, $q_k^{\text{sample},n} = F_{\text{sample},n}^{-1}(p_k)$ is the p_k quantile of the sample histogram, and $q_k^{\text{ref},n}$ is the corresponding quantile of the reference histogram.

[39] Figure 9 plots histogram deviation as a function of sampling density (98% mass removal case). As more observations are collected, the sample histograms on average gradually approach the reference histogram. Unfortunately, the literature offers no guidance on the acceptable level of this deviation. The Q-Q plot in Figure 8 suggests that even with 50 samples (sampling density of 6%), the sample histogram may still not be representative. A rough estimate of the acceptable deviation can be obtained from Figure 9 by examining the average histogram deviation at a sampling density of 6% (the histogram is not representative, as shown in Figure 8). The average histogram deviation associated with this sampling density is about 2% of the range of variation, suggesting that the acceptable level of the deviation could be even lower than 2%. Moreover, the deviation decreases at a slower rate with sampling densities beyond 6%. All of these findings based upon Figure 9 suggest that most field applications cannot provide sufficient sample data to infer a representative histogram.

[40] Similarly, the departure between the sample semivariogram and the reference semivariogram can be calculated as:

$$D_s = \frac{1}{NL} \sum_{n=1}^{N} \sum_{l=1}^{L} \left| \gamma_l^{\text{sample},n} - \gamma_l^{\text{ref},n} \right|$$

[41] Here N is the number of reference fields, L is the number of lags used for the calculation of experimental semivariograms, $\gamma_i^{\text{ref},n}$ is the reference semivariogram value at lag i for the nth reference field, and $\gamma_i^{\text{sample},n}$ is the corresponding sample semivariogram value at the same lag.
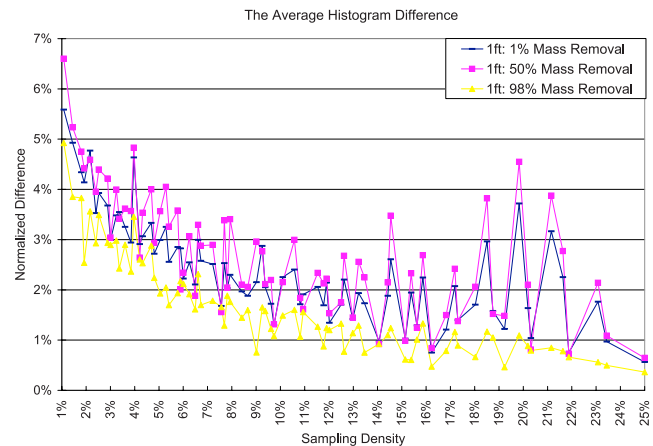

Figure 9. Average deviation of sample histogram from the reference histogram (normalized by the range of variation).
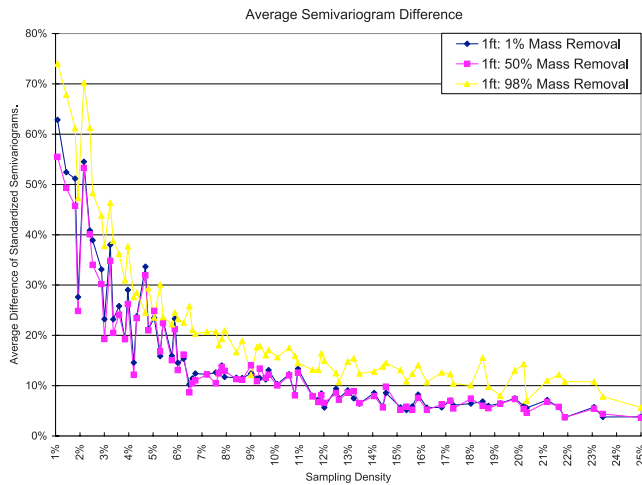
Figure 10. Average deviation of sample semivariogram from the reference semivariogram (standardized).

