

## Contaminant discharge and uncertainty estimates from passive flux meter measurements

Harald Klammler,<sup>1,2,3</sup> Kirk Hatfield,<sup>2,3</sup> Joana Angélica Guimarães da Luz,<sup>1</sup> Michael D. Annable,<sup>3,4</sup> Mark Newman,<sup>2,3</sup> Jaehyun Cho,<sup>3,4</sup> Aaron Peacock,<sup>5</sup> Valerie Stucker,<sup>6</sup> James Ranville,<sup>6</sup> Steven A. Cabaniss,<sup>7</sup> and P. S. C. Rao<sup>8</sup>

Received 9 February 2011; revised 28 November 2011; accepted 7 January 2012; published 14 February 2012.

[1] The passive flux meter (PFM) measures local cumulative water and contaminant fluxes at an observation well. Conditional stochastic simulation accounting for both spatial correlation and data skewness is introduced to interpret passive flux meter observations in terms of probability distributions of discharges across control planes (transects) of wells. An estimator of the effective number of independent data is defined and applied in the development of two significantly simpler approximate methods for estimating discharge distributions. One method uses a transformation of the  $t$  statistic to account for data skewness and the other method is closely related to the classic bootstrap. The approaches are demonstrated with passive flux meter data from two field sites (a trichloroethylene [TCE] plume at Ft. Lewis, WA, and a uranium plume at Rifle, CO). All methods require that the flux heterogeneity is sufficiently represented by the data and maximum differences in discharge quantile estimates between methods are  $\sim 7\%$ .

**Citation:** Klammler, H., et al. (2012), Contaminant discharge and uncertainty estimates from passive flux meter measurements, *Water Resour. Res.*, 48, W02512, doi:10.1029/2011WR010535.

### 1. Introduction

[2] Subsurface contaminants are a well-known threat to groundwater at source locations and as contaminant plumes that may emerge. Identification, monitoring, risk assessment (including transport modeling), and the design of remediation strategies are fundamental engineering tasks related to this problem. Over the past decades these tasks and have received significant research efforts from the theoretical to the field level [*U.S. Environmental Protection Agency (EPA)*, 1998, 2004; *Einarson and Mackay*, 2001; *Rao et al.*, 2002; *Stroo et al.*, 2003]. Historically, estimates of contaminant source mass remaining and/or contaminant concentration in a plume have been used to characterize contaminated sites. However, more recently, contaminant mass discharge from a source zone at a given control plane (transect) has been recognized as more useful, since it

directly quantifies the contaminant mass released from a source per unit time, which may impact a potential receptor [*Einarson and Mackay*, 2001; *Interstate Technology and Regulatory Council (ITRC)*, 2010].

[3] Current field measurements supporting estimation of contaminant mass discharge employ one of three basic approaches: integral pump tests (IPT) [*Teutsch et al.*, 2000; *Bockelmann et al.*, 2001, 2003], multilevel sampling (MLS) [*Borden et al.*, 1997; *Einarson and Mackay*, 2001], and passive flux meter measurements (PFM) [*Hatfield et al.*, 2004; *Annable et al.*, 2005]. IPT uses inverse modeling on monitored contaminant concentrations at one or more pumping wells. This approach has the advantage of interrogating a large and continuous portion of the aquifer; however, it requires prior knowledge (or separate measurement/assumption) of certain aquifer properties. In addition, significant alterations of groundwater flow and contaminant transport regimes are induced, such that contaminated groundwater is often extracted for surface disposal. MLS uses separate measurements of local contaminant concentrations, hydraulic conductivities, and the hydraulic gradient at a control plane. Estimates of contaminant mass discharges require spatial interpolation of concentrations and conductivities as well as assumptions about the hydraulic gradient for the calculation and summation of local mass fluxes. While these are potential sources of uncertainty, less hydrogeological knowledge or assumptions are required than for IPT and the natural flow field remains practically undisturbed. The PFM provide direct measurements of local cumulative (i.e., time integrated or averaged) contaminant mass fluxes as depth profiles along monitoring wells, without necessarily requiring hydraulic conductivity and gradient data. Besides inherent measurement errors in

<sup>1</sup>Department of Environmental Sciences and Sustainable Development, Federal University of Bahia, Barreiras, Bahia, Brazil.

<sup>2</sup>Department of Civil and Coastal Engineering, University of Florida, Gainesville, Florida, USA.

<sup>3</sup>Inter-Disciplinary Program in Hydrologic Sciences, University of Florida, Gainesville, Florida, USA.

<sup>4</sup>Department of Environmental Engineering Sciences, University of Florida, Gainesville, Florida, USA.

<sup>5</sup>Microbial Insights, Inc., Rockford, Tennessee, USA.

<sup>6</sup>Department of Chemistry and Geochemistry, Colorado School of Mines, Golden, Colorado, USA.

<sup>7</sup>Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico, USA.

<sup>8</sup>Department of Civil Engineering, Purdue University, West Lafayette, Indiana, USA.

local fluxes, spatial interpolation of fluxes over unsampled portions of a transect is the main source of uncertainty in PFM-based mass discharge estimates.

[4] None of the above approaches are capable of delivering error-free estimates of mass discharges. As a result, *Jarsjö et al.* [2005] perform uncertainty analysis for an IPT, while *Li et al.* [2007], *Schwede and Cirpka* [2010], *Trolborg et al.* [2010], and *Cai et al.* [2011] use geostatistical conditional stochastic simulation approaches for quantification of uncertainty in mass discharge estimates from MLS measurements. To reduce estimation uncertainty, *Li and Abriola* [2009] present a multistage and multicriteria strategy for multilevel sampling. Comparative theoretical and field studies of different methods are available by *Kübert and Finkel* [2006], *Herold et al.* [2008], *Goltz et al.* [2009], and *Belland-Pelletier et al.* [2011]. Taken together, these studies recommend pump test methods in highly heterogeneous settings where “point” sample methods suffer from large interpolation uncertainties, and identify PFM measurements as less error prone than MLS methods. The latter is related to the fact that PFMs directly measure local mass fluxes along continuous vertical sampling lines rather than concentrations (to be multiplied by collocated and simultaneously measured or estimated water fluxes) at a finite number of MLS points.

[5] The objective of the present study is to investigate methods for estimating probability distributions of water and contaminant mass discharges from PFM measurements. If the empirical distribution of the sampled PFM fluxes is representative of the transect (i.e., if it accurately indicates the frequency of occurrence of each sampled and unsampled flux value in the transect), then the arithmetic sample mean gives an unbiased estimate of the spatial mean of all fluxes in the transect. In the presence of irregular sampling patterns, this representativeness can be achieved by application of a declustering technique [*Goovaerts*, 1997], which assigns smaller weights to more closely spaced data (“clusters”) and larger weights to more largely spaced data; hence, the name “declustering.” Classical and bootstrap statistics provide tools for quantifying uncertainty about mean estimates of sample data; however, the inherent assumption of independent and identically distributed (i.i.d.) data is commonly violated by the presence of spatial correlation [*Journel*, 1994] between observed fluxes.

[6] In contrast, the principles of kriging are founded on correlated random variables. However, the definition of probability distributions for estimation errors requires either the assumption of multivariate Gaussianity [*Goovaerts*, 1997] or data transformations, e.g., indicator kriging or as in the work of *Kitanidis and Shen* [1996]. The typically large ranges of variability (e.g., coefficients of variation near or above one) of contaminant mass fluxes, in combination with the physical requirement of non-negativity, lead to strong deviations from Gaussianity characterized by positive skewnesses. Data transformations required to obtain Gaussian distributions are necessarily nonlinear leading to biased estimates of mean transect fluxes when back-transformed after averaging. Methods to determine probability distributions of mean transect fluxes in the presence of irregular sampling patterns, spatial correlation, and non-Gaussianity are limited. Geostatistics offers the conditional

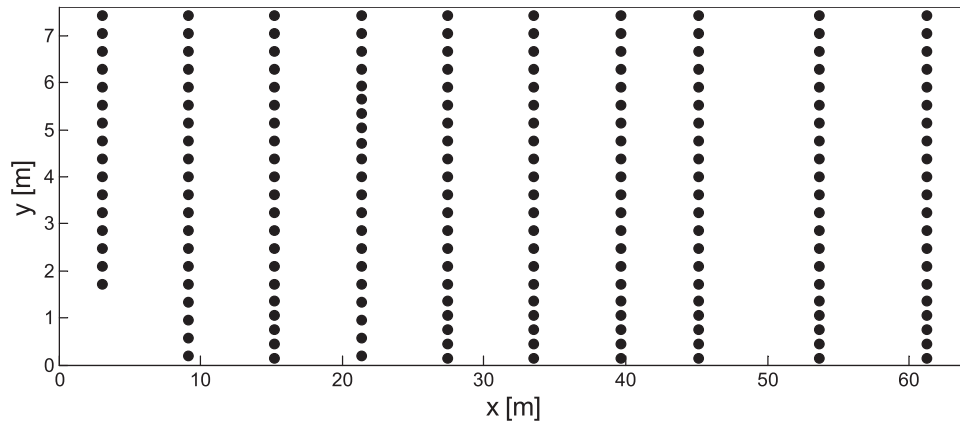
stochastic simulation approaches cited above as well as a method using resampling from conditional stochastic simulation [*Journel*, 1994]. Moreover, *Zhu and Morgan* [2004] and *Chen and Shao* [1999] apply the block bootstrap method for inference of distributions of spatial means in the case of data correlation. However, no explicit guidelines for the choice of bootstrap block size and shape in dependence of sample number, pattern (especially irregular and sparse patterns), and correlation structure are presently available, neither may the actual shape and size of the transect be accounted for.

[7] Here we present and demonstrate a conditional stochastic simulation approach and two simpler approximate methods for determining probability distributions of contaminant mass discharges from PFM field data. Section 2 focuses on stochastic simulation: section 2.1 uses the unconditional stochastic simulation paradigm to introduce the concepts of randomization and ordinary block kriging as well as fundamental assumptions and the required variables. This includes the derivation of a so-called “weighted data dispersion variance.” Section 2.2 is a step-by-step outline of the conditional simulation approach taken in the case studies and emphasizes guidelines for reliable variogram inference as well as for the validation of results. A parameter termed “missing variance”  $\Delta\sigma^2$  [-] is proposed to quantify the representativeness of PFM flux data with respect to local flux heterogeneity over a transect. Section 3 is dedicated to the introduction and discussion of an “effective number of independent data”  $n_e$  [-], which forms the basis for the development of the two approximate methods in section 4. These approximate methods avoid performing stochastic simulation and may represent more practical tools for field engineers. Field data from a TCE plume at Ft. Lewis, WA, and from a uranium plume at Rifle, CO, are used in section 5 for demonstration and a discussion of the results including a validation of the approximate methods against conditional stochastic simulation. Also, an example of how to quantify expected benefits in terms of uncertainty reduction due to additional PFM sampling is presented. Finally, section 6 contains an evaluation of the assumptions initially taken based on the case study results.

## 2. Stochastic Simulation Approach

### 2.1. Fundamental Concepts, Assumptions, and Variables

[8] Without actually generating any realizations, we use the unconditional stochastic simulation (Monte Carlo) paradigm in section 2.1 to review the concepts of randomization and ordinary block kriging as well as to define the fundamental assumptions and variables required. An example of a geometrical setup of the problem at hand is illustrated in Figure 1. The goal is to estimate the probability distribution of mass discharge  $Q$  [M/T] across a rectangular transect  $T$  containing  $n$  [-] local (“quasi-point”) PFM measurements  $q_{i\text{PFM}}$  ( $i = 1, 2, \dots, n$ ) [M/(L<sup>2</sup>T)] of mass fluxes along observations wells. Hereby, it is assumed that  $q_{i\text{PFM}}$  are error-free. Unconditional stochastic simulation can deliver realizations  $q(\mathbf{u})$  of a random function (or stochastic process) representing possible spatial distributions of local fluxes  $q$  at locations described by coordinate vector  $\mathbf{u}$  inside a transect. The unconditional realizations are



**Figure 1.** PFM sampling locations over transect at Ft. Lewis, WA.

interpreted as (partially) unknown scenarios of the real world with an equal likelihood of occurrence, where one single realization is assumed to perfectly match the real-world situation. The problem is that this single realization may not be identified and the outcomes of different realizations are used to make probabilistic statements about the unknown real-world situation. The constraints typically imposed in unconditional stochastic simulation are those of a given point or univariate distribution  $\text{pdf}(q)$  and a particular spatial correlation structure (variogram  $\gamma[\mathbf{h}]$  or covariance function  $C[\mathbf{h}]$ , where  $\mathbf{h}$  is a spatial separation vector). Up to second order,  $\text{pdf}(q)$  is defined by its mean  $\mu_0$  and variance  $\sigma_0^2 = C(0)$ , where the random function  $q$  is called second order stationary, if  $\mu_0$  and  $C(\mathbf{h})$  are location independent. Not considered in the present work are deterministic trend functions that may be combined with random function  $q(\mathbf{u})$ ; nor do we consider uncertainties in the type and parameters of  $C(\mathbf{h})$ . For the remainder of section 2.1  $C(\mathbf{h})$  does not have to be specified, while it is inferred from sample data  $q_{i\text{PFM}}$  for each practical example presented below. In cases where the spatial variance in  $q$  increases continuously with the size of the domain investigated, such that process variance  $\sigma_0^2$  and, hence, covariance function  $C(\mathbf{h})$  are not defined, the variogram  $\gamma(\mathbf{h})$  and the intrinsic hypothesis (i.e., stationarity of the increments [Journal and Huijbregts, 1978]) may still be used in what follows.