Figure 12. Semivariogram comparisons (50% mass removal). Solid line, semivariogram of the hot spot area only; dash line, semivariogram of all the data.

[42] The calculated semivariogram deviation is depicted in Figure 10. This figure shows that for the best situation (at a sampling density of 25%), the sample semivariogram still deviates by about 5% of the variance. Figure 11 presents an example for which the experimental semivariogram is not representative at a sampling density of 6% for the 98% mass removal case. Furthermore, Figure 10 shows that the deviation of the semivariogram drops quickly as sampling density increases, but that the deviation decreases more gradually and exhibits a long tail when the sampling density exceeds about 6~7%. This suggests that beyond this density, the detection of the spatial structure improves little with more samples. This 6~7% density is consistent with the results of the accuracy plots in section 4.3. Figure 10 also indicates that it is more difficult to capture the spatial variability for the control planes that are associated with high levels of source mass removal, which is the type of
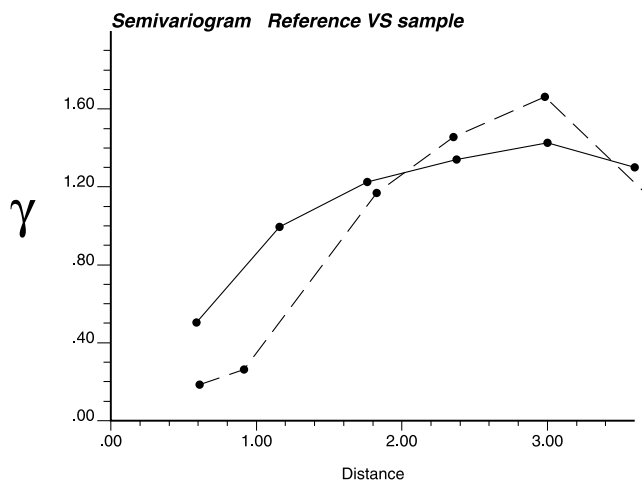
concentration distribution that has few hot spots and large areas of near-zero concentration (Figure 2c).

4.6. Impact of Nonstationary Concentration Field

[43] When spatial variables are modeled by random functions (RF), second-order stationarity or at least the intrinsic hypothesis [*Journel and Huijbregts*, 1978] is required for geostatistical applications. The intrinsic hypothesis ensures the RF increments have finite variance and are independent of location. The synthetic hydraulic conductivity fields in this work are stationary (because of the construction process), while the resulting transect concentration fields have different degrees of nonstationarity, related to the percentage of mass removal. For example, the concentration fields associated with low levels of mass removal (Figures 2a and 2b) have higher degrees of nonstationarity in comparison to those associated with high levels of mass removal (Figure 2c). It is common in stochastic hydrology to detrend nonstationary permeability fields whenever geological information supports the use of a specific mathematical form for the deterministic trend model [e.g., *Rehfeldt et al.*, 1992; *Indelman and Rubin*, 1995]. In this paper, interpolation was conducted using ordinary indicator kriging which implicitly accounts for global trends by estimating the local mean within each search window and limits the assumption of stationarity to that window [*Goovaerts*, 1997]. More complicated trend modeling was not considered here because (1) there is no other information to justify a reasonable deterministic trend model, and a model derived purely from a limited number of samples may bring in large uncertainty; and (2) theoretical and experimental studies showed that compared to ordinary kriging, trend modeling using kriging with a trend (universal kriging) has an impact on kriging results only for the extrapolation situation [*Journel and Rossi*, 1989; *Zimmerman et al.*, 1999; *Reed et al.*, 2004]. However, in that case, the sample data alone cannot justify the choice of a trend model [*Deutsch and Journel*, 1998].

[44] For the concentration distributions with large contaminated areas (1 and 50% mass removal case), Figure 12 shows that the semivariogram computed with only the hot

Figure 11. The comparison of the semivariograms of 50 samples (0.3048-m screen) and the reference field (98% mass removal). Solid line, reference semivariogram; dash line, sample semivariogram.
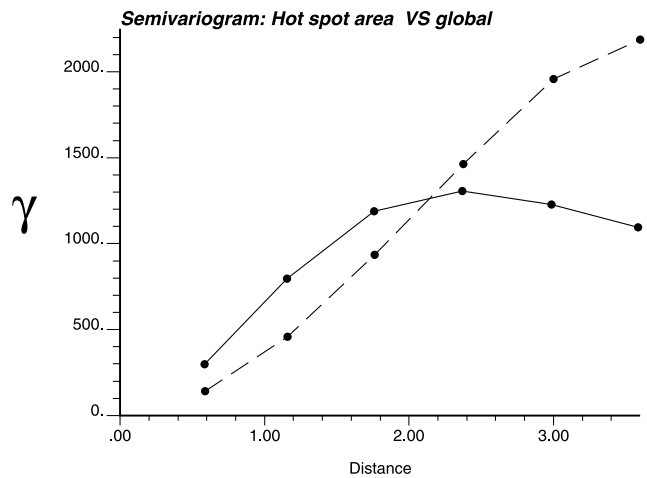
spot data is quite different from the semivariogram computed with all of the sample data. The nonstationarity of the variance threatens the intrinsic hypothesis for the geostatistics analysis. To solve this problem, kriging within strata [e.g., *Goovaerts*, 1997] has been suggested. However, it is difficult to classify the sample data into "hot spot strata" and "nonhot spot strata" in this analysis because the boundary of the hot spot is usually unknown. Therefore one-stage sampling strategy cannot address the difficulty of nonstationarity encountered here. The above discussion suggests that a stage-sampling strategy should be explored for this problem.

5. Conclusion

[45] The proposed geostatistical approach has been demonstrated to estimate the contaminant mass discharge and quantify the associated uncertainty from multilevel transect measurements of hydraulic conductivity and contaminant concentration. The approach does not require that measurements are collocated, and the measurement supports can be different for the two attributes. The simulation-based approach provides a probability distribution of mass discharge, which can serve for risk-based source-plume remediation strategies. Multiple realizations of the spatial distribution of mass flux values can be used for decision making, for example for mapping the probability that the local mass flux exceeds a given threshold.

[46] Application of the approach on 48 numerically simulated transects suggests that, for all of the sampling densities used (1∼25%), the approach provides an accurate estimate of mass discharge and a conservative model of mass discharge uncertainty for control planes associated with low levels of source mass removal. For high levels of mass removal, a minimum sampling density of 6∼7% was required for an accurate model of uncertainty (regular sampling pattern). Note that the required sampling density is typically not achieved in field applications (see examples in section 4.3); however, concentration distributions associated with high levels of mass removal are more likely to be encountered in the field. Thus most field applications to date may not have been based upon a sample size sufficient to accurately quantify the uncertainty of mass discharge, and the estimated mass discharge may have large errors.

[47] Comparison of results for different screen lengths suggests that longer screen lengths may not lead to better modeling of the uncertainty of mass discharge.

[48] These conclusions regarding sampling density are based on the analysis of synthetic transects, where the hydraulic conductivity field is relatively homogeneous. The model is applicable to more heterogeneous hydraulic conductivity fields; however, the required sampling density could be even larger under such circumstances and is expected to more closely correspond to the 98% removal case. A more efficient sampling strategy (rather than one-stage regular sampling) will be investigated in future studies to reduce the required sampling density.

providing synthetic transects as the reference fields for the geostatistical analysis.

References

Abriola, L. M., C. D. Drummond, E. J. Hahn, K. F. Hayes, T. C. G. Kibbey, L. D. Lemke, K. D. Pennell, E. A. Petrovskis, C. A. Ramsburg, and K. M. Rathfelder (2005), Pilot-scale demonstration of surfactant-enhanced PCE solubilization at the Bachman road site. 1. Site characterization and test design, *Environ. Sci. Technol.*, 39, 1778−1790.