[9] Each unconditional realization may be fictitiously sampled for fluxes  $q_i$  [ $\text{M}/(\text{L}^2\text{T})$ ] at the  $n$  locations of actual PFM field measurements  $q_{i\text{PFM}}$  and a discharge may be estimated. Similarly, a respective true discharge  $Q$  across the transect may be computed for each unconditional realization resulting in a specific estimation error (i.e., estimated minus true value) per realization. Over many unconditional realizations the estimation errors form a distribution, which may be used as a performance (uncertainty) metric of the discharge estimate adopted. In what follows, rather than working with discharge  $Q$  over transect  $T$  of known size  $A$  [ $\text{L}^2$ ], the mean transect flux  $q_T = Q/A$  [ $\text{M}/(\text{L}^2\text{T})$ ] over  $T$  will be used as the magnitude to be estimated. Thus, for each unconditional realization, a value of  $q_T$  may be computed by

$$q_T = \frac{1}{A} \int_T q(\mathbf{u}) d\mathbf{u}, \quad (1)$$

which follows a probability distribution of variance  $\sigma_{q_T}^2$  given by [Journal and Huijbregts, 1978]

$$\sigma_{q_T}^2 = \frac{1}{A^2} \int_T \int_T C(\mathbf{u} - \mathbf{v}) d\mathbf{u} d\mathbf{v}, \quad (2)$$

where  $\mathbf{v}$  is a coordinate vector analogous to  $\mathbf{u}$ . Note that equation (2) is nothing but the average value of  $C(\mathbf{h})$  over all possible pairs of locations within  $T$ . In the same way as for  $q_T$ , the fictitiously sampled fluxes  $q_i$  of each unconditional realization are used to compute estimates  $\bar{q}$  of  $q_T$  through the linear operator

$$\bar{q} = \sum_{i=1}^n \lambda_i q_i, \quad (3)$$

where in the present work  $\lambda_i$  are taken to be the ordinary block kriging weights [Journal and Huijbregts, 1978; Goovaerts, 1997; Kitaniadis, 1997]. Moreover, the ensemble variance  $\sigma_{\bar{q}}^2$  of  $\bar{q}$  is known to be

$$\sigma_{\bar{q}}^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{u}_i - \mathbf{u}_j), \quad (4)$$

which, in analogy to equation (2), is nothing but the weighted average of  $C(\mathbf{h})$  over all possible pairs of sampling locations  $\mathbf{u}_i$  and  $\mathbf{u}_j$ . Finally, the covariance  $\sigma_{q_T \bar{q}}^2$  between  $q_T$  and  $\bar{q}$  results as

$$\sigma_{q_T \bar{q}}^2 = \frac{1}{A} \sum_{i=1}^n \lambda_i \int_T C(\mathbf{u}_i - \mathbf{v}) d\mathbf{v}, \quad (5)$$

which represents the weighted average of  $C(\mathbf{h})$  over all possible combinations of one sampling location  $\mathbf{u}_i$  and a location  $\mathbf{v}$  in  $T$ . With this, over many unconditional realizations, the random estimation error  $e = \bar{q} - q_T$  possesses an ensemble mean value of zero (unbiasedness) and an ensemble variance  $\sigma_e^2$  of

$$\sigma_e^2 = \sigma_{\bar{q}}^2 + \sigma_{q_T}^2 - 2\sigma_{q_T \bar{q}}^2, \quad (6)$$

which is obtained as the variance of the difference of two correlated random variables  $q_T$  and  $\bar{q}$ . Using ordinary block

kriging weights in equation (3) assures that  $\sigma_e^2$  is a minimum under the constraint that  $\sum_{i=1}^n \lambda_i = 1$ . The denomination “ordinary kriging” (versus “simple kriging,” for example) refers to the fact that  $\sum_{i=1}^n \lambda_i = 1$  is used to impose unbiasedness, such that the mean  $\mu_0$  of  $\text{pdf}(q)$  of the stochastic process does not appear in the estimator (equation (3)). In turn, “block kriging” (versus “point kriging,” for example) indicates that the estimated variable  $q_T$  is defined as a spatial average over a domain  $T$  (“block” in traditional mining language) rather than at a single point.

[10] In addition to evaluating the estimator in equation (3) for each unconditional realization, one may also compute a respective spatial variance  $s^2$  of the weighted histogram of the (fictitiously) measured local fluxes  $q_i$  in each realization. Hereby,  $\lambda_i$  are to be interpreted as declustering weights [Journal and Huijbregts, 1978; Goovaerts, 1997] used to estimate the frequency or probability of occurrence of each value  $q_i$  over the transect (rather than over the sampling locations), such that

$$s^2 = \sum_{i=1}^n \lambda_i (q_i - \bar{q})^2. \quad (7)$$

[11] The mean of  $s^2$  over many unconditional realizations is called the “weighted data dispersion variance”  $\sigma_{D\lambda}^2$  equal to

$$\sigma_{D\lambda}^2 = \sigma_0^2 - \sigma_q^2, \quad (8)$$

as derived in the appendix and with  $\sigma_q^2$  from equation (4). For generating unconditional realizations it is important that  $\sigma_{D\lambda}^2$  be close to  $\sigma_0^2$ , such that the variability of the underlying process is well represented by the data. In the same way, an unweighted dispersion variance of local fluxes over the transect may be defined as [Journal and Huijbregts, 1978]

$$\sigma_{DT}^2 = \sigma_0^2 - \sigma_{qT}^2, \quad (9)$$

with  $\sigma_{qT}^2$  from equation (2). Thus,  $\sigma_{DT}^2$  is the mean variance of local fluxes (i.e., dispersion) inside transect  $T$  taken over many unconditional realizations. Note that the difference between equations (2) and (9) is that  $\sigma_{qT}^2$  is the ensemble (i.e., over many realizations) variance of a spatial mean, while  $\sigma_{DT}^2$  is the ensemble mean of a spatial variance; the sum of both is equal to  $\sigma_0^2$ . Analogous observations apply to  $\sigma_q^2$  and  $\sigma_{D\lambda}^2$ . Equations (6), (8), and (9) exist under the intrinsic hypothesis and respective expressions in terms of the variogram are obtained by simply substituting  $-\gamma(\mathbf{h})$  for  $C(\mathbf{h})$  and, consequently, zero for  $\sigma_0^2$  [Journal and Huijbregts, 1978].

## 2.2. Conditional Simulation for Discharge Estimation

[12] As mentioned in the introduction (section 1), different conditional simulation approaches have recently been applied for contaminant discharge estimation. Conditional

simulation differs from unconditional simulation in that all conditional realizations honor (reproduce) the measured data values at their locations. Moreover, for the conditional simulation near PFM sampling locations it is sufficient that  $\sigma_{D\lambda}^2$  is close to  $\sigma_{DT}^2$  (equations (8) and (9)). This is a less restrictive condition than for unconditional simulation, since  $\sigma_{DT}^2 < \sigma_0^2$  (spatial data dispersion grows with domain size), and a dimensionless parameter  $\Delta\sigma^2$  termed “missing variance” is defined here as

$$\Delta\sigma^2 = \frac{\sigma_{DT}^2 - \sigma_{D\lambda}^2}{\sigma_{DT}^2}. \quad (10)$$

[13] This magnitude quantifies the relative portion of variability contained in the domain of interest  $T$ , which is not actually described by the available data after declustering. A graphical illustration of all steps involved in the conditional simulation approach is given in Figure 2 with further descriptions given here:

### 2.2.1. Variogram Analysis of Raw Data $q_{i\text{PFM}}$ (Step A)

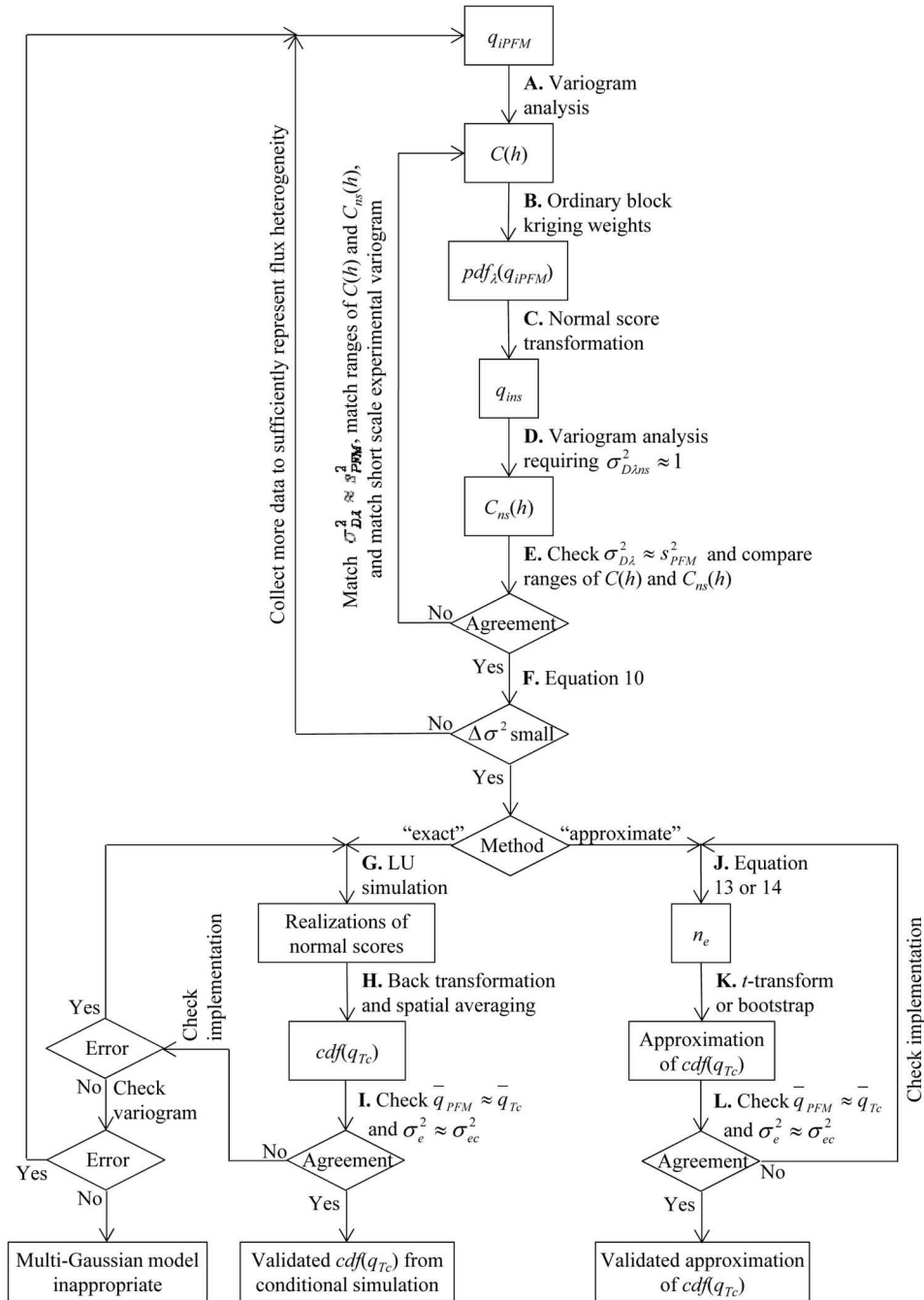
[14] Figure 1 shows an example of a rectangular transect with monitoring wells and PFM flux measurement intervals (typically  $\sim 30$  cm in length), indicated by dots at the center of each interval. For each dot, spatial coordinates and an observed value of  $q_{i\text{PFM}}$  are available for the generation of an experimental variogram and structural analysis [Journal and Huijbregts, 1978] to obtain a first estimate of the spatial covariance function  $C(\mathbf{h})$ . The experimental variogram points for shorter lag distances  $|\mathbf{h}|$  are typically better defined (based on more data pairs) and contain more important information for spatial interpolation (simulation). As a consequence, a good variogram fit near the origin is considered more valuable than at larger lag distances [Kitanidis, 1997].

### 2.2.2. Data Declustering Using Ordinary Block Kriging Weights (Step B)

[15] In the present work, the transect is identical to the simulation domain and it is roughly defined as extending one half-well separation distance beyond the lateral-most wells in the horizontal direction and from the bottom of the bottom-most to the top of the top-most PFM sampling interval. To achieve a representative flux distribution over the transect from irregularly spaced data, declustering is required. In the present work, the ordinary block kriging weights  $\lambda_i$  from equation (3) are used for this purpose [Journal and Huijbregts, 1978] as they accommodate the spatial arrangement of the sampling locations, the spatial correlation structure, as well as the position and size of the transect. The declustered data histogram is denoted by  $\text{pdf}_\lambda(q_{i\text{PFM}})$  to indicate that the declustered frequency of occurrence of local flux  $q_{i\text{PFM}}$  is equal to  $\lambda_i$  rather than  $1/n$ .

### 2.2.3. Conversion to Gaussianity Through Normal Score Transformation (Step C)

[16] Because of typically skewed data distributions, this step is required in order to take advantage of the properties of Gaussianity and to allow for using a relatively simple algorithm for the generation of Gaussian realizations. Theoretically, Gaussianity has to be achieved for the joint distribution of the random function at all locations (multi-Gaussianity), however, as a consequence of generally



**Figure 2.** Flowchart for conditional simulation approach (section 2.2) and approximate methods (section 4).

limited amounts of data for verification, we limit our attention to univariate Gaussianity (i.e., the transformation of  $pdf_{\lambda}[q_{iPFM}]$  to a normal distribution). The present work uses normal score transformation [Goovaerts, 1997], which is a numerical technique using a piecewise linear transformation function to achieve Gaussianity. In other words, the declustered data histogram is converted to a standard normally distributed histogram (zero mean and unit variance) by simply mapping respective quantiles onto each other. This is, if all measured fluxes  $q_{iPFM}$  are arranged in ascending order, then flux  $q_{jPFM}$  corresponds to a weighted cumulative distribution

function value  $cdf_{\lambda}(q_{jPFM}) \approx \frac{\lambda_j}{2} + \sum_{i=1}^{j-1} \lambda_i$  [Goovaerts, 1997], and the respective quantile of the normal distribution yields the normal score transform  $q_{ins}$  of  $q_{iPFM}$ . Thus, a correspondence table between  $q_{iPFM}$  and  $q_{ins}$  is constructed.