Almeida, A. S., and A. G. Journel (1994), Joint simulation of multiple-variables with a Markov-type coregionalization model, *Math. Geol.*, 26, 565−588.

Basel, M. D., and C. H. Nelson (2000), Overview of in-situ chemical oxidation: Status and lessons learned, in *Second International Conference on Remediation of Chlorinated and Racalcitrant Compounds; Monterey, CA; USA; 22−25 May 2000*, Monterey, CA, USA, 117−124, Battelle Press.

Beck, M. B. (1987), Water-quality modeling—A review of the analysis of uncertainty, *Water Resour. Res.*, 23, 1393−1442.

Bockelmann, A., T. Ptak, and G. Teutsch (2001), An analytical quantification of mass fluxes and natural attenuation rate constants at a former gasworks site; Natural attenuation of organic pollutants in groundwater, *J. Contam. Hydrol.*, 53, 429−453.

Bockelmann, A., D. Zamfirescu, T. Ptak, P. Grathwohl, and G. Teutsch (2003), Quantification of mass fluxes and natural attenuation rates at an industrial site with a limited monitoring network; a case study, *J. Contam. Hydrol.*, 60, 97−121.

Borden, R. C., R. A. Daniel, L. E. LeBrun IV, and C. W. Davis (1997), Intrinsic biodegradation of MTBE and BTEX in a gasoline-contaminated aquifer, *Water Resour. Res.*, 33, 1105−1115.

Christ, J. A., L. D. Lemke, and L. M. Abriola (2005), Comparison of two-dimensional and three-dimensional simulations of dense nonaqueous phase liquids (DNAPLs): Migration and entrapment in a nonuniform permeability field, *Water Resour. Res.*, 41, W01007, doi:10.1029/2004WR003239.

Christ, J. A., C. A. Ramsburg, K. D. Pennell, and L. M. Abriola (2006), Estimating mass discharge from DNAPL source zones using upscaled mass transfer coefficients: An evaluation using multiphase numerical simulations, *Water Resour. Res.*, 42, W11420, doi:10.1029/2006WR004886.

Desbarats, A. J., and R. Dimitrakopoulos (2004), Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors, *Math. Geol.*, 32, 919−942.

Deutsch, C. V. (1997), Direct assessment of local accuracy and precision, *Geostatistics Wollongong '96*, 1 and 2(8), 115−125.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed., Oxford University Press, New York.

Dungan, J. (1998), Spatial prediction of vegetation quantities using ground and image data, *Int. J. Remote Sens.*, 19, 267−285.

Einarson, M. D., and D. M. Mackay (2001), Predicting impacts of ground-water contamination, *Environ. Sci. Technol.*, 35, 66A−73A.

Feenstra, S., J. A. Cherry, B. L. Parker, J. F. Pankow, and J. A. Cherry (1996), Conceptual models for the behavior of dense non-aqueous phase liquids (DNAPLs) in the subsurface. Dense chlorinated solvents and other DNAPLs in groundwater; history, behavior, and remediation, Waterloo Press, Portland, OR, United States, USA.

Freeze, R. A., and D. B. McWhorter (1997), A framework for assessing risk reduction due to DNAPL mass removal from low-permeability soils, *J. Groundwater*, 35, 111−123.

Froidevaux, R. (1993), Probability field simulation, in *Fourth International Geostatistics Congress, Troia, Portugal, Sept. 1992*, 73−84, Springer, New York.

Grimmett, G. R., and D. R. Stirzaker (2001), *Probability and Random Processes*, 3rd. ed., 596 pp., Oxford Univ. Press, New York.

Gómez-Hernández, J. J., and A. G. Journel (1993), Joint sequential simulation of multigaussian fields, *Geostat. Troia 92*, 1 and 2(5), 85−94.

Gómez-Hernández, J. J., and R. M. Srivastava (1990), ISIM3D—An ANSI-C 3-dimensional multiple indicator conditional simulation program, *Comput. Geosci.*, 16, 395−440.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, vol. 1. Oxford Univ. Press, New York.

Goovaerts, P. (1999), Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties, *Stochastic Environ. Res. Risk Assess.*, 13, 161−182.

Goovaerts, P. (2001), Geostatistical modelling of uncertainty in soil science, *Geoderma*, 103, 3−26.

Goovaerts, P. (2002), Geostatistical modelling of spatial uncertainty using *p*-field simulation with conditional probability fields, *Int. J. Geogr. Inf. Sci.*, 16, 167–178.

Goulard, M., and M. Voltz (1992), Linear coregionalization model—Tools for estimation and choice of cross-variogram matrix, *Math. Geol.*, 24, 269–286.

Gutjahr, A., B. Bullard, S. Hatch, and L. Hughson (1994), Joint conditional simulations and the spectral approach for flow modeling, *Stochastic Environ. Res. Risk Assess.*, 8, 79–108.

Hatfield, K., M. D. Annable, S. Kuhn, P. S. C. Rao, and T. Campbell (2002), A new method for quantifying contaminant flux at hazardous waste sites; Groundwater quality; natural and enhanced restoration of groundwater pollution., in *Groundwater Quality; Natural and Enhanced Restoration of Groundwater Pollution*, pp. 25–31, International Association of Hydrological Sciences, International (III), Sheffield, United Kingdom.

Hatfield, K., M. Annable, J. Cho, P. S. Rao, and H. Klammler (2004), A direct passive method for measuring water and contaminant fluxes in porous media, *J. Contam. Hydrol.*, 75, 155–181.

Heuvelink, G. B. M., and E. J. Pebesma (1999), Spatial aggregation and soil process modelling; Pedometrics '97, *Geoderma*, 89, 47–65.

Holder, T. H., G. P. T. Teutsch, and R. Schwarz (1998), A new approach for source zone characterization; The Neckar Valleystudy, in *Groundwater Quality: Remediation and Protection*, pp. 49–55, International Association of Hydrological Sciences, [Louvain], International (III), Tubingen, Federal Republic of Germany.

Indelman, P., and Y. Rubin (1995), Flow in heterogeneous media displaying a linear trend in the log conductivity, *Water Resour. Res.*, 31, 1257–1265.

Interstate Technology and Regulatory Council (ITRC) (2004), Strategies for Monitoring the Performance of DNAPL Source Zone Remedies.

Isaaks, E. H., and R. M. Srivastava (1989), *An Introduction to Applied Geostatistics*, 561 pp, Oxford Univ. Press, New York.

Jarsjo, J., M. Bayer-Raich, and T. Ptak (2005), Monitoring groundwater contamination and delineating source zones at industrial sites: Uncertainty analyses using integral pumping tests, *J. Contam. Hydrol.*, 79, 107–134.

Journel, A. G. (1995), Probability fields: Another look and a proof: Unpublished report of Stanford Center for Reservoir Forecasting, Stanford, CA.

Journel, A. G., and C. V. Deutsch (1993), Entropy and spatial disorder, *Math. Geol.*, 25, 329–355.

Journel, A. G., and C. J. Huijbregts (1978), *Mining Geostatistics*, Elsevier, New York.

Journel, A. G., and M. E. Rossi (1989), When do we need a trend model in Kriging, *Math. Geol.*, 21, 715–739.

Kao, C. M., and Y. S. Wang (2001), Field investigation of the natural attenuation and intrinsic biodegradation rates at an underground storage tank site, *Environ. Geol.*, 40, 622–631.

King, M. W. G., J. F. Barker, J. F. Devlin, and B. J. Butler (1999), Migration and natural fate of a coal tar creosote plume; 2, Mass balance and biodegradation indicators; Transport and fate of coal-tar compounds in the subsurface, *J. Contam. Hydrol.*, 39, 281–307.