**2.2.4. Variogram Analysis of Normal Score Data  $q_{ins}$  (Step D)**

[17] In the same way as under step A, variogram analysis is performed on  $q_{ins}$  to fit a respective covariance function  $C_{ns}(h)$ . Because step C converts the declustered data into

a standard normal distribution of unit variance,  $C_{ns}(\mathbf{h})$  is fitted, such that the declustered normal score dispersion variance  $\sigma_{D\lambda_{ns}}^2$ , obtained from equations (4) and (8) with  $C_{ns}(\mathbf{h})$  instead of  $C(\mathbf{h})$ , is  $\sim 1$ . This helps in assuring that  $C_{ns}(\mathbf{h})$  is compatible with the conditioning data and that the simulated normal scores over the transect are close to standard normally distributed. The latter is required, such that  $\text{pdf}_\lambda(q_{i\text{PFM}})$ , which was particularly determined for the transect, is reproduced by the conditional simulation output after back-transformation in step H.

### 2.2.5. Joint Variogram Analysis of Raw Data and Normal Scores With Adjustment of $C(\mathbf{h})$ to Meet Weighted Data Dispersion Variance (Step E)

[18] Distributions of contaminant fluxes are typically positively skewed with many small data values and few very large ones. It is well known (and because of the squaring of flux differences involved) that the experimental variogram of the raw data is quite sensitive to large values, which may complicate a reliable fitting of  $C(\mathbf{h})$  through step A. More robust measures of dispersion (avoiding squaring) or variogram analysis on transformed data (avoiding large values, but preserving ranks) do not affect the spatial correlation features along the lag distance axis [Deutsch and Journel, 1992], and may thus be used to assist with inferring spatial correlation lengths. This also applies to normal score data [Chiles and Delfiner, 1999] as obtained from step C and as such, we perform a joint modeling of  $C(\mathbf{h})$  and  $C_{ns}(\mathbf{h})$  to assure that they have the same correlation lengths. Moreover, given the declustered PFM flux histogram  $\text{pdf}_\lambda(q_{i\text{PFM}})$ , its variance  $s_{\text{PFM}}^2$  is obtained by equations (3) and (7) with  $q_{i\text{PFM}}$  instead of  $q_i$ , and  $s_{\text{PFM}}^2$  is used as an estimator of its ensemble mean  $\sigma_{D\lambda}^2$  [Journel and Huijgrebts, 1978] to further improve the fit of  $C(\mathbf{h})$ . In other words,  $C(\mathbf{h})$  is modified (e.g., through adjusting horizontal and vertical variogram sills), such that  $\sigma_{D\lambda}^2$  from equation (8) is approximately equal to  $s_{\text{PFM}}^2$ . A convenient property of this approach is that a scaling of the variance axis of  $C(\mathbf{h})$  does not affect  $\lambda_i$  and, consequently,  $s_{\text{PFM}}^2$  at all. Because of the normal score transformation,  $C_{ns}(\mathbf{h})$  is also relatively robust with respect to changes in  $\lambda_i$  through range adjustments and a single manual iteration for fitting of  $C(\mathbf{h})$  is performed in this work. Overall, as apparent from Figure 2,  $C(\mathbf{h})$  is determined based on three sources of information: the experimental variogram of raw data  $q_{i\text{PFM}}$  for variogram fitting near the origin, the weighted data dispersion variance  $\sigma_{D\lambda}^2$  estimated by  $s_{\text{PFM}}^2$  to adjust scaling on the variance axis, and the correlation length(s) of the normal score variogram to adjust scaling along the distance axis.

### 2.2.6. Check for Representative Flux Sampling (Step F)

[19] The “missing variance”  $\Delta\sigma^2$  from equation (10) can be computed prior to any conditional simulation and elevated values may indicate that local flux heterogeneity  $\text{pdf}(q)$  over the transect is not sufficiently represented by the data collected. If this is the case, it may still be reasonable to quantify discharge and uncertainty in terms of an error variance  $\sigma_e^2$  without, however, being able to make reliable statements about the exact distribution of the estimation error (by any method considered herein). To overcome this limitation, additional PFM measurements or alternative approaches including complementary data, such

as contaminant concentrations, hydraulic head, and/or conductivities [Schwede and Cirpka, 2010; Trolldborg et al., 2010] may be required. While further investigation is needed to see whether a maximum admissible value for  $\Delta\sigma^2$  may be defined, the case studies in section 5 indicate that  $\Delta\sigma^2$  smaller than  $\sim 5\%$  leads to satisfactory results for discharge distributions.

### 2.2.7. Conditional Simulation Using $q_{\text{ins}}$ and $C_{ns}(\mathbf{h})$ (Step G)

[20] LU decomposition (lower-upper triangle or also known as Cholesky decomposition) [Goovaerts, 1997] is chosen as a fast algorithm for multi-Gaussian conditional simulation of up to several thousand grid points. It possesses the advantage that all data and grid locations are considered simultaneously as a whole, rather than as little portions inside local neighborhoods. This avoids approximation errors, which may deteriorate the simulation performance in terms of histogram or variogram reproduction.

### 2.2.8. Back-Transformation and Spatial Averaging (Step H)

[21] The resulting conditional realizations of normal scores are back-transformed into fluxes by the inverse of the transformation in step C, which is quantile matching (with linear interpolation) in the inverse direction. At the tails, i.e., outside of the range of  $q_{i\text{PFM}}$  measured, extrapolation models have to be chosen (“distributional extrapolation”). It was found, however, that with sufficiently dense PFM sampling, enough data is available, such that this choice does not influence results (i.e., the missing variance  $\Delta\sigma^2$  and consequent probability of simulating local flux values outside the range of measured fluxes are sufficiently small). Hence, the lower tail is linearly extrapolated to zero, which is generally equal to the minimum data value if transects span entire plumes, while the upper tail is limited to the maximum data value. Given the conditionally simulated fluxes at all grid points, spatial averaging is performed on each realization to obtain conditionally simulated values of transect average fluxes  $q_{Tc}^{(j)}$  (by using a discrete form of equation (1)). The additional subscript “c” indicates that  $q_T$  is conditioned to measured fluxes and the index “j” denotes the number of the realization that  $q_{Tc}$  corresponds to. Knowing that a total number  $N$  (i.e.,  $j = 1, 2, \dots, N$ ) of conditional realizations are generated, a discrete approximation of the conditional cumulative distribution function  $\text{cdf}(q_{Tc})$  may be obtained. The latter is regarded as the probability distribution of the unknown transect average flux  $q_T$  in the field [Goovaerts, 1997]. The ensemble mean  $\bar{q}_{Tc}$  of  $\text{cdf}(q_{Tc})$  approximated by

$$\bar{q}_{Tc} = \frac{1}{N} \sum_{j=1}^N q_{Tc}^{(j)}, \quad (11)$$

is used as a conditional simulation based estimate of the true  $q_T$ , while the respective estimation error variance is given by

$$\sigma_{ec}^2 = \frac{1}{N} \sum_{j=1}^N \left( q_{Tc}^{(j)} - \bar{q}_{Tc} \right)^2. \quad (12)$$

[22] In the present work,  $N = 1000$  conditional realizations are used, which results in insignificant uncertainty

(because of a finite number of realizations) in equations (11) and (12) as well as in  $\text{cdf}(q_{Tc})$  over the principal range of interest (i.e., approximately between 0.01 and 0.99 quantiles). Once  $\text{cdf}(q_{Tc})$  is known, a multiplication of quantiles by transect area  $A$  produces the cdf of transect discharge  $Q$  as  $\text{cdf}(Q) = \text{cdf}(Aq_{Tc})$ .

### 2.2.9. Validation of Results (Step I)

[23] As a consequence of good histogram reproduction,  $\bar{q}_{Tc}$  from equation (11) should be in good agreement with the mean of  $\text{pdf}_\lambda(q_{i\text{PFM}})$ . By step B, the mean of  $\text{pdf}_\lambda(q_{i\text{PFM}})$  is again equal to the ordinary block kriging estimate  $\bar{q}_{\text{PFM}}$  obtained from equation (3) with actual sample values of  $q_{i\text{PFM}}$ . Furthermore, Gaussian conditional simulation algorithms are based on the properties of simple kriging [Goovaerts, 1997]. It is known that these properties are well approximated by those of ordinary kriging, if a sufficient amount of data is available within the domain of interest (i.e., small  $\Delta\sigma^2$  [Emery, 2007]). Thus, the error variance  $\sigma_{ec}^2$  from equation (12) should be in good agreement with  $\sigma_e^2$  from equation (6). Note that both variances are independent of higher order moments (e.g., distribution skewness of  $\text{pdf}[q_{i\text{PFM}}]$ ) and that issues related to conditional bias (i.e., overestimation of variance in areas of low variability and underestimation of variance in areas of high variability) do not arise in the present situation. This is because estimation is over the whole transect encompassing all data and both high and low variability zones. Thus, as indicated by Figure 2, the output of the conditional simulation method may be validated to some degree by verifying  $\bar{q}_{\text{PFM}} \approx \bar{q}_{Tc}$  and  $\sigma_e^2 \approx \sigma_{ec}^2$ . Significant discrepancies may indicate errors in the implementation of the conditional simulation routine or the possibility of improvements in variogram fits (within the requirements of steps A, D, and E). If none of the above is the case, then a conditional simulation approach based on transformation of a multi-Gaussian field may be inadequate for the data at hand.

## 3. Effective Number of Independent Data $n_e$

### 3.1. Purpose, Definition, and Basic Properties

[24] The implementation of the conditional simulation approach of step G may not be a straightforward task for practitioners. In an effort to offer simpler alternatives, which closely approximate well-calibrated conditional simulation results (i.e., successfully validated in step I), an effective number of independent data  $n_e$  is defined and proposed for use with two estimation methods from classic independent and identically distributed (i.i.d.) statistics. The parameter  $n_e$  contains information about the transect and sampling geometries as well as of the spatial correlation structure. Similar to Kitanidis [1997, p. 157], Deutsch [2004], or Klammler *et al.* [2011] it is defined by

$$n_e = \frac{\sigma_{D\lambda}^2}{\sigma_e^2}, \quad (13)$$

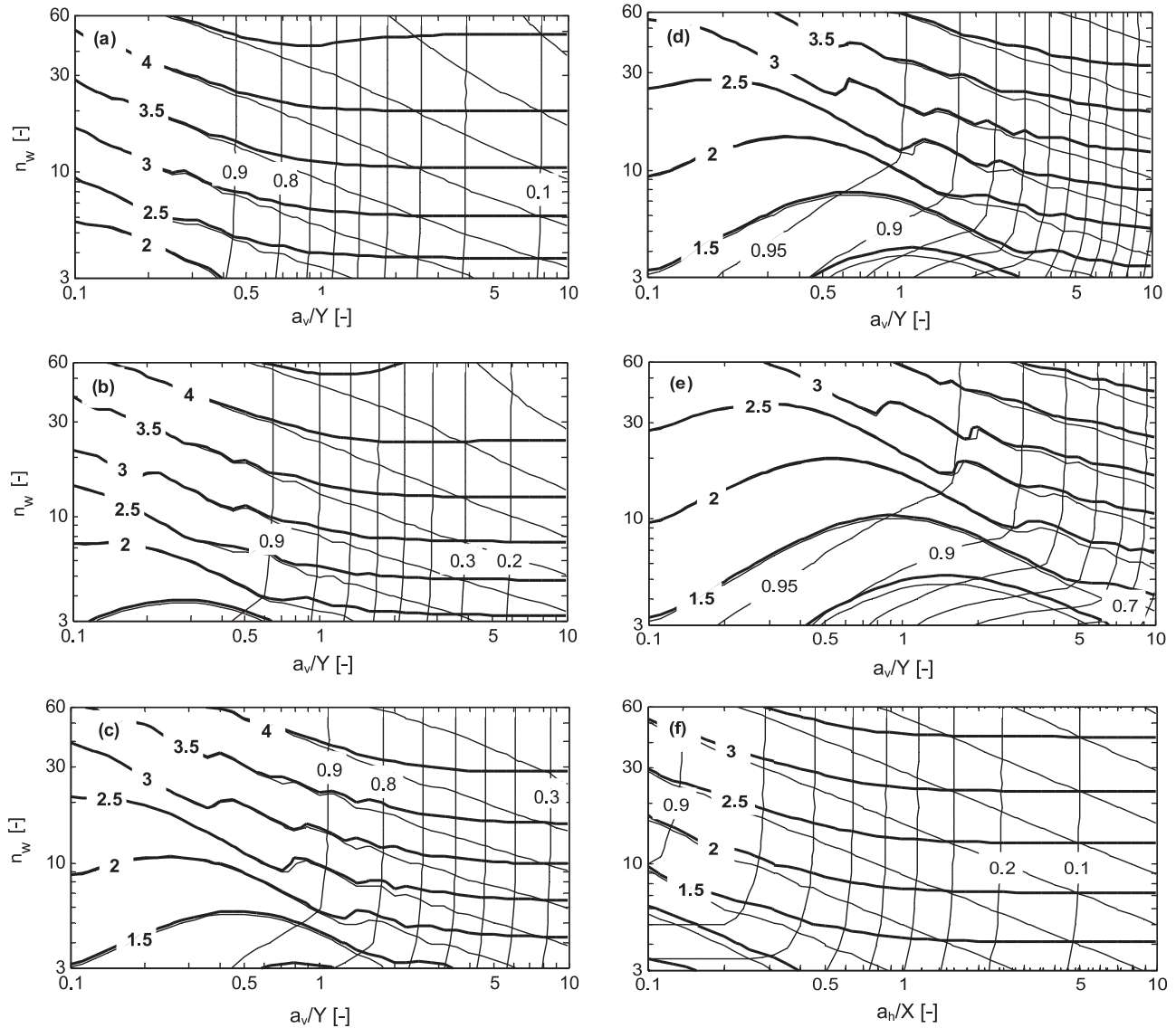
where  $\sigma_e^2$  is the ordinary block kriging variance of equation (6) and  $\sigma_{D\lambda}^2$  is the weighted data dispersion variance of equation (8). In analogy to classical statistics of uncorrelated data,  $n_e$  acts as a proportionality factor between the variance of observed data ( $\sigma_{D\lambda}^2$  or its estimator  $s_{\text{PFM}}^2$ , see the appendix) and the variance of a mean estimate ( $\sigma_e^2$ ).