Kübert, M., and M. Finkel (2006), Contaminant mass discharge estimation in groundwater based on multi-level point measurements: A numerical evaluation of expected errors, *J. Contam. Hydrol.*, 84, 55–80.

Lemke, L. D., and L. M. Abriola (2003), Predicting DNAPL entrapment and recovery; the influence of hydraulic property correlation, *Stochastic Environ. Res. Risk Assess.*, 17, 408–418.

Lemke, L. D., L. M. Abriola, and P. Goovaerts (2004), Dense nonaqueous phase liquid (DNAPL) source zone characterization: Influence of hydraulic property correlation on predictions of DNAPL infiltration and entrapment, *Water Resour.*, 40.

Luster, G. R. (1985), Raw materials for Portland cement: Applications of conditional simulation of coregionalization, doctoral dissertation, Stanford University, Stanford, CA.

Mackay, D. M., and J. A. Cherry (1989), Groundwater contamination; limitations of pump-and-treat remediation, *Environ. Sci. Technol.*, 23, 630–636.

Mao, S. G., and A. G. Journel (1999), Conditional 3D simulation of lithofacies with 2D seismic data, *Comput. Geosci.*, 25, 845–862.

Painter, S. (2001), Flexible scaling model for use in random field simulation of hydraulic conductivity, *Water Resour. Res.*, 37, 1155–1163.

Parker, J. C., and E. Park (2004), Modeling field-scale dense nonaqueous phase liquid dissolution kinetics in heterogeneous aquifers, *Water Resour.*, 40.

Pennell, K. D., and L. M. Abriola (1997), Surfactant enhanced aquifer remediation: Fundamental processes and practical application, in *Bioremediation: Principles and Practice*, edited by S. K. Sikdar and R. L. Irvine, pp. 693–750, Technomic Publ., Lancaster, PA.

Rao, P. S. C., J. W. Jawitz, C. G. Enfield, R. W. Falta Jr., M. D. Annable, and A. L. Wood (2002), Technology integration for contaminated site remediation; clean-up goals and performance criteria; Groundwater quality; natural and enhanced restoration of groundwater pollution, *IAHS-AISH Publ.*, 275, 571–578.

Reddy, K. R., S. Kosgi, and J. Zhou (1995), A review of in-situ air sparging for the remediation of VOC-contaminated saturated soils and groundwater, *Hazard. Waste Hazard. Mater.*, 12, 97–118.

Reed, P. M., T. R. Ellsworth, and B. S. Minsker (2004), Spatial interpolation methods for nonstationary plume data, *J. Groundwater*, 42, 190–202.

Rehfeldt, K. R., J. M. Boggs, and L. W. Gelhar (1992), Field-study of dispersion in a heterogeneous aquifer. 3. Geostatistical analysis of hydraulic conductivity, *Water Resour. Res.*, 28, 3309–3324.

Robin, M. J. L., A. L. Gutjahr, E. A. Sudicky, and J. L. Wilson (1993), Cross-correlated random-field generation with the direct fourier-transform method, *Water Resour. Res.*, 29, 2385–2397.

Ruan, F., and D. McLaughlin (1998), An efficient multivariate random field generator using the fast Fourier transform, *Adv. Water Resour.*, 21, 385–399.

Saito, H., and P. Goovaerts (2002), Accounting for measurement error in uncertainty modeling and decision-making using indicator kriging and p-field simulation: application to a dioxin contaminated site, *Environmetrics*, 13, 555–567.

Sale, T. C., and D. B. McWhorter (2001), Steady state mass transfer from single-component dense nonaqueous phase liquids in uniform flow fields, *Water Resour. Res.*, 37, 393–404.