This means that the ratio in equation (13) between the spatial variance of the  $n$  correlated flux data and the ensemble variance of the error of the transect mean estimate is equal to the ratio of the variance of  $n_e$  hypothetical data values (which are statistically independent) and the variance of their mean estimate (compare “standard error equation” from i.i.d. statistics). Thus, in terms of the error variance of a mean estimate,  $n$  correlated data achieve the same variance reduction as  $n_e$  uncorrelated data. It may be observed that the effective number of independent data  $n_e$  is unaffected by the scaling of  $C(\mathbf{h})$  along the variance axis, since both  $\sigma_{D\lambda}^2$  and  $\sigma_e^2$  are scaled by the same factor. Furthermore, in the same way as  $\sigma_{D\lambda}^2$  and  $\sigma_e^2$ ,  $n_e$  exists under the intrinsic hypothesis and can range from zero to infinity. That is,  $n_e = 0$  is obtained for a single-point sample leading to the trivial case of  $\sigma_{D\lambda}^2 = 0$ , while  $n_e \rightarrow \infty$  is obtained when the entire transect is sampled. The latter corresponds to complete sampling of a finite population in classical statistics, where uncertainty is totally removed. Intuitively speaking, in the range  $n_e < n$ , redundancy between nearby sampling locations prevails, while for  $n_e > n$  the favorable effect of correlation between sampled and unsampled locations is predominant, which decreases the interpolation uncertainty between monitoring wells.

### 3.2. Properties of $n_e$ Related to PFM Discharge Estimation

[25] From above, it may be understood that  $n_e$  acts as a kind of effective parameter quantifying the efficiency of a sampling pattern in terms of variance reduction between data dispersion (observed flux heterogeneity) and estimation error of the spatial mean. In order to analyze some general aspects related to  $n_e$  and PFM discharge estimation, section 3.2 assumes regularly spaced monitoring wells along a rectangular transect with a constant number of 10 vertical PFM sampling intervals. Evaluating equations (2), (4), and (5) for the computation of  $n_e$  generally requires numerical methods [Journal and Huijbregts, 1978] and direct statements about the quantitative behavior of  $n_e$  for practical purposes are difficult to make. For this reason, Figure 3 (bold continuous lines) shows a normalized representation of the order of magnitude of  $n_e$  as a function of the number of PFM wells  $n_w$  [-], horizontal and vertical correlations lengths  $a_h$  and  $a_v$  [L], as well as transect width  $X$  [L] ( $x$  direction) and depth  $Y$  [L] ( $y$  direction). The spherical covariance model [Goovaerts, 1997] is used for  $C(\mathbf{h})$  and integration is performed on discretized transects of  $10 \times 200$  (depth per width) rectangular cells. This implies that each PFM well corresponds to a grid column and the sampling over depth is exhaustive as it is typical for PFM measurements (every cell in a PFM grid column is considered “sampled” corresponding to a PFM sampling interval). Each chart in Figure 3 corresponds to a particular value of  $a_v X / (a_h Y)$  representing the aspect ratio of the transect after scaling to isotropic correlation length. In Figure 3f  $a_v = \infty$ , which reduces the problem to one-dimension in the  $x$  direction and, hence,  $a_h / X$  is used for the abscissa as opposed to  $a_v / Y$  in Figures 3a–3e. For  $a_h = \infty$ , in contrast, the problem is reduced to the  $y$  direction, which is entirely sampled, thus leading to  $\sigma_e^2 = 0$  and  $n_e = \infty$  (no uncertainty left). Both cases correspond to zonal anisotropies [Goovaerts, 1997] and may be relevant for PFM applications. The former





**Figure 3.** Contour lines of  $\sigma_{D\lambda}^2/\sigma_0^2$  (dashed lines),  $-\log_{10}\sigma_e^2/\sigma_0^2$  (thin continuous lines), and  $\log_{10}n_e$  (bold continuous lines) for PFM transects of  $a_vX/(a_hY) = \{1, 2, 5, 10, 20, \infty\}$  in (a)–(f), respectively. PFM wells are assumed to be regularly spaced ( $\lambda_i = 1/n$ ) and possess 10 vertical sampling intervals. Spherical model for  $C(\mathbf{h})$ .

( $a_v = \infty$ ) is known as a random areal trend and can reflect the fact that certain wells are located in high- and others in low-flux zones (e.g., plume core versus periphery). The latter ( $a_h = \infty$ ) typically reflects random stratigraphic layering. Apparent irregularities of contour lines for  $\sigma_e^2/\sigma_0^2$  and  $n_e$  over certain ranges in Figure 3 are caused by a type of periodicity appearing when  $a_h$  reaches multiples of the PFM well-separation distance. Without disturbing the general tendency of  $n_e$ , the strongest effect occurs at  $a_h = X/n_w$ . The periodicity is found to be an artifact of the spherical covariance model adopted and it is not observed with the smoother exponential covariance function. For either model of  $C(\mathbf{h})$ , the contour lines of  $n_e$  share similar characteristics and become horizontal as  $a_h > X$  and  $a_v > Y$ , i.e., when  $C(\mathbf{h})$  becomes linear over the extent of the transect. In such

a case (e.g., Rifle case study in section 5.2),  $n_e$  becomes independent of the individual values of  $a_h$  and  $a_v$  reducing to a function of the anisotropy ratio  $a_h/a_v$  only.

[26] The structure of  $C(\mathbf{h})$  may be nested, which means it consists of a number  $m$  [-] of substructures  $C_j(\mathbf{h})$  of different geometric or zonal anisotropies such that  $C(\mathbf{h}) = \sum_{j=1}^m C_j(\mathbf{h})$ . According to the principle of factorial kriging [Goovaerts, 1997], the substructures may be regarded as related to mutually independent stochastic sub-processes, which add up to the composite process corresponding to  $C(\mathbf{h})$ . As a consequence, dispersion and estimation variances of the subprocesses ( $\sigma_{D\lambda_j}^2$  and  $\sigma_{e_j}^2$ , respectively) also add up to the respective variances of the



composite process, and  $n_e$  from equation (13) can be expressed more generally as

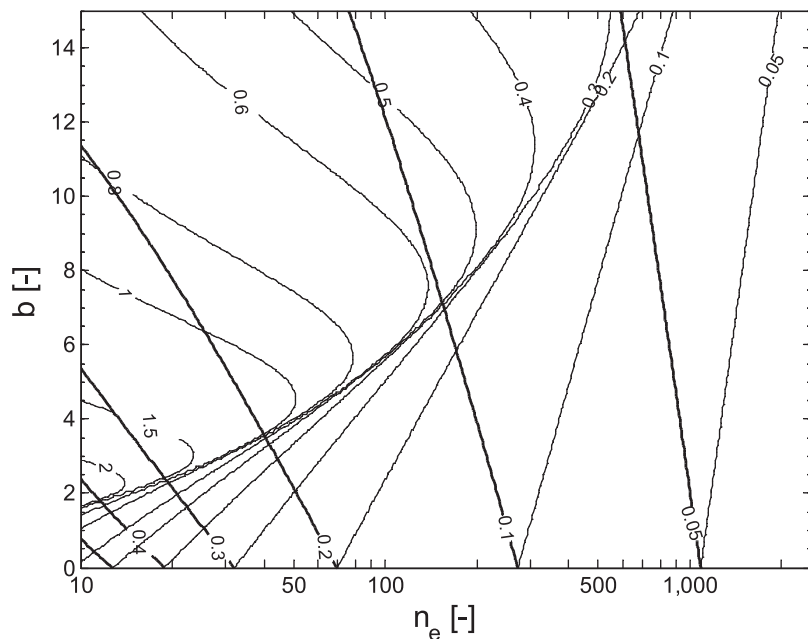
$$n_e = \frac{\sum_{j=1}^m \sigma_{D\lambda j}^2}{\sum_{j=1}^m \sigma_{e j}^2} = \frac{\sum_{j=1}^m \sigma_{e j}^2 n_{e j}}{\sum_{j=1}^m \sigma_{e j}^2}. \quad (14)$$

[27] Note that the same declustering weights obtained for the compound covariance function  $C(\mathbf{h})$  under step B in section 2.2 are used for all variogram components in equation (14). If data are approximately regularly spaced, the contour lines of  $\sigma_{D\lambda}^2/\sigma_0^2$  (dashed) and  $\sigma_e^2/\sigma_0^2$  (thin continuous) in Figure 3 may be applied to individual substructures of a nested  $C(\mathbf{h})$  to approximate  $n_e$  through equation (14). For each substructure,  $\sigma_0^2$  in Figure 3 has to be taken as the variance of that substructure only. The last expression in equation (14) shows that the compound  $n_e$  may also be represented as a weighted sum of all substructures'  $n_{e j}$ , where weighting is proportional to the relative contribution  $\sigma_{e j}^2/\sigma_e^2$  of each structure to the total estimation error  $\sigma_e^2$ .

[28] Figure 3 provides a means to evaluate the sensitivity of  $n_e$  with respect to the (somewhat subjective) choice of variogram ranges and the number of PFM wells installed over a transect. The latter is of interest for the evaluation of the benefit of additional sampling in terms of uncertainty reduction in discharge estimates. Under the hypothesis that  $C(\mathbf{h})$  is not significantly different after additional sampling, Figure 3 shows, for example, that over most of the range represented (especially for  $a_v/Y > 0.5$ ), doubling the number  $n_w$  of PFM wells approximately increases  $n_e$  by a factor of  $10^{0.5} \approx 3$ . For shorter correlation lengths (e.g., approaching a nugget variogram and data independence),  $n_e \approx n$  is known from section 3.1 and a proportional increase in  $n_w$  leads to the

same proportional increase in  $n$  and, hence,  $n_e$ . Thus, it may be stated that doubling the number  $n_w$  of wells results in an overproportional increase in the effective number of independent data  $n_e$  by a factor between  $\sim 2$  and 3. By virtue of the final expression in equation (14), this remains valid also for nested variogram structures. However, by inspection of equation (15) or Figure 4 in section 4.1, for example, it is seen that the width of confidence intervals is (inversely) proportional to  $\sqrt{n_e}$ . This means that doubling  $n_w$  may be expected to decrease the width of confidence intervals about a discharge estimate underproportionally by a factor between  $\sim \sqrt{2} \approx 1.4$  and  $\sqrt{3} \approx 1.7$ . This is in agreement with Kübert and Finkel [2006], who numerically simulate contaminant transport and PFM measurements in heterogeneous aquifers and who observe that the reduction in estimation uncertainty due to increasing  $n_w$  decreases as  $n_w$  grows.

[29] Further inspection of Figure 3 shows that the relative data dispersion variance  $\sigma_{D\lambda}^2/\sigma_0^2$  increases with  $n_w$  until PFM well separation becomes smaller than  $a_h$ . At this point the observed data represent practically all of the variability  $\sigma_{DT}^2/\sigma_0^2$  contained in the transect and data dispersion cannot increase any more with  $n_w$ . Based on results of the case studies below, it may be tentatively stipulated, for example, that distributional extrapolation is acceptable as long as the declustered (weighted) data dispersion variance accounts for at least 95% of the transect dispersion variance (i.e.,  $\Delta\sigma^2 < 0.05$ ). Lines for  $\Delta\sigma^2 = 0.05$  in Figure 3 are not shown for clearness of the plots, but they are found to be close to the contour lines of  $n_e = 30$ , such that the approximate condition  $n_e \geq 30$  may be given to contain distributional extrapolation within 5%. For a transect of given parameters  $X/Y$ ,  $n_w$ , and  $a_h$ , reducing  $a_v$  decreases  $a_v X/(a_h Y)$  and results in an increase in  $n_e$ , which may be attributed to less correlation (less data redundancy) in the vertically aligned data. In contrast, a decrease in  $a_h$  increases  $a_v X/$



**Figure 4.** Contour lines of  $t'_{n_e-1,1-0.05}/\sqrt{n_e}$  (thick lines) and  $-t'_{n_e-1,1-0.95}/\sqrt{n_e}$  (thin lines) as functions of an effective number of independent data  $n_e$  (equation (13) or (14)) and declustered data coefficient of skewness  $b$  (equation (18)) for use in equation (15).

$(a_h Y)$  and decreases  $n_e$ , which reflects the reduced correlation between sampled and unsampled (to be interpolated) locations. In other words, short-vertical and large-horizontal correlation scales, as typical for hydrogeological field situations, tend to increase  $n_e$ . This acts favorably on the uncertainty of discharge estimates, when sampling in the vertical direction is dense and indicates, in agreement with *Kübert and Finkel* [2006], a significant potential for the PFM method to deliver reliable discharge estimates.

#### 4. Approximate Methods Using $n_e$

[30] Section 4 presents two relatively simple methods from classical i.i.d. statistics, which both make use of  $n_e$  to approximate conditional simulation results. One applies a transformation of the  $t$  statistic, which is capable of accounting for distribution skewness and the other method uses the classic bootstrap on the full  $\text{pdf}_\lambda(q_{i\text{PFM}})$ . Steps J and K in Figure 2 illustrate how  $n_e$  and the approximate methods are implemented in the present work and how they circumvent the need for conditional simulation to arrive at an approximation of  $\text{cdf}(q_{Tc})$ . The idea of the approximate methods is based on the fact that the averaging of  $n_e$  independent data following a distribution  $\text{pdf}_\lambda(q_{i\text{PFM}})$  leads to the same mean estimate and error variance as averaging over the spatially correlated random function through the conditional simulation approach. This means that the first- and second-order statistics of spatial averaging are exactly honored by using  $n_e$ , while the influence of spatial correlation on higher-order moments is approximately incorporated through  $n_e$  and its effect due to the central limit theorem. Similar to step I above, step L in Figure 2 validates results up to second order, but may only indicate errors in the implementation of the approximate methods.

##### 4.1. Transformation of the $t$ Statistic

[31] *Willink* [2005], *Zhou and Dinh* [2005], as well as *Yanagihara and Yuan* [2005] independently present identical methods for improved determination of nonparametric confidence intervals for mean estimates of a wide range of skewed i.i.d. samples. The method is based on a transformation of the common Student's  $t$  statistic leading to a parameter denoted as  $t_2$ ,  $T_1$ , and  $\tilde{t}$  by the above authors, respectively. For the present purpose, we propose the use of this method in combination with  $\text{pdf}_\lambda(q_{i\text{PFM}})$  and  $n_e$  instead of  $\text{pdf}(q_{i\text{PFM}})$  and  $n$  as in the i.i.d. case. With this, the quantile  $q_\alpha$  of the distribution of the spatial mean estimate, which is not exceeded by a probability  $\alpha$ , is approximated by

$$q_\alpha = \bar{q}_{\text{PFM}} - \frac{t'_{n_e-1,1-\alpha} s_{\text{PFM}}}{\sqrt{n_e}}, \quad (15)$$

where  $s_{\text{PFM}}$  is the declustered data standard deviation obtained as the square root of equation (7) by using  $q_{i\text{PFM}}$  instead of  $q_i$  and  $\bar{q}_{\text{PFM}}$  instead of  $\bar{q}$ . Thus, a two-sided  $1-2\alpha$  confidence interval about the mean estimate  $\bar{q}_{\text{PFM}}$  from equation (3) results as  $[q_\alpha, q_{1-\alpha}]$ , where  $t'_{n_e-1,1-\alpha}$  in equation (15) is a cubic function of the classic Student's  $t$  variable  $t_{n_e-1,1-\alpha}$  for  $n_e-1$  degrees of freedom [*Willink*, 2005],

$$t'_{n_e-1,1-\alpha} = \frac{\sqrt[3]{1 + 6a(t_{n_e-1,1-\alpha} - a)} - 1}{2a}. \quad (16)$$

[32] This transformation is graphically represented in Figure 4 and incorporates distribution asymmetry in the parameter  $a$  in terms of the declustered sample coefficient of skewness  $b$ ,

$$a = \frac{b}{6\sqrt{n_e}}, \quad (17)$$

$$b = \frac{\sum_{i=1}^n \lambda_i (q_{i\text{PFM}} - \bar{q}_{\text{PFM}})^3}{s_{\text{PFM}}^3}. \quad (18)$$

[33] Equation (17) reflects the consequence of the central limit theorem that both variance and skewness of the distribution of the mean estimate decrease with  $n_e$ . For symmetric distributions ( $b = 0$ ) or large  $n_e$ ,  $a$  becomes small and equation (16) reduces to  $t'_{n_e-1,1-\alpha} = t_{n_e-1,1-\alpha}$ . Otherwise, for positive skewness ( $b > 0$ ), equation (16) results in asymmetric confidence intervals, which reflect a larger probability of underestimation and which work toward the requirement of non-negativity (confidence intervals shifted to right). For two-sided confidence intervals of  $1-2\alpha \leq 0.99$  and i.i.d. samples, *Willink* [2005] concludes that equation (16) is appropriate for  $n_e \geq 30$ .

##### 4.2. Bootstrap

[34] The bootstrap [*Efron and Tibshirani*, 1998] is a method based on distribution resampling, which is applicable to a diverse range of statistical problems. In its simplest form it consists of randomly drawing (with replacement)  $n$  values from an i.i.d. sample to compute a desired sample statistics (e.g., mean). Repeated performance of this process populates a distribution, from which a mean estimate and uncertainty intervals can be inferred. From a geostatistical standpoint, bootstrapping may be considered as a degenerate form of unconditional stochastic simulation over  $n$  locations, which are spatially uncorrelated. Similar to *Klammler et al.* [2011], in the present work, we propose and demonstrate the use of the bootstrap by resampling  $n_e$  times with equal likelihood from the declustered data distribution  $\text{pdf}_\lambda(q_{i\text{PFM}})$ . For computational convenience, it is noted that this is equivalent to resampling  $n_e$  times from the unweighted (i.e., raw) sample distribution  $\text{pdf}(q_{i\text{PFM}})$ , where each datum is drawn with probability  $\lambda_i$ .

## 5. Application to Estimate Discharges From PFM Field Measurements

### 5.1. Estimating TCE Discharge at Ft. Lewis, WA, Military Base

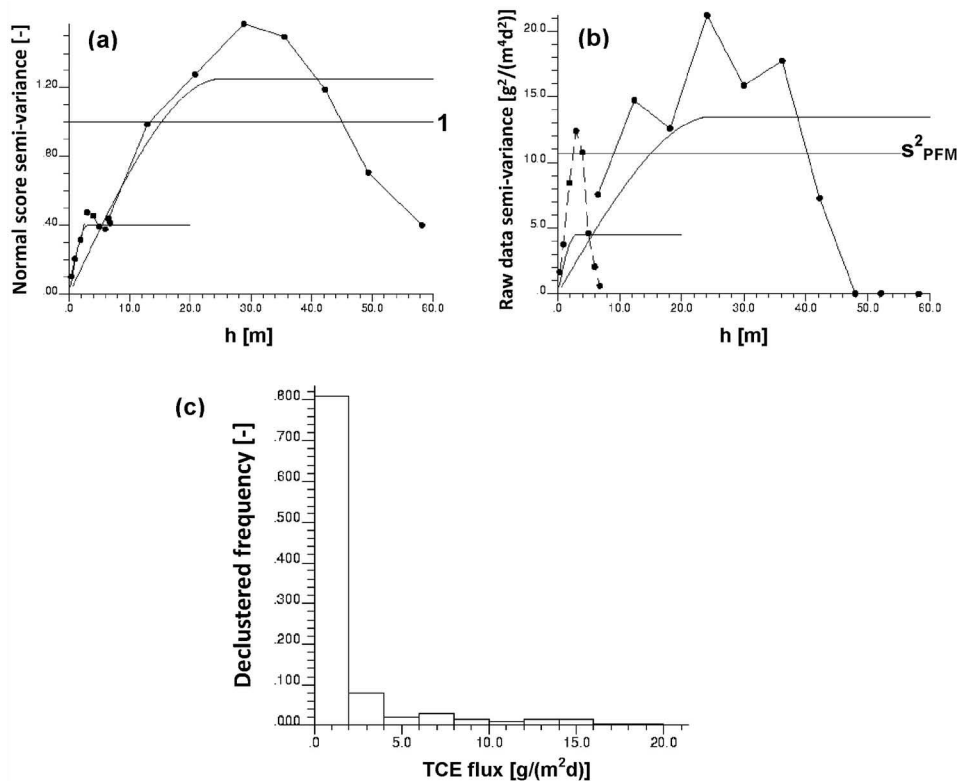
#### 5.1.1. Ten Well Scenario

[35] At Ft. Lewis, WA, Military Base extensive use of TCE as a degreasing substance led to the emergence of numerous TCE plumes in the groundwater. For one of the plumes investigated, transect limits (width  $X = 64.3$  m, depth  $Y = 7.6$  m, area  $A = XY = 490$  m<sup>2</sup>), PFM well locations, and segmentation are illustrated in Figure 1. Approximately uniform well spacing of 6 m results in  $n_w = 10$  wells with  $\sim 21$  PFM sampling intervals per well of 35 cm in length leading to a total of  $n = 204$  local flux measurements. The experimental variogram points for the horizontal

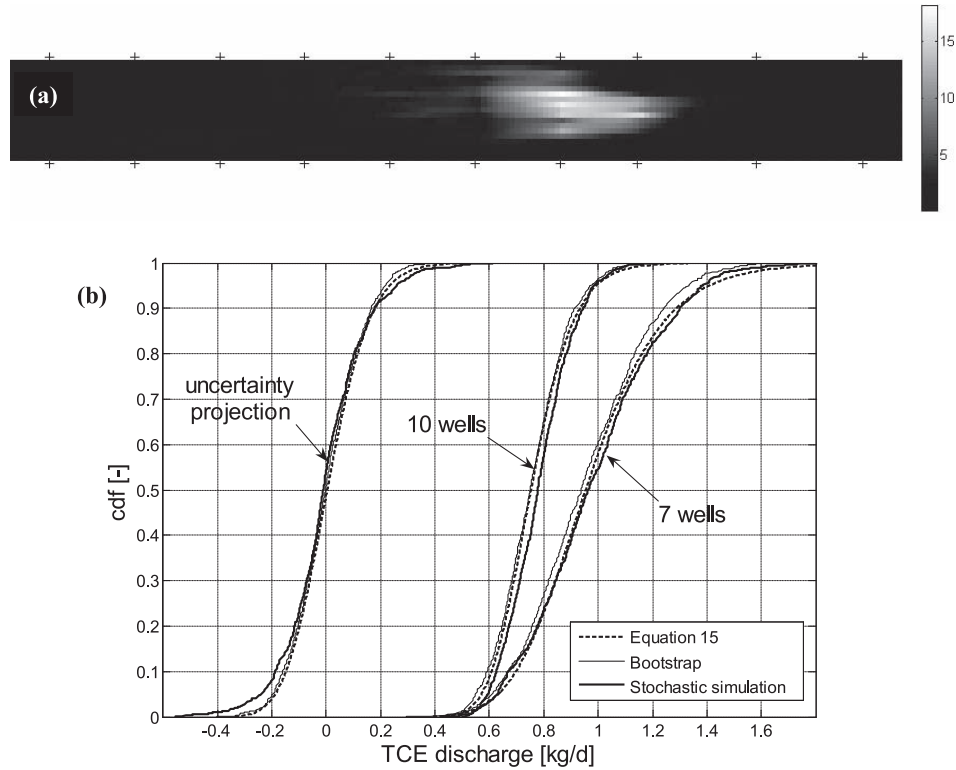
and vertical directions are depicted in Figures 5a and 5b for the normal scores and the raw data, respectively. Performing steps A through E of the proposed method (see section 2.2 and Figure 2), the sum of two spherical models is chosen to represent  $C_{ns}(\mathbf{h})$  and the following parameters are fitted:  $\sigma_{0ns1}^2 = 0.4$ ,  $a_{h1} = 25$  m,  $a_{v1} = 3$  m, and  $\sigma_{0ns2}^2 = 0.85$ ,  $a_{h2} = 25$  m,  $a_{v2} = \infty$ . The same correlation ranges are adopted for  $C(\mathbf{h})$  and the respective sills are fitted as  $\sigma_{01}^2 = 4.5$  g<sup>2</sup>/(m<sup>4</sup> d<sup>2</sup>) and  $\sigma_{02}^2 = 9$  g<sup>2</sup>/(m<sup>4</sup> d<sup>2</sup>). Both  $\sigma_{D\lambda}^2 \approx s_{PFM}^2$ , as indicated by the horizontal lines in Figures 5a and 5b, are satisfied to approximately  $\pm 1\%$  and the fit to the experimental variogram points, especially for short lag distances, appears to be quite good. At larger lag distances the fit becomes worse; however, the variogram sills do represent the correct average degree of variability. Using  $C(\mathbf{h})$  from Figure 5b and  $\lambda_i$  from ordinary block kriging, equations (3), (7), and (18) (with  $q_{iPFM}$  instead of  $q_i$ ) give an estimated mean TCE flux over the transect of  $\bar{q}_{PFM} = 1.53$  g/(m<sup>2</sup> d), a declustered data dispersion variance  $s_{PFM}^2 = 10.7$  g<sup>2</sup>/(m<sup>4</sup> d<sup>2</sup>), and a coefficient of skewness  $b = 3.10$ . The respective declustered histogram is depicted in Figure 5c with a “missing variance”  $\Delta\sigma^2$  from step F amounting to  $\sim 3\%$ .

[36] The variogram analysis indicates both larger statistical continuity as well as a higher degree of variability in the horizontal direction. The former may be attributed to horizontally more continuous hydrogeological conditions of aquifer and source zone, while the latter reflects variability in local fluxes between plume center and lateral periphery.

Applying Figure 3 to the first variogram structure with  $a_v X / (a_h Y) = 1.0$ ,  $n_w = 10$  and  $a_{v1} / Y = 0.4$  gives  $\sigma_{D\lambda 1}^2 / \sigma_{01}^2 = 0.91$  and  $\sigma_{e1}^2 / \sigma_{01}^2 = 10^{-3.2}$  (Figure 3a). For the second structure with  $a_{h2} / X = 0.4$  and  $n_w = 10$ , respective variances  $\sigma_{D\lambda 2}^2 / \sigma_{02}^2 = 0.72$  and  $\sigma_{e2}^2 / \sigma_{02}^2 = 10^{-2.2}$  are obtained (Figure 3f). From this  $n_e \approx 177$  is determined from equation (14), which is close to  $n_e = 169$  from direct application of equation (13) with  $\lambda_i$  from ordinary block kriging. Besides rounding errors (e.g., chart readings are only to the first decimal of the exponents), the difference is due to the assumption of equal well and sample spacing in Figure 3. The assumption of 10 vertical PFM intervals per well in Figure 3 can be shown to be of minor significance, because for the range depicted ( $a_v / Y \geq 0.1$ ) additional samples are increasingly redundant (correlated to previous samples). Finally, validation by step I (see Figure 2) indicates acceptable properties of the conditional simulation output: The discrepancies between  $\bar{q}_{PFM}$  and  $\bar{q}_{TC}$  and between  $\sigma_e^2$  and  $\sigma_{ec}^2$  amount to  $\sim 4\%$  each. Figure 6a is a representation of the spatial distribution of the ensemble means of local fluxes (20 vertical  $\times$  160 horizontal grid elements) from 1000 conditional realizations and illustrates an estimated position and shape of the plume core. Figure 6b contains the estimated cumulative distribution functions (cdf's) of TCE discharge  $Q = Aq_{TC}$  illustrating good agreement between conditional simulation and the approximation methods. Means and standard deviations of all methods are listed in the top part of Table 1 for comparison to “target values” from block kriging. Table 1 further lists the 5%



**Figure 5.** Horizontal (continuous lines) and vertical (dashed lines) experimental variograms of (a) normal scores and (b) raw data with spherical model fits for TCE fluxes from 10 wells at Ft. Lewis, WA. (c) Declustered data histogram.



**Figure 6.** (a) Distribution of ensemble means of local fluxes in  $\text{g}/(\text{m}^2 \text{d})$  from all conditional realizations. Crosses indicate PFM well locations according to Figure 1. (b) Cumulative distribution functions of TCE discharge. Graphs denoted by “7 wells” and “10 wells” compare results from conditional simulation, equation (15), and bootstrap for the respective number of observation wells indicated. Graphs called “uncertainty projection” compare unconditional simulation, equation (15), and bootstrap for predicting variance reduction due to additional sampling. One thousand realizations were used for all stochastic simulations and bootstrap.

and 95% quantiles of estimated discharge distributions showing maximum discrepancies of  $\sim 7\%$  between conditional simulation and the approximation methods.

**5.1.2. Seven Well Scenario**

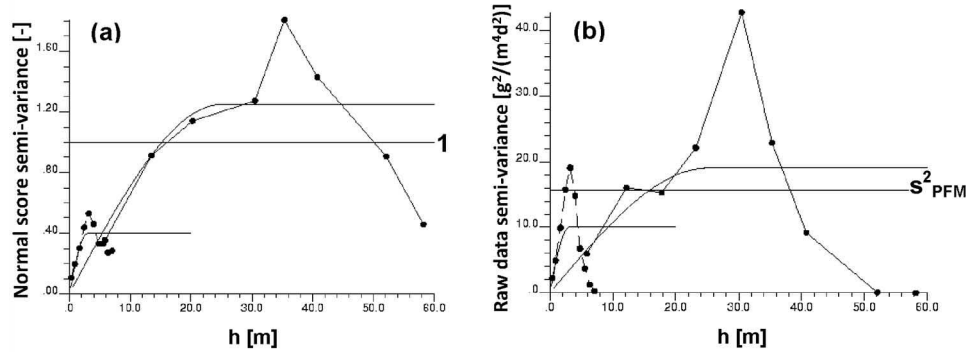
[37] Although the above confidence intervals are slightly asymmetric about the respective means, the large value of  $n_e$  causes the discharge distributions to be quite close to normal (indicated by a small value of  $a = 0.04$  in equation (16)). In general, however, discharge distributions may strongly

deviate from normality, and to demonstrate the performance of the proposed approximate methods under such scenarios the above analysis is repeated after excluding the local flux data from three wells (third, fifth, and ninth well from left to right in Figure 1) resulting in  $n = 141$  data points. Following again the method of section 2.2 (see Figure 2), for variogram model fitting it is found that all parameters may be maintained, except for  $\sigma_{01}^2 = 10.0 \text{ g}^2/(\text{m}^4 \text{d}^2)$ , which is illustrated by Figure 7. This results in a mean, variance, and coefficient of skewness of the declustered data

**Table 1.** Comparison of  $n_e$ ,  $\Delta\sigma^2$ , and Selected Statistics of Estimated Discharges From All Methods and Sites Investigated<sup>a</sup>

Site	Discharge Parameter	Block Kriging	Conditional Simulation	<i>t</i> Transformation	Bootstrap
Ft. Lewis (10 wells) $n_e = 169 \Delta\sigma^2 = 3\%$	Mean	750	777	765	746
	Standard deviation	124	121	130	117
	5% quantile	–	595	570	550
	95% quantile	–	985	992	975
	Mean	946	973	977	948
Ft. Lewis (7 wells) $n_e = 77 \Delta\sigma^2 = 4\%$	Standard deviation	221	230	241	217
	5% quantile	–	630	632	610
	95% quantile	–	1385	1408	1335
	Mean	19.0	18.8	19.1	19.1
	Standard deviation	1.33	1.38	1.41	1.41
Rifle $n_e = 250 \Delta\sigma^2 = 1\%$	5% quantile	–	16.5	16.8	16.7
	95% quantile	–	21.0	21.5	21.5

<sup>a</sup>Parameters of TCE discharge at Ft. Lewis are in  $\text{g d}^{-1}$  and parameters of uranium discharge at Rifle are in  $\text{mg d}^{-1}$ .



**Figure 7.** Analogous to Figures 5a and 5b, but for data from seven wells only.

histogram of  $\bar{q}_{\text{PFM}} = 1.93 \text{ g}/(\text{m}^2 \text{ d})$ ,  $s_{\text{PFM}}^2 = 15.6 \text{ g}^2/(\text{m}^4 \text{ d}^2)$ ,  $b = 2.51$ , and a significantly reduced value of  $n_e = 77$  from equation (13). Although well spacing is even less uniform in this case than in the 10 well scenario above, application of Figure 3 with  $n_w = 7$  to both variogram structures of Figure 7b and using equation (14) leads to an excellent approximation of  $n_e \approx 80$  (this calculation is analogous to the procedure for the 10 wells above and not shown here). Step F results in a “missing variance”  $\Delta\sigma^2$  of  $\sim 4\%$  and validation step I provides relative differences between  $\bar{q}_{\text{PFM}}$  and  $\bar{q}_{Tc}$  and between  $\sigma_{ec}^2$  and  $\sigma_{ec}^2$  of  $\sim 4\%$  and  $7\%$ , respectively. The probability distributions of TCE discharge for this scenario are superimposed in Figure 6b and are seen to be in comparably good agreement similar to the 10 well scenario considered previously. The center part of Table 1 further confirms a good matching of the “target values” from block kriging. However, the degree of skewness of the discharge distributions is more pronounced than in the 10 well scenario. This is reflected by the right shift (with respect to mean) of the 5% and 95% quantiles as well as a slight increase in parameter  $a$  to 0.05. It is noted that declustered sample means ( $946$  versus  $750 \text{ g d}^{-1}$ ) and TCE discharge estimates from the seven and 10 well scenarios are significantly different. However, the confidence interval between the 5% and 95% quantiles of the seven well scenario almost fully contains the corresponding interval of the 10 well scenario.

### 5.1.3. Uncertainty Projection

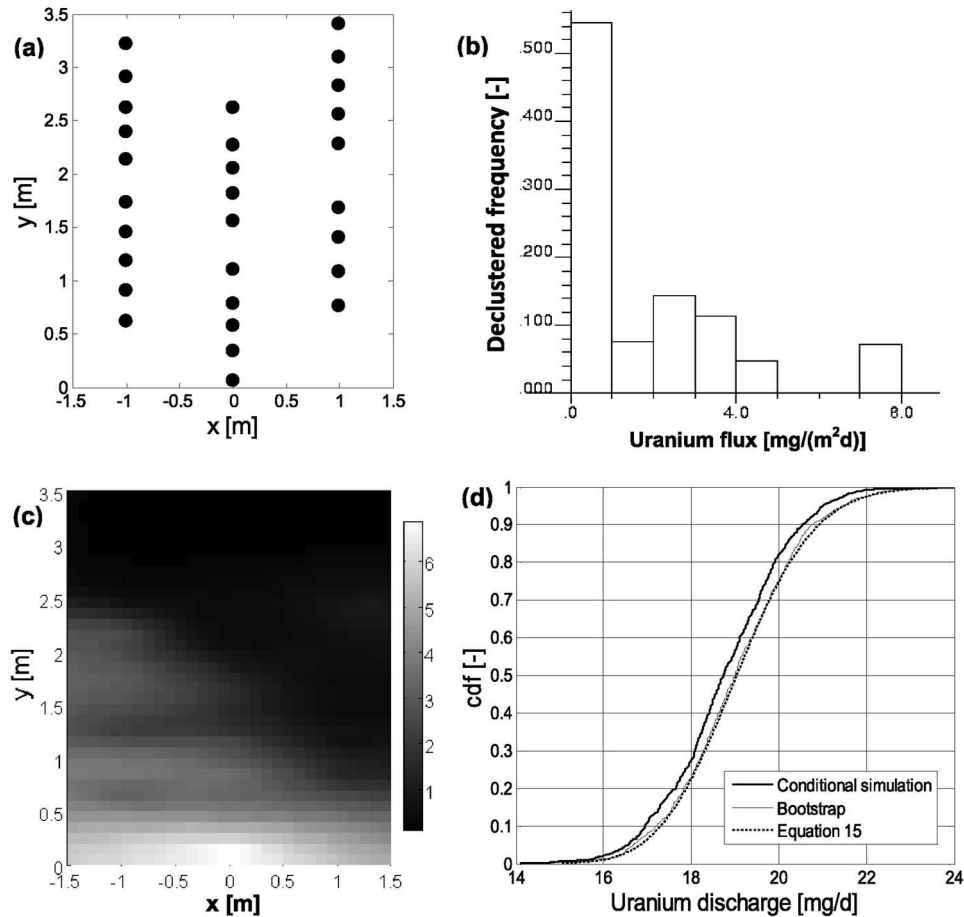
[38] Finally, the above methods for uncertainty quantification may also be adopted to assess the benefits of obtaining additional PFM measurements in terms of uncertainty reduction on the discharge estimate. This is termed “uncertainty projection” and is demonstrated by maintaining the scenario from above, where data is available from seven observation wells, and where it is intended to perform complementary measurements at three additional wells, such that the sampling pattern of Figure 1 is again complete. The principle is most easily illustrated by inspection of equation (15) revealing that any measure of dispersion or uncertainty (e.g., confidence interval width on both sides of the mean estimate or variance of discharge cdf) around a mean estimate  $\bar{q}_{\text{PFM}}$  is purely determined by the last term of equation (15). Assuming that additional sampling may change  $\bar{q}_{\text{PFM}}$ , but does not significantly alter declustered data variance, skewness, and spatial correlation structure, the last term in equation (15) becomes a pure function of  $n_e$

and may be determined for future sampling events knowing only the locations (and not actual values) of additional observations. In the present case, the projected value of  $n_e$  becomes 238 (from equation (13) with sampling locations and variogram from Figures 1 and 7b, respectively). Results correspond to the dashed line in the group of graphs denoted by “uncertainty projection” in Figure 6b. Since this cdf is independent of  $\bar{q}_{\text{PFM}}$ , it is arbitrarily shifted to  $\bar{q}_{\text{PFM}} = 0$ .

[39] In similar ways the bootstrap and stochastic simulation methods may be used for the same purpose. For the bootstrap approximation,  $n_e$  is simply increased to the projected value after additional sampling, and the declustered sample distribution, from which it is drawn, may be arbitrarily shifted (here to  $\bar{q}_{\text{PFM}} = 0$ ). The result is shown by the thin continuous line in the group “uncertainty projection” of Figure 6b. With the stochastic simulation approach conditional realizations cannot be generated as long as the additional data values are unknown. However, following section 2.1, unconditional realizations may be generated and fictitiously sampled at the existing and additional sample locations to obtain respective values of  $\bar{q}$  and  $q_T$  for each realization (equations (1) and (3)). From this, a distribution of the estimation errors  $e = \bar{q} - q_T$  may be built, which corresponds to the augmented sampling network. The respective result is also superimposed in Figure 6b (the bold continuous line in the group “uncertainty projection”) and shows good agreement with the  $t$  transformation and bootstrap methods as well as with the shape of the actual discharge distributions for the 10 well scenario.

## 5.2. Estimating Uranium Discharge at La Quinta Gallery, Rifle, CO

[40] At the La Quinta Gallery site in Rifle, CO, a former uranium mill resulted in the persistence of a groundwater uranium plume on the northern border of the Colorado River. Measurements of local uranium fluxes in a transect of  $X = 3 \text{ m}$  and  $Y = 3.5 \text{ m}$  are available from three wells ( $n = 29$ ;  $A = 10.5 \text{ m}^2$ ) as depicted in Figure 8a. By the same procedure as above (Figure 2), declustering weights are found by ordinary block kriging resulting in the histogram of Figure 8b with  $\bar{q}_{\text{PFM}} = 1.81 \text{ mg}/(\text{m}^2 \text{ d})$ ,  $s_{\text{PFM}}^2 = 4.4 \text{ mg}^2/(\text{m}^4 \text{ d}^2)$ , and  $b = 1.53$ . A conditional stochastic simulation (1000 realizations; 31 horizontal  $\times$  36 vertical grid elements) is used to obtain the spatial distribution of ensemble means of local fluxes in Figure 8c as well as the probabilistic uranium



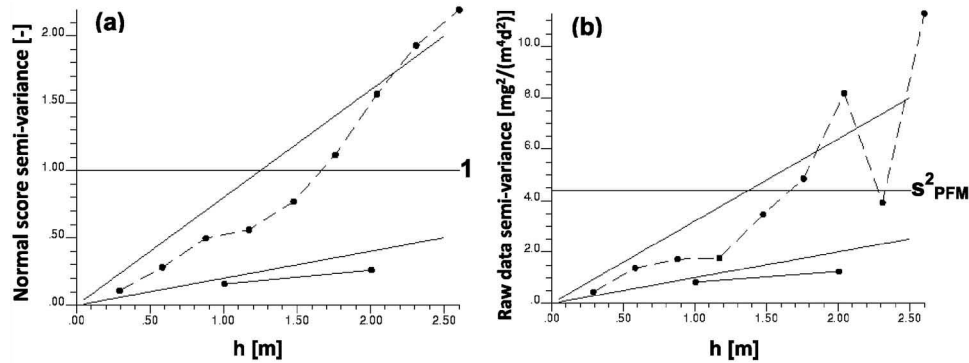
**Figure 8.** (a) PFM sampling locations over transect at Rifle, CO., with (b) declustered data histogram. (c) Distribution of ensemble means of local fluxes in  $\text{mg}/(\text{m}^2\text{d})$  from 1000 conditional realizations. (d) Uranium discharge distributions from the methods discussed.

discharge distribution of Figure 8d (thick continuous line). Simulation output is validated by step I, which results in relative differences between  $\bar{q}_{\text{PFM}}$  and  $\bar{q}_{Tc}$  and between  $\sigma_{ec}^2$  and  $\sigma_{ec}^2$  of  $\sim 1\%$  and  $8\%$ , respectively; the “missing variance”  $\Delta\sigma^2$  is  $\sim 1\%$ . Figure 8d also contains cdf’s from the two approximate methods and agreement in terms of quantile differences is seen to be within  $\sim 3\%$ . Selected statistics of the discharge distributions are again summarized in Table 1 (bottom part).

[41] The experimental normal score and raw data variograms (connected dots) with linear variogram model fits (simple lines) used in the analysis are depicted in Figures 9a and 9b, respectively. The slopes adopted are  $0.2$  and  $0.8 \text{ m}^{-1}$  for normal scores and  $1$  and  $3.2 \text{ mg}^2/(\text{m}^5 \text{ d}^2)$  for raw data (horizontal and vertical, respectively) resulting in  $n_e = 250$  from equation (13). In order to be able to apply Figure 3 for obtaining an approximate value of  $n_e$ , it is assumed that  $a_h/X$  and  $a_v/Y$  are much larger than one, such that the shape of the spherical variogram inside the transect is basically linear. Because the slopes are known and proportional to the reciprocals of the respective ranges, it is found that  $a_v/a_h = 1/3.2$ , leading to  $a_v X/(a_h Y) = 0.27$ , which is beyond the range of charts in Figure 3. However, Figure 3a for  $a_v X/(a_h Y) = 1$  with  $a_v/Y \gg 1$  may be used to arrive at a lower bound of  $n_e$  as  $\sim 200$ . Because of the small

amount of data available, fewer data pairs exist within the lag classes and experimental variogram values are less reliable (especially in the horizontal direction). For a reasonable histogram reproduction of the conditional simulation method it is again observed that  $\sigma_{D\lambda}^2 \approx 1$  (horizontal line in Figure 9a; step D in Figure 2) is an important criterion. While the latter is achieved to  $\pm 1\%$ ,  $\sigma_{D\lambda}^2/s_{\text{PFM}}^2 \approx 0.9$  (step E in Figure 2) is used as a compromise in the raw data variogram fit of Figure 9b to avoid larger discrepancies between experimental and theoretical variograms. The problem in this case seems to be related to the particular sampling configuration of Figure 8a, where the center well is shifted downward such that the largest data values at its bottom and the smallest data values at the top of the lateral wells (compare Figure 8c, which honors data values) receive the largest declustering weights. Thus, the declustering process does not affect the data mean very much, but does significantly inflate  $s_{\text{PFM}}^2$  to almost  $50\%$  above its raw value before declustering. The experimental variogram, in contrast, is computed in the classic way from unweighted data and does not account for this fact. Hence, variogram fits slightly above the experimental variogram points are justified in agreement with the requirements of steps D and E in Figure 2. The linearity of the variograms is a consequence of the smooth trend apparent in Figure 8c. The





**Figure 9.** Horizontal (continuous lines) and vertical (dashed lines) experimental variograms of (a) normal scores and (b) raw data with linear model fits for uranium fluxes at Rifle, CO.

point variance (sill)  $\sigma_0^2$  remains unknown, such that equations (6), (8), and (9) have to be evaluated in terms of the variogram  $\gamma(\mathbf{h})$  rather than  $C(\mathbf{h})$  (compare to the remark at the end of section 2.1). Further evident from Figure 8c is that the transect does not contain the entire plume such that results in terms of uranium discharge are limited to the transect studied, rather than the whole plume.

## 6. Evaluation of Assumptions and Further Observations

### 6.1. Modeling of a Spatial Trend

[42] The approach presented here is generally valid under the intrinsic hypothesis and no spatial trends are explicitly modeled. This may be contrary to the intuition that expected fluxes are higher near the plume core than at its periphery, for example. However, the fact that stochastic simulation is used with sufficient PFM data, such that all locations in a transect are effectively conditioned (correlated) to observed data, results in the automatic modeling of a spatial trend function. This means that the conditional simulation output is not stationary [Chiles and Delfiner, 1999] as is nicely reflected by Figures 6a and 8c showing the spatial distribution of ensemble means of local fluxes. An analogous argument applies to the stationarity of flux variance. Thus, the present work takes advantage of sufficient available data to implicitly account for spatial trends, which avoids the need for explicit trend modeling with respective uncertainties involved. This is further supported by a closely related discussion of trend modeling related to kriging by Journel and Rossi [1989]. However, if not all transect locations were correlated to one or more data locations, then a relatively small value of  $\Delta\sigma^2$  (e.g.,  $< 5\%$ ) would still be sufficient to perform the conditional simulation approach. The latter assures that data is representative of the flux distribution over the transect, which is what ultimately matters for discharge estimation (rather than the exact location and shape of the plume within the transect).

### 6.2. Uncertainties in Choice of Variogram Model and Parameters

[43] The present work does not account for uncertainties in the choice of variogram models and their parameters. If these uncertainties can be reliably quantified, then a Bayesian approach similar to that of Trolborg *et al.* [2010] may

be combined with the present one. However, the discussion related to Figure 3 gives an idea about how sensitive  $n_e$ , the fundamental parameter influencing discharge uncertainty, is with respect to changes (or uncertainties) in variogram ranges. Moreover, theoretical variogram models are not simply fitted to experimental variogram points. Rather, joint analysis of normal score and raw data variograms contributes to a reliable choice of variogram ranges and honoring data dispersion variances assists in fitting variogram sills or slopes (steps D and E in Figure 2). Eventually, if conditional simulation is performed, successful validation of the output (step I in Figure 2) may confer additional confidence to variogram models and parameters used. It is further noted in this context that the choice of variogram models used (e.g., spherical in Ft. Lewis case study versus a possible exponential fit) is rather arbitrary and not expected to significantly influence results. The reason for this is that what matters most for the present problem are integral properties of the variogram over the extent of the transect (compare equations (2), (4), and (5)), rather than its exact shape. Meeting these integral properties is assured by matching the dispersion variances in steps D and E. An extreme example is the linear variogram of the Rifle case study, which may be modeled by any variogram type that is linear near the origin and with appropriate and large enough sills and ranges (beyond the transect scale). As a further consequence of this, the present work also neglects the circumstance [Chiles and Delfiner, 1999] that the raw data and normal score variograms  $C(\mathbf{h})$  and  $C_{ns}(\mathbf{h})$  should be related depending on the properties of the normal score transform applied (i.e., they would generally not be of the same type as assumed here).

### 6.3. Uncertainty in Local PFM Measurements

[44] Measured contaminant fluxes by the PFM method are associated with a random measurement error (e.g., related to sampling and/or analysis), which is assumed negligible in this study. Because of its causes, measurement uncertainty is typically considered spatially uncorrelated and manifests as a nugget variance component in the data variogram [Kitanidis, 1997]. It may be accounted for in the conditional simulation approach by conditioning to uncertain data [Trolborg *et al.*, 2010] as well as in the ordinary block kriging system [Kitanidis, 1997]. For the case studies presented, especially from the more reliable vertical variograms

(Figures 5, 7, and 9), no significant nugget variance is apparent. This demonstrates the fact that flux measurements predominantly reflect spatial variability, which is over several orders of magnitude, rather than measurement errors within a certain percentage range. In addition to this and in agreement with a conclusion by *Cai et al.* [2011], measurement errors also undergo a process of averaging such that neglecting them is justified.

#### 6.4. Additional Information

[45] The present work assumes that the only data available are local PFM flux measurements inside a transect of interest. In principal, the conditional simulation approach using PFM measurements may be generalized to accommodate additional types of “indirect” information (e.g., hydraulic conductivity, head and/or concentration measurements, contaminant source properties, etc.) as demonstrated by *Schwede and Cirpka* [2010] or *Trolborg et al.* [2010]. However, along the lines of section 6.1, we argue that if  $\Delta\sigma^2$  is sufficiently small, i.e., if the plume is effectively sampled by the “direct” PFM flux measurements (in the sense that they are part of the discharge to be estimated), then the expected benefit of including additional “indirect” information is not very significant. This may especially be the case if rather complex nonlinear stochastic simulation and conditioning methods have to be adopted. A closely related advantage of working with flux data only is the complete linearity of the problem, which allows for using straightforward ordinary block kriging to define an effective number of independent data for use in combination with one of the simple approximation methods presented. Overall, it may be observed at this point that the methods presented herein are generally applicable to any linearly averaging quantity measured inside a domain of interest. This includes the estimation of water discharges from local water flux measurements (e.g., through PFM) as well as the estimation of mean contaminant concentration over a transect from MLS measurements (e.g., for alternative discharge estimation).

### 7. Summary and Conclusions

[46] The problem of determining probability distributions or confidence intervals for mean estimates of correlated skewed data is known in many fields of science and engineering. In the present work, estimation and uncertainty quantification of contaminant mass discharges in the groundwater from local contaminant flux data (based on passive flux meter, “PFM,” measurements) over a transect are discussed and demonstrated. Conditional stochastic simulation appears to represent the state-of-the-art method for this purpose. It requires a representative flux histogram over the transect, which is found through data declustering using ordinary block kriging weights (which account for spatial correlations as well as transect limits). In order to achieve reliable histogram reproduction over many realizations, the importance of considering data dispersion variances in the process of variogram-model fitting is highlighted. The declustered data histogram is assumed to correctly reflect the frequency of occurrence of local fluxes over the transect. As a consequence, its mean represents an unbiased estimate of the mean flux over the transect and should be honored by any discharge estimation method. Based on the

stochastic simulation paradigm, an effective number of independent data  $n_e$  is further introduced as an expected factor of variance reduction between the declustered data variance and the ordinary block kriging variance. As such,  $n_e$  should also be closely honored by conditional simulation results. Based on  $n_e$  we further develop two simple “approximate” methods for estimating discharge and uncertainty. One uses a transformation of the  $t$  statistic to account for skewness, while the other is closely related to the classic bootstrap method and uses the full declustered data histogram. Moreover, by means of  $n_e$  it is shown that one of the advantages of PFM measurements is the complete sampling of fluxes in the vertical direction, which is exactly the direction of typically larger spatial variability (shorter correlation lengths). This leads to less redundancy between data values and a higher degree of representativeness of flux variability in the plume.

[47] All three methods presented explore two fundamental properties of the problem: (1) linearity, i.e., the integral of local fluxes over the transect gives discharge, and (2) the data sufficiently well represents local flux variability over the transect. As a consequence of (1), the approaches are equally appropriate for estimating water discharges from PFM water flux measurements or for estimating average concentrations over an MLS transect with respective uncertainties. In order to evaluate (2), the “missing variance”  $\Delta\sigma^2$  is introduced and found to be smaller than 5% in the case studies presented. Large values of  $\Delta\sigma^2$  indicate that (2) may not be satisfied (whether an approximate limiting value exists and how large it is, is still an open question). In such a case, quantification of an estimation error variance may still be reliable, while the confidence in discharge quantile estimates becomes questionable. In order to overcome the latter, additional PFM sampling or application of more comprehensive (possibly nonlinear) approaches including complementary site data [*Schwede and Cirpka*, 2010; *Trolborg et al.*, 2010] may be required. Whenever linearity and data requirements are satisfied, the two approximate methods may be of significant practical value for estimating the probability distributions of discharges as they circumvent the need for implementing the much more complex tool of conditional stochastic simulation (compare Figure 2), and do not require the assumption of multi-Gaussianity of normal scores. Although not utilized in the present work, the latter also alleviates the need for normal score transformation, provided that other robust measures of spatial variability (e.g., variogram of logarithms of fluxes) are used for joint fitting of variogram ranges. Conditional simulation, in contrast, may be useful for additional validation of adopted variogram models (step I in Figure 2), and in addition to discharge distributions, it delivers spatial information about plume shape and location.

[48] All of the methods discussed are demonstrated for two different field sites with PFM data from TCE (Ft. Lewis, WA) and uranium (Rifle, CO) plumes. Besides the different types of contaminants, the two selected sites vary in transect size, number of observation wells, and spatial correlation structure. The latter may be the case because of different geological (aquifer heterogeneity) and contaminant source conditions as well as the scale of the study (entire plume versus a portion thereof). Results are based on 1000 conditional realizations and represented in the

form of spatial distributions of ensemble means of local fluxes as well as discharge probability distributions (cdf's). For both sites, quantile estimates of discharge from the different methods are seen to agree within  $\sim 7\%$ . As an additional hypothetical scenario, some observation wells at Ft. Lewis were ignored to illustrate the effect that having fewer data available will increase estimation uncertainty as well as the skewness of the estimation error distribution. The same hypothetical scenario is also used to successfully predict uncertainty reduction due to additional sampling by the three methods investigated.

## Appendix A: Weighted Data Dispersion Variance

[49] The weighted data dispersion variance  $\sigma_{D\lambda}^2$  is the ensemble mean (expectation) of the spatial variance  $s^2$  of the weighted (declustered) data histogram over many unconditional realizations, which are fictitiously sampled for fluxes  $q_i$  at the actual measurement locations. Substituting equation (3) into (7) and taking the square leads to

$$\begin{aligned} s^2 &= \sum_{i=1}^n \lambda_i \left( q_i - \sum_{j=1}^n \lambda_j q_j \right)^2 \\ &= \sum_{i=1}^n \lambda_i \left[ q_i^2 - 2q_i \sum_{j=1}^n \lambda_j q_j + \left( \sum_{j=1}^n \lambda_j q_j \right)^2 \right]. \end{aligned} \quad (\text{A1})$$

[50] Multiplying  $\lambda_i$  into the brackets and splitting up the summation gives

$$s^2 = \sum_{i=1}^n \lambda_i q_i^2 - 2 \sum_{i=1}^n \lambda_i q_i \sum_{j=1}^n \lambda_j q_j + \sum_{k=1}^n \lambda_k \left( \sum_{j=1}^n \lambda_j q_j \right)^2. \quad (\text{A2})$$

[51] Knowing that  $\left( \sum_{j=1}^n \lambda_j q_j \right)^2 = \sum_{i=1}^n \lambda_i q_i \sum_{j=1}^n \lambda_j q_j = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j q_i q_j$  further yields

$$s^2 = \sum_{i=1}^n \lambda_i q_i^2 - \left( 2 - \sum_{k=1}^n \lambda_k \right) \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j q_i q_j. \quad (\text{A3})$$

[52] Taking the expectation  $E[\ ]$  and with  $\sum_{k=1}^n \lambda_k = 1$  as imposed with equation (3) leads to

$$\sigma_{D\lambda}^2 = E[s^2] = \sum_{i=1}^n \lambda_i E[q_i^2] - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E[q_i q_j], \quad (\text{A4})$$

where the weighting constants are written outside the expectation operator. Using  $E[q_i^2] = \sigma_0^2 + \mu_0^2$  and  $E[q_i q_j] = C(\mathbf{u}_i - \mathbf{u}_j) + \mu_0^2$  (based on stationarity of process mean  $\mu_0$  and variance  $\sigma_0^2$ ) with  $\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j = 1$  finally results in

$$\sigma_{D\lambda}^2 = \sigma_0^2 - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{u}_i - \mathbf{u}_j), \quad (\text{A5})$$

where the last term is equal to  $\sigma_{\bar{q}}^2$  from equation (4).

## List of Variables

### Dimensionless

$n$	number of local flux data collected
$n_e$	effective number of independent data; numerical index “ $j$ ,” if present, denotes different variogram components
$n_w$	number of wells in transect
$m$	number of nested variogram components
$N$	number of realizations in conditional stochastic simulation
$i, j, k$	index variables
$\lambda_i$	data declustering weights; ordinary block kriging weights for estimating $q_T$
$t$	Student's $t$ variable
$t'$	cubic transformations of $t$ accounting for distribution skewness
$b$	weighted data coefficient of skewness; coefficient of skewness of PFM measured flux histogram after declustering
$a$	auxiliary variable
$T$	transect (spatial averaging domain for which discharge is sought)
$\Delta\sigma^2$	“missing variance”; portion of flux variability $\sigma_{DT}^2$ over transect not represented by declustered data variance $\sigma_{D\lambda}^2$

### Lengths and Areas

$x, y$	horizontal and vertical coordinates
$\mathbf{u}_{(i,j)}, \mathbf{v}$	spatial coordinate vectors of components $x$ and $y$
$\mathbf{h}$	spatial separation vector between two locations
$a_h, a_v$	horizontal and vertical correlation lengths of $q()$ ; numerical index “ $j$ ,” if present, denotes different variogram components
$X, Y$	transect length and depth
$A$	transect area

### Fluxes and Discharges

$q$	spatially random function representing local fluxes within transect
$\mu_0$	expectation (ensemble mean) of $q$ ; mean of pdf( $q$ )
$q_{i\text{PFM}}$	measured local fluxes from PFM field deployment
$q_{\text{ins}}$	measured local fluxes from PFM field deployment after normal score transformation
$\bar{q}_{\text{PFM}}$	Weighted mean of $q_{i\text{PFM}}$ ; ordinary block kriging estimate of $q_T$ from field data; mean of declustered PFM data histogram; obtained from equation (3) by using $q_{i\text{PFM}}$ instead of $q_i$
$q_i$	fictitiously sampled fluxes from unconditional realizations at actual PFM measurement locations
$\bar{q}$	weighted mean of $q_i$ ; ordinary block kriging estimate of $q_T$ for each unconditional realization
$q_T$	spatial mean of local fluxes over transect for each unconditional realization
$q_{Tc}$	spatial mean of local fluxes over transect for each conditional realization
$e$	estimation error (difference between estimate $\bar{q}$ and true value $q_T$ for each unconditional realization)
$\bar{q}_{Tc}$	mean of $q_{Tc}$ obtained from averaging over many conditional realizations

- $q_\alpha$   $\alpha$ -quantile of  $q_{Tc}$  ( $q_{Tc}$  which is not exceeded with probability  $\alpha$ )  
 $Q$  discharge across transect

### Flux and Normal Score Variances

- $C()$  spatial covariance function of  $q$ ; numerical index “ $j$ ,” if present, denotes different variogram components  
 $C_{ns}()$  spatial covariance function of normal scores (i.e., of local fluxes after transformation to standard normal distribution)  
 $\gamma()$  (semi) variogram of  $q$   
 $\sigma_0^2$  variance of  $q$ ; variogram sill if existing; numerical index “ $j$ ,” if present, denotes different variogram components  
 $\sigma_{0ns}^2$  same as  $\sigma_0^2$ , but for normal scores of fluxes  
 $\sigma_{\bar{q}}^2$  variance of  $\bar{q}$ ; ensemble variance of the declustered data mean; corresponds to calculating  $\bar{q}$  from equation (3) for each realization using unconditionally simulated data at the sampling locations and then taking the variance over many realizations  
 $\sigma_{q_T}^2$  variance of  $q_T$ ; ensemble variance of the transect mean of  $q$ ; corresponds to calculating  $q_T$  from equation (1) for each unconditional realization and then taking the variance over many realizations  
 $\sigma_{\bar{q}q_T}^2$  covariance between  $\bar{q}$  and  $q_T$ ; corresponds to calculating  $\bar{q}$  and  $q_T$  from equations (1) and (3) for each unconditional realization and then taking the covariance between them over many realizations  
 $\sigma_e^2$  variance of estimation error; corresponds to calculating  $e = \bar{q} - q_T$  for each unconditional realization and then taking the variance over many realizations; numerical index “ $j$ ,” if present, denotes different variogram components  
 $\sigma_{DT}^2$  dispersion variance of  $q$  over transect; ensemble mean of the spatial variance of local fluxes in the transect; corresponds to taking the variance of all  $q$  inside  $T$  for each unconditional realization and then averaging the variances over many realizations  
 $\sigma_{D\lambda}^2$  declustered data dispersion variance; ensemble mean of the spatial variance of the declustered data; corresponds to taking the variance of the declustered data, which are unconditionally simulated at the sampling locations, for each realization and then averaging the variances over many realizations; numerical index “ $j$ ,” if present, denotes different variogram components  
 $\sigma_{D\lambda ns}^2$  same as  $\sigma_{D\lambda}^2$ , but for normal scores of fluxes  
 $\sigma_{ec}^2$  variance of  $q_{Tc}$  from conditional realizations  
 $s^2$  weighted variance of  $q_i$ ; declustered data variance for each unconditional realization  
 $s_{PFM}^2$  weighted variance of  $q_{iPFM}$ ; variance of the PFM measured flux histogram after declustering; obtained from equation (7) by using  $q_{iPFM}$  instead of  $q_i$  and  $\bar{q}_{PFM}$  instead of  $\bar{q}$

[53] **Acknowledgments.** This research was partially funded by the first author’s fellowship of the Bahia State Science Foundation (FAPESB; DCR 0001/2009), Brazil, the U.S. National Science Program (award 0804134), the Environmental Remediation Science Program (ERSP), U.S. Department of Energy (grant DE-FG02-08ER64585), and the U.S. Department of Defense (project ER0831) under the Environmental Security Technology Compliance Program (ESTCP).

### References

- Annable, M. D., K. Hatfield, J. Cho, H. Klammler, B. Parker, J. Cherry, and P. S. C. Rao (2005), Field-scale evaluation of the passive flux meter for simultaneous measurement of groundwater and contaminant fluxes, *Environ. Sci. Technol.*, 39(18), 7194–7201.
- Beland-Pelletier, C., M. Fraser, J. Barker, and T. Ptak (2011), Estimating contaminant mass discharge: A field comparison of the multilevel point measurement and the integral pumping investigation approaches and their uncertainties, *J. Contam. Hydrol.*, 122(1–4), 63–75.
- Bockelmann, A., T. Ptak, and G. Teutsch (2001), An analytical quantification of mass fluxes and natural attenuation rate constants at a former gasworks site, *J. Contam. Hydrol.*, 53(3–4), 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(1–2), 97–121.
- Borden, R. C., R. A. Daniel, L. E. LeBrun, and C. W. Davis (1997), Intrinsic biodegradation of MTBE and BTEX in a gasoline-contaminated aquifer, *Water Resour. Res.*, 33(5), 1105–1115.
- Cai, Z., R. D. Wilson, M. A. Cardiff, and P. K. Kitanidis (2011), Increasing confidence in mass discharge estimates using geostatistical methods, *Ground Water*, 49(2), 197–208.
- Chen, L., and J. Shao (1999), Bootstrap minimum cost estimation of the average chemical concentration in contaminated soils, *Environmetrics*, 10(2), 153–161.
- Chiles J. P., and P. Delfiner (1999), *Geostatistics—Modeling Spatial Uncertainty*, 720 pp., Wiley-Interscience, N.Y.
- Deutsch, C. V. (2004), “A statistical resampling program for correlated data: spatial\_bootstrap,” Center for Computational Geostatistics Annual Report 6, Univ. of Alberta, Edmonton, available at [http://www.uofaweb.ualberta.ca/ccg/pdfs/2004%20401-Spatial\\_Bootstrap.pdf](http://www.uofaweb.ualberta.ca/ccg/pdfs/2004%20401-Spatial_Bootstrap.pdf).
- Deutsch, C. V., and A. G. Journel (1992), *GSLIB—Geostatistical Software Library and User’s Guide*, 340 pp., Oxford Univ. Press, N.Y.
- Efron, B., and R. J. Tibshirani (1998), *An Introduction to the Bootstrap*, 456 pp., CRC Press, N.Y.
- Einarson, M. D., and D. M. Mackay (2001), Predicting impacts of groundwater contamination, *Environ. Sci. Technol.*, 35(3), 66A–73A.
- Emery, X. (2007), Conditioning simulations of Gaussian random fields by ordinary kriging, *Math. Geol.*, 39(6), 607–623.
- Goltz, M. N., M. E. Close, H. Yoon, J. Huang, M. J. Flintoft, S. Kim, and C. Enfield (2009), Validation of two innovative methods to measure contaminant mass flux in groundwater, *J. Contam. Hydrol.*, 106(1–2), 51–61.
- Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, 496 pp., Oxford Univ. Press, N.Y.
- Hatfield, K., M. D. Annable, J. Cho, P. S. C. Rao, and H. Klammler (2004), A direct method for measuring water and contaminant fluxes in porous media, *J. Contam. Hydrol.*, 75(3–4), 155–181.
- Herold, M., T. Ptak, and P. Grathwohl (2008), A comparison of mass flow rate estimates based on point scale and integral measurements, *Grundwasser*, 13(4), 231–250 (in German).
- Interstate Technology and Regulatory Council (ITRC) (2010), Use and Measurement of Mass Flux and Mass Discharge, MASSFLUX-1, 154 pp., Integrated DNAPL Site Strategy Team, Washington, D.C., available at <http://www.itrcweb.org>.
- Jarsjö, 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(3–4), 107–134.
- Journel, A. G. (1994), Resampling from stochastic simulations, *Environ. Ecol. Stat.*, 1(1), 63–91.
- Journel, A., and C. J. Huijbregts (1978), *Mining Geostatistics*, 660 pp., Academic, London, U.K.
- Journel, A. G., and M. E. Rossi (1989), When do we need a trend model in kriging?, *Math. Geol.*, 21(7), 715–739.
- Kitanidis, P. K. (1997), *Introduction to Geostatistics: Applications to Hydrogeology*, 272 pp., Cambridge Univ. Press, Cambridge, U.K.
- Kitanidis, P. K., and K. F. Shen (1996), Geostatistical interpolation of chemical concentration, *Adv. Water Resour.*, 19(6), 369–378.
- Klammler, H., K. Hatfield, M. McVay, and J. Luz (2011), Approximate up-scaling of geo-spatial variables applied to deep foundation design, in *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 5, pp. 163–172, Taylor and Francis, Oxon, U.K. doi:10.1080/17499518.2010.546266.
- Kübert, M., and M. Finkel (2006), Contaminant mass discharge estimation in groundwater based on multi-level measurements: A numerical evaluation of expected errors, *J. Contam. Hydrol.*, 84(1–2), 55–80.

- Li, K. B., and L. M. Abriola (2009), A multistage multicriteria spatial sampling strategy for estimating contaminant mass discharge and its uncertainty, *Water Resour. Res.*, *45*, W06407, doi:10.1029/2008WR007362.
- 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.
- 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 contaminant site remediation: Cleanup goals and performance criteria, in *Groundwater Quality: Natural and Enhanced Restoration of Groundwater Pollution*, edited by S. F. Thornton and S. E. Oswald, pp. 571–578, IAHS Publication 275, Oxfordshire, U.K.
- Schwede, R. L., and O. A. Cirpka (2010), Stochastic evaluation of mass discharge from pointlike concentration measurements, *J. Contam. Hydrol.*, *111*(1–4), 36–47.
- 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*(11), 224A–230A.
- Teutsch, G., T. Ptak, R. Schwarz, and T. Holder (2000), Ein neues integrales Verfahren zur Quantifizierung der Grundwasserimmission: I. Beschreibung der Grundlagen, *Grundwasser* *4*(5), 170–175 (in German).
- Troldborg, M., W. Nowak, N. Tuxen, P. L. Bjerg, R. Helmig, and P. J. Binning (2010), Uncertainty evaluation of mass discharge estimates from a contaminated site using a fully Bayesian framework, *Water Resour. Res.*, *46*, W12552, doi:10.1029/2010WR009227.
- U.S. Environmental Protection Agency (EPA) (1998), Technical protocol for evaluating natural attenuation of chlorinated solvents in groundwater, Cincinnati, Ohio, available at <http://www.epa.gov/superfund/health/commmedia/gwdocs/protocol.htm>.
- U.S. Environmental Protection Agency (EPA) (2004), The DNAPL remediation challenge: Is there a case for source depletion?, 88 pp., Cincinnati, Ohio.
- Willink, R. (2005), A confidence interval and test for the mean of an asymmetric distribution, *Comm. Stat. Theory Methods*, *34*(4), 753–766.
- Yanagihara, H., and K. H. Yuan (2005), Four improved statistics for contrasting means by correcting skewness and kurtosis, *British J. Math. Statist. Psychol.*, *58*(2), 209–237.
- Zhou, X. H., and P. Dinh (2005), Nonparametric confidence intervals for the one- and two-sample problems, *Biostatistics*, *6*(2), 187–200.
- Zhu, J., and G. D. Morgan (2004), A nonparametric procedure for analyzing repeated measures of spatially correlated data, *Environ. Ecol. Stat.*, *11*(4), 431–443.
- 
- M. D. Annable and J. Cho, Department of Environmental Engineering Sciences, University of Florida, 217 Black Hall, Gainesville, FL 32611-6450, USA.
- S. A. Cabaniss, Department of Chemistry and Chemical Biology, University of New Mexico, MSC03 2060, Albuquerque, NM 87131, USA.
- J. A. Guimarães da Luz and H. Klammler, Department of Environmental Sciences and Sustainable Development, Federal University of Bahia, Rua José Seabra, Barreiras, Bahia, 47805-100 Brazil. (haki@gmx.at)
- K. Hatfield and M. Newman, Department of Civil and Coastal Engineering, University of Florida, 365 Weil Hall, Gainesville, FL 32611-6450, USA.
- A. Peacock, Microbial Insights, Inc., 2340 Stock Creek Boulevard, Rockford, TN 37853-3044, USA.
- J. Ranville and V. Stucker, Department of Chemistry and Geochemistry, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA.
- P. S. C. Rao, Department of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2045, USA.