Schwarz, R., T. Ptak, T. H. Holder, and G. Teutsch (1998), Groundwater risk assessment at contaminated sites: A new investigation approach, in *Groundwater Quality: Remediation and Protection*, 68–71, International Association of Hydrological Sciences [Louvain], International (III), Tubingen, Federal Republic of Germany.

Semprini, L., P. K. Kitanidis, D. H. Kampbell, and J. T. Wilson (1995), Anaerobic transformation of chlorinated aliphatic hydrocarbons in a sand aquifer based on spatial chemical distributions, *Water Resour. Res.*, 31, 1051–1062.

Soga, K., J. W. E. Page, and T. H. Illangasekare (2004), A review of NAPL source zone remediation efficiency and the mass flux approach, *J. Hazard. Mater.*, 110, 13–27.

Srivastava, R. M. (1992), Reservoir characterization with probability field simulation, in *SPE Annual Conference and Exhibition, Washington, D. C.*, number 24753, Washington, D. C., 927–938, Society of Petroleum.

Stroo, H. F., M. Unger, C. H. Ward, M. C. Kavanaugh, C. Vogel, A. Leeson, J. A. Marqusee, and B. P. Smith (2003), Remediating chlorinated solvent source zones, *Environ. Sci. Technol.*, 37, 224A–230A.

Udell, K. S. (1997), Thermally enhanced removal of liquid hydrocarbon contaminants from soils and groundwater, in *Subsurface Restoration*, edited by C. H. Ward, J. A. Cherry, and M. R. Scalf, pp. 251–270, Sleeping Bear Press, Chelsea, MI.

U.S. Environmental Protection Agency (EPA) (2003), *DNAPL Remediation Challenge: Is There a Case for Source Depletion.* 68 Environmental Pollution & Control. EPA600R03143; PB2005110799, 134 pp.

U.S. Environmental Protection Agency (EPA) (2004), Site Characterization Technologies for DNAPL Investigations. 77G Nuclear Science and Technology: Radioactive Wastes and Radioactivity; 68 Environmental Pollution and Control. EPA542R04017; PB2005100347, 170 pp.

U.S. National Research Council (NRC) (1994), Alternatives for ground water cleanup, 22, *Environmental Geology.*

U.S. National Research Council (NRC) (2004), Contaminants in the subsurface: Source zone assessment and remediation, *Water Science and Technology Board.*

Vargas-Guzman, J. A. (2004), Fast modeling of cross-covariances in the LMC: A tool for data integration, *Stochastic Environ. Res. Risk Assess.*, 18, 91–99.

Verly, G. W. (1993), Sequential Gaussian cosimulation—A simulation method integrating several types of information, *Geostatistics Troia 92*, 1 and 2(5), 543–554.

Wilson, J. T., J. S. Cho, B. H. Wilson, and J. A. Vardy (2000), Natural attenuation of MTBE in the subsurface under methanogenic conditions, *U. S. Environmental Protection Agency.*

Yao, T. T., and A. Chopra (2000), Integration of seismic attribute map into 3D facies modeling, *J. Pet. Sci. Eng.*, 27, 69–84.

Yao, T. T., and A. G. Journel (1998), Automatic modeling of (cross) covariance tables using fast Fourier transform, *Math. Geol.*, *30*, 589–615.

Zimmerman, D., C. Pavlik, A. Ruggles, and M. P. Armstrong (1999), An experimental comparison of ordinary and universal kriging and inverse distance weighting, *Math. Geol.*, *31*, 375–390.

Zeru, A., and G. Schafer (2005), Analysis of groundwater contamination using concentration-time series recorded during an integral pumping test: Bias introduced by strong concentration gradients within the plume, *J. Contam. Hydrol.*, *81*, 106–124.

————————————

L. M. Abriola, Department of Civil and Environmental Engineering, Tufts University, 105 Anderson Hall, Medford, MA 02155, USA. (linda.abriola@tufts.edu)

P. Goovaerts, BioMedware, Inc., Ann Arbor, MI, USA.

K. B. Li, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA.