

# Conditional stochastic analysis of solute transport in heterogeneous, variably saturated soils

Thomas Harter

Department of Land, Air, and Water Resources, University of California, Davis

T.-C. Jim Yeh

Department of Hydrology and Water Resources, University of Arizona, Tucson

**Abstract.** A method is developed for the conditional (Monte Carlo) simulation of steady state flow and transient transport from point sources in variably saturated porous media. It combines the geostatistical method, a linearized approximation of the soil water tension perturbation solution, and a finite element numerical model. The method is used to investigate the usefulness of conditional simulation for predicting solute transport under a variety of sampling network designs applied to a number of hypothetical soils. Saturated hydraulic conductivity data yield the largest reduction of conditional uncertainty in relatively wet soils with mild heterogeneities. In highly heterogeneous soils or under dry conditions, soil water tension data by themselves, taken at a sampling density of one to two correlation scales along the expected mean travel path, can greatly reduce prediction uncertainty about solute concentration. Parameter uncertainty about statistical properties of the independent random variables becomes less important as the number of conditioning data increases. However, even with a very high number of sampling data, uncertainty of predicted concentration levels remains significant.

## 1. Introduction

Stochastic analysis is a useful approach in assessing the uncertainty associated with predicting migration of contaminants in heterogeneous geologic media under variably saturated conditions. Unconditional numerical stochastic simulations have been introduced and compared to existing analytical techniques in a companion paper [Harter and Yeh, this issue]. These methods exploit the overall statistical information contained in a data set of those field parameters which constitute a hypothetical unsaturated hydraulic conductivity function, that is, saturated hydraulic conductivity,  $K_s$ , and soil pore size distribution parameter,  $\alpha$ . But they do not account for local, deterministic field measurements of  $K_s$  and  $\alpha$  (“direct” conditional information); nor can they account for important information contained in measured data related to  $K_s$  and  $\alpha$ , such as the soil water tension  $h$  (“indirect” conditional information).

Conditional stochastic analysis, which takes advantage of field data at specified locations, has been developed for and applied to a number of groundwater problems [Clifton and Neuman, 1982; Dagan, 1982, 1984; Delhomme, 1979; Graham and McLaughlin, 1989a, b; Rubin, 1991; Smith and Schwartz, 1981a, b; Zhang and Neuman, 1995a, b, c, d]. To date, however, no attempt has been made to analyze either unsaturated flow or unsaturated transport with conditional stochastic methods.

Conditional simulation of unsaturated flow and transport distinguishes itself from the conditional simulation of saturated systems. Prediction uncertainty depends not only on knowledge about saturated hydraulic conductivity but also on knowledge about entire functional relationships such as the unsaturated hydraulic conductivity function. Hence the infor-

mation content of various field data, that is, the degree to which prediction uncertainty is reduced by one type of field data or another, is expected to strongly depend on soil water tension (or soil moisture content).

The objectives of this paper are to develop an efficient conditional simulation algorithm, to investigate the role of both indirect information (soil water tension data,  $h$ ) and direct information (data of  $f$  and  $a$ , where  $f = \ln K_s$  and  $a = \ln \alpha$ ), and to analyze the role of sampling network design with respect to reducing uncertainties of conditional stochastic predictions of nonreactive solute transport under variably saturated conditions.

## 2. Conditional Monte Carlo Simulation: Methods

The general implementation of Monte Carlo simulation (MCS) has been described by Harter and Yeh [this issue]. Here we generate conditional realizations of  $f$  and  $a$  together with a conditional approximate solution  $h$  (conditional ASIGNing; see below). The realization of each of these three random space functions (RSFs) is passed to the flow and transport model (modified method of characteristics (MMOC)) [Yeh *et al.*, 1993], which computes steady state soil water tension through a finite element approximation of Richards equation, flux distribution through a finite element approximation of Darcy’s law, and transient solute transport by using a modified method of characteristics. After completing either 150 or 300 realizations (see Table 2), the appropriate statistical sample parameters are computed from the output of the MCS.

### 2.1. Generating Conditional Random Fields

The conditional random field generator developed for this study is based on geostatistical work by Matheron [1973], Jour-

nel [1974], *Journal and Huijbregts* [1978], and *Myers* [1982]. To conduct the conditional simulation, we assume that the unconditional RSFs  $f$ ,  $a$ , and  $h$  are Gaussian stationary processes. For notational convenience,  $f$ ,  $a$ , and  $h$  are combined into a RSF vector  $\mathbf{X}$ . The term “random field” is used to denote individual realizations of the RSF  $\mathbf{X}$ . We assume that the moments of  $f$  and  $a$  as well as the mean,  $H$ , of soil water tension,  $h$ , and the geometric mean of  $\alpha$ ,  $\Gamma$ , are known. The problem of parameter uncertainty is addressed later in this study.

An estimate of conditional mean  $\langle \mathbf{X} \rangle^k$  and conditional covariance  $\mathbf{E}_{\mathbf{X}}$  is obtained by simple cokriging [*Dagan*, 1982; *Myers*, 1982]. Available measurement data of  $f$ ,  $a$ , and  $h$  are combined in a vector of known data  $\mathbf{X}_1 = (\mathbf{X}_{1f}, \mathbf{X}_{1a}, \mathbf{X}_{1h})^T$ , where  $T$  indicates transpose, and subscripts  $f$ ,  $a$ , and  $h$  indicate the type of data. Covariance matrix  $\mathbf{C}_{11} = \langle \mathbf{X}_1 \mathbf{X}_1^T \rangle$  is computed using appropriate covariance and cross-covariance functions as described below. Denoting the vector of unknown data  $\mathbf{X}_2 = (\mathbf{X}_{2f}, \mathbf{X}_{2a}, \mathbf{X}_{2h})^T$  and defining an unconditional covariance matrix  $\mathbf{C}_{12} \equiv \langle \mathbf{X}_1 \mathbf{X}_2^T \rangle$ , the (simple) cokriging system of equations is

$$\mathbf{C}_{11} \mathbf{\Lambda}_{12} = \mathbf{C}_{12} \quad (1)$$

where  $\mathbf{\Lambda}_{12}$  is the matrix of cokriging weights, from which we obtain  $\langle \mathbf{X}_2 \rangle^k$  through the cokriging equation

$$\langle \mathbf{X}_2 \rangle^k = \langle \mathbf{X}_2 \rangle + \mathbf{\Lambda}_{21} (\mathbf{X}_1 - \langle \mathbf{X}_1 \rangle) \quad (2)$$

where  $\langle \mathbf{X}_1 \rangle$  and  $\langle \mathbf{X}_2 \rangle$  are unconditional ensemble means of  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , and  $\mathbf{\Lambda}_{21} \equiv \mathbf{\Lambda}_{12}^T$ . The “estimation error covariance,” or conditional covariance,  $\mathbf{E}_{22}$  at nonmeasured locations corresponds to the simple cokriging covariance and is given by

$$\mathbf{E}_{22} = \mathbf{C}_{22} - \mathbf{\Lambda}_{21} \mathbf{C}_{12} \quad (3)$$

Note that individual entries in the conditional or error covariance matrix  $\mathbf{E}_{22}$  are equal to or less than entries in the unconditional covariance matrix  $\mathbf{C}_{22}$ ; that is, conditional variability decreases with additional information. At measurement locations, error variance is zero, provided that measurement errors are negligible.

Conditional simulation of  $\mathbf{X}$  with moments (2) and (3) proceeds as follows. After first generating unconditional random fields of  $f$  and  $a$  with a spectral random field generator [*Gutjahr*, 1989] and determining an unconditional solution  $h$ , we define the unconditional random field  $\mathbf{X}_s \equiv (\mathbf{X}_{sf}, \mathbf{X}_{sa}, \mathbf{X}_{sh})^T$ . Conditional random fields  $\mathbf{X}_s^c$  that are consistent with (1)–(3) are constructed through

$$\mathbf{X}_s^c = \langle \mathbf{X} \rangle^k + (\mathbf{X}_s - \langle \mathbf{X}_s \rangle^k) = \langle \mathbf{X} \rangle^k + \mathbf{e}_s \quad (4)$$

where  $\langle \mathbf{X}_s \rangle^k$  is a cokriged estimate of  $\mathbf{X}$  given “measurement” data  $\mathbf{X}_{1s}$ .  $\langle \mathbf{X}_s \rangle^k$  is the simulated equivalent to  $\langle \mathbf{X} \rangle^k$ : It preserves data  $\mathbf{X}_{1s}$  in the unconditionally generated random fields at and only at locations  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ , where measurements are available in the real field site as well. It gives cokriged estimates  $\langle \mathbf{X}_{2s} \rangle^k$  at all other locations  $\{\mathbf{x}_{m+1}, \dots, \mathbf{x}_n\}$  given unconditionally simulated data  $\mathbf{X}_{1s}$  from known data locations. The term  $(\mathbf{X}_s - \langle \mathbf{X}_s \rangle^k)$  is a realization  $\mathbf{e}_s$  of the estimation error incurred by estimating simulated data  $\mathbf{X}_s$  through kriged values  $\langle \mathbf{X}_s \rangle^k$ . Simulated error  $\mathbf{e}_s$  is added to the originally estimated conditional mean  $\langle \mathbf{X} \rangle^k$  to obtain one possible conditional random field  $\mathbf{X}_s^c$ . It can be shown that the simulated estimation error  $\mathbf{e}_s$  has the same conditional moments (3) as the real estimation error  $\mathbf{e} = (\mathbf{X} - \langle \mathbf{X} \rangle^c)$  because unconditional

probability density functions (pdfs) of real and simulated fields are identical (neglecting the possibility of measurement and moment estimation errors), and because conditioning occurs at the same locations at the field site and in the simulations [*Journal*, 1974; *Delhomme*, 1979]. For a large number of samples thus obtained the sample mean and sample variance of  $\mathbf{X}_s^c$  will converge in the mean square to the true conditional mean and variance, that is, the kriging estimate (2) and kriging variance (3) of  $\mathbf{X}$ .

In our model, following the approach by *Gutjahr et al.* [1992], (4) is rearranged and kriging equation (2) is used to explicitly write kriged terms in (4) such that

$$\mathbf{X}_{2s}^c = \mathbf{X}_{2s} + \mathbf{\Lambda}_{21} (\mathbf{X}_1 - \mathbf{X}_{1s}) \quad (5)$$

where  $\mathbf{X}_{2s}$  is the unconditionally generated mean removed realization of  $f$ ,  $a$ , and  $h$ ;  $\mathbf{X}_{1s}$  is the array of unconditionally simulated data at the particular locations, where measurements of the same variable are available in the field site (also unconditional mean removed); and  $\mathbf{X}_{2s}^c$  is the conditional mean removed realization  $f^c$  or  $a^c$ . Equation (5) is evaluated once for each realization of each RSF.

## 2.2. Conditional ASIGNing

If head measurements are part of  $\mathbf{X}_1$  (indirect conditioning), unconditional realizations of  $h$  must be computed from unconditional realizations of  $f$  and  $a$  to fill the one-dimensional array  $\mathbf{X}_{1s}$ . Conditional flow simulation therefore requires that the unsaturated flow equation be solved twice: once to obtain the unconditional random field  $h$  from the unconditionally generated realizations  $f$  and  $a$ , and, eventually, a second time to obtain the conditional nonlinear solution  $h^c$  from the conditional realizations  $f^c$  and  $a^c$ .

Normally, unconditional head solutions  $h$  would be computed using standard finite difference or finite element models. For steady state unsaturated conditions, an efficient method is ASIGNing [*Harter and Yeh*, 1993], which combines the spectrally derived first-order, linear approximation of head,  $h_L$ , with the finite element model MMOC2. For this study, an even more efficient approximation is chosen by applying the linear approximation  $h_L(f, a)$  directly to fill  $\mathbf{X}_{1s}$ . A linear head solution  $h_L$  is cogenerated with random fields  $f$  and  $a$  by explicitly solving for the spectral representation of  $h$ ,  $dZ_h(\mathbf{k}) = f(dZ_f(\mathbf{k}), dZ_a(\mathbf{k}), H, \Gamma)$  and applying a numerical Fourier transform [*Harter*, 1994]. Using a linearized unconditional solution  $h_L$  in the conditioning process is consistent with the linear estimation procedure (5).

Moreover, a conditional linear approximation  $h_L^c$  of the conditional head solution is obtained by estimating  $h^c$  with (5), that is, by including  $h_L(f, a)$  into  $\mathbf{X}_{2s}$ . Then  $h_L^c = \mathbf{X}_{2s}^c$  in the array  $\mathbf{X}_{2s}^c$ , in which  $h_L^c$  is used as the initial approximation of the nonlinear finite element solution  $h_{FE}^c$  to the unsaturated flow equation given the conditional realizations  $f^c$  and  $a^c$ . This is in principle equivalent to the ASIGNing procedure for unconditional flow simulations and is therefore called conditional ASIGNing. A flow chart of conditional ASIGNing and MCS is given in Figure 1.

Conditional random fields  $h_L^c$  obtained from (5) are accurate enough to allow the numerical algorithm to converge efficiently to the conditional finite element solution  $h_{FE}^c(f^c, a^c)$ . In this work, conditional ASIGNing has successfully and consistently been applied to soils with  $\sigma_y^2 \leq 3.2$ , where  $\sigma_y^2$  is the unconditional variance of the logarithm  $y = \ln K$  of the unsaturated

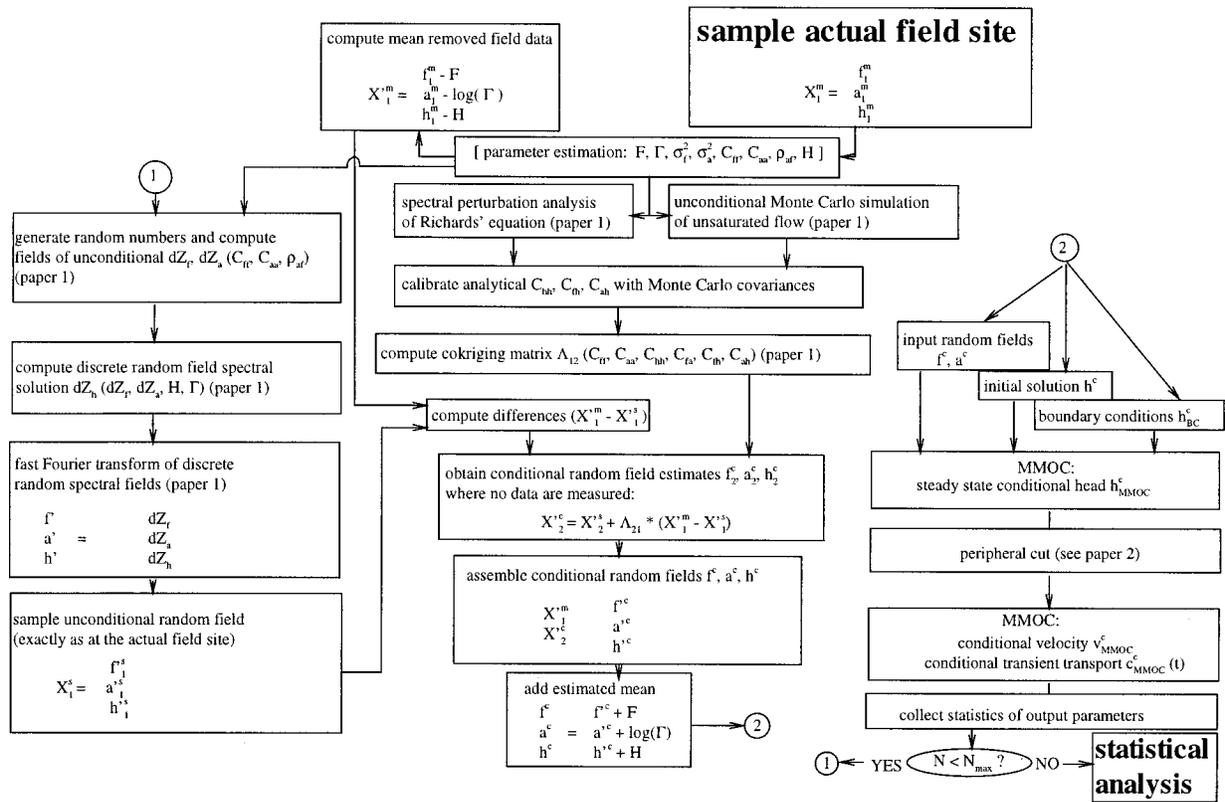


Figure 1. Schematic overview of flowchart of the conditional simulation technique including conditional ASIGNing. “Paper 1” refers to that by Harter [1994]; “paper 2” refers to that by Harter and Yeh [this issue].

hydraulic conductivity  $K$ . Use of the spectrally derived linear head solution has two advantages in the conditioning algorithm (5). It allows for a very efficient evaluation of the unconditional head field needed to obtain  $X_{1s}$  (increase in efficiency over use of FE solutions by 3 orders of magnitude). Further, by including it into  $X_{2s}$ , an initial approximation of the conditional head field is obtained such that the conditional finite element solution converges approximately 2 orders of magnitude faster than without such an initial approximation. The total computational savings are so enormous that an entire conditional Monte Carlo simulation of unsaturated steady state flow with several hundred realizations can be carried out more efficiently CPU-wise than a single conditional realization based on finite element solutions alone (i.e., without using the linear, spectral head solution). Because computational efforts to generate conditional random fields and conditional initial head solutions are minor compared to the CPU time for numerical flow and transport computations, the actual CPU time for a single realization (including conditional random field generation, flow, and transport computation) is only slightly more than that of unconditional realizations.

### 2.3. Covariances and Cross Covariances for the Cokriging Matrix $\Lambda_{12}$

Cokriging system (1) has a solution only if  $C_{11}$  is a positive definite matrix [Journel and Huijbregts, 1978]. In kriging (direct estimation) positive definiteness is assured by fitting field data to a valid functional form of the covariance, such as the exponential, spherical, or Gaussian models [Isaaks and Srivastava, 1989]. In cokriging (indirect estimation),  $C_{11}$  contains cross-covariance terms. An empirical method to estimate cross-

covariances is described by Myers [1982]. In our problem, covariance functions of  $f$  and  $a$  as well as mean head,  $H$ , and geometric mean of  $\alpha$ ,  $\Gamma$ , are assumed to be known; hence cross-covariance functions between  $f$  and  $h$ ,  $C_{fh}$ , and between  $a$  and  $h$ ,  $C_{ah}$ , as well as the covariance function of  $h$ ,  $C_{hh}$ , in the covariance matrices  $C_{11}$  and  $C_{12}$  can be determined by first-order stochastic analysis or by implementation of MCSs of unsaturated flow [Harter, 1994].

We found that analytically derived covariance and cross-covariance functions yield invertible (positive definite) matrices  $C_{11}$  and  $C_{12}$ , but may have significant errors in their magnitude, as shown by Harter [1994]. On the other hand, covariance matrices,  $C_{11}$  and  $C_{12}$ , derived from MCSs are stochastically accurate to within the sampling error of such simulations, but they may not be invertible, particularly if head and flux variances are very large, owing to inherent random fluctuations in the sample (cross-) covariance fields. For our conditional simulation we therefore use a modified analytical (cross-) covariance solution based on calibration of the invertible analytical (cross-) covariances against numerically determined sample (cross-) covariance functions: Analytical functions are multiplied by a correction factor (defined separately for each (cross-) covariance function and for each example soil) such that modified analytical solutions match nonlinear MCS sample (cross-) covariance functions with minimal error. Calibration of analytical cross-covariance functions is implemented by visual matching. Head covariance is calibrated such that the modified analytical solution for the variance exactly matches the numerically determined variance. Preliminary experiments were implemented and it was found that results

**Table 1.** Input Parameters for the Various Hypothetical Soil Field Sites Used in the Conditional Transport Analysis

Site	General Description	$\sigma_f^2$	$\sigma_a^2$	$\rho$	$\Gamma$	$H$	$\Delta x$	$\Delta z$	$\lambda_x$
3	$\sigma_v^2 = 0.85$ ; wet, isotropic	1.0	0.01	0	0.0	-1.5	0.1	0.1	0.5
12	$\sigma_v^2 = 0.53$ ; wet, anisotropic	...	...	1	...	...	0.3	...	3.0
21	$\sigma_v^2 = 3.2$ ; dry, anisotropic	...	...	1	...	-30.0	0.3	...	3.0
15	$\sigma_v^2 = 1.5$ ; dry, anisotropic	...	...	...	...	-10.0	0.3	...	3.0
28	$\sigma_v^2 = 1.8$ ; wet, anisotropic	2.25	0.04	...	...	...	0.3	...	3.0

Here  $\sigma_f^2$ , variance of  $f = \log K_s$  (log: natural logarithm);  $\sigma_a^2$ , variance of  $a = \log \alpha$ ;  $\rho_{af}$ , correlation coefficient between  $f$  and  $a$ ;  $\Gamma$ , geometric mean of  $\alpha$  [ $\text{m}^{-1}$ ];  $\Delta x$ , horizontal discretization of finite elements [m];  $\Delta z$ , vertical discretization of finite elements [m];  $\lambda_{fx}$ , horizontal correlation length of  $f$  [m]. Vertical correlation length of  $f$  is 50 cm in all example soils. Where not otherwise indicated, parameter values are identical to those of 3.

from the Monte Carlo simulation changed insignificantly within the range of potential error in the calibration procedure. However, significant changes in the shape of the conditional mean and variance contour lines of solute concentration were observed in one example (soil 15), if the correction factor for  $C_{ah}$  was doubled. While not a rigorous sensitivity test, this shows that it is important to adjust the analytical (cross-) covariances to match those obtained from MCS. Further research is needed to quantify exactly potential calibration errors.

#### 2.4. Nodal and Elemental Properties in the Finite Element Model Versus Grid Properties in the Spectral Random Field Generator

In those finite element realizations that serve as hypothetical field sites for this study (see below), head,  $h$ , and concentration,  $c$ , are nodal values, while saturated hydraulic conductivity,  $K_s$ , and pore size distribution parameter,  $\alpha$ , (and hence  $f$  and  $a$ ) are element properties. In contrast, the spectral random field generator and conditioning algorithm assume identical grid size and support for all variables. For the purpose of conditional simulation it is simply assumed that the support scales of both "measured" nodal and "measured" elemental properties are identical, and that the lower left node of each element has the same support and location as the element itself. This introduces a small error in the computation of cross covariances (which are functions of support and of distances between measurement points). The error is negligible since the element discretization is rather small compared to the correlation scale. To be consistent, the subsequent assignment of nodal and elemental properties in the finite element model from the conditional random field realizations  $f^c$  and  $a^c$  and the initial head  $h^c$  follows the reverse order: The  $f$ ,  $a$ , and  $h$  values at the  $i$ th column in the  $j$ th row of the conditional random fields are assigned to the  $i$ th element in the  $j$ th element row ( $f$ ,  $a$ ) and to the  $i$ th node in the  $j$ th nodal row ( $h$ ), which is the lower left node to the  $i$ th element in the  $j$ th element row.

### 3. "Field Test Sites" and Sampling Strategies: Methodology

#### 3.1. "Field Test Sites"

The so-called field sites that are investigated here are computer-generated hypothetical soil cross sections. In these artificial field sites, "field" hydraulic properties and the movement of the contamination plume can be perfectly sampled. Physical processes governing flow and transport of the so-called real plume and the heterogeneous properties of the soil are per-

fectly known. Measurement errors, parameter estimation errors, and upscaling problems are neglected for the moment. Only such computer-generated field sites allow a rigorous analysis of the information content of measurement data that can be retrieved through conditional stochastic simulation.

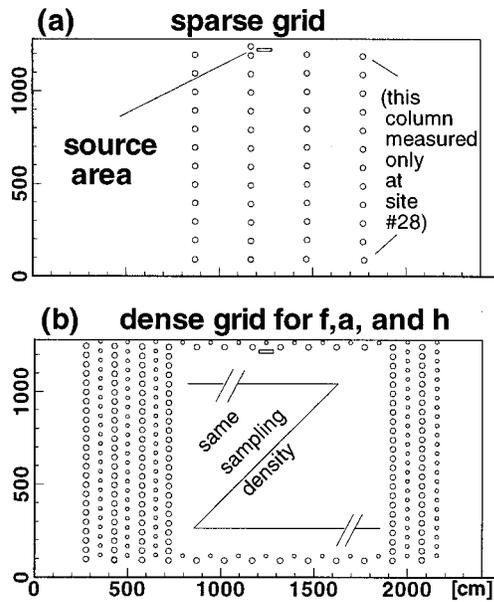
Five synthetic field sites were selected to evaluate effects of conditioning using relevant field data. Table 1 summarizes the characteristics of the five field sites. Detailed descriptions of these types of field sites are reported by *Harter and Yeh* [1995]. Each field site is a single, randomly chosen realization from an unconditional MCS. Four of the sites are anisotropic with  $\nu = \lambda_{fx}/\lambda_{fz} = 6$ , where  $\lambda_{fx}$  and  $\lambda_{fz}$  are correlation scales of  $f$  in the horizontal and vertical directions, respectively. One site is isotropic. Mean soil water tension ranges from -1.50 to -30.0 m. Together with the range of textural variability the resulting log unsaturated hydraulic conductivity variance at the five sites ranges from 0.5 to 3.2.

#### 3.2. Sampling Strategies

We assume that one, two, or all three types of measurements are available:  $K_s$  measurements; measurements that determine the unsaturated hydraulic conductivity function, that is,  $K_s$  and  $\alpha$  measurements; and measurements of soil water tension, which is assumed to be at steady state. For practical purposes, steady state flow conditions are achieved when temporal variability in  $h$  is much smaller than spatial variability in  $h$ , generally at depths of at least one to several meters depending on soil type and climate. The method is therefore strictly applicable only where the thickness of the unsaturated zone is considerable, for example, in arid and semiarid climates. It is under such conditions that assessment of unsaturated solute transport is particularly important.

Two basic sampling networks were designed for each of the three physical variables  $f$ ,  $a$ , and  $h$ : a "sparse" sampling network and a "dense" one (Figure 2). The sparse sampling network (Figure 2a) consists of measurement locations along three vertical "boreholes" (four "boreholes" at site 28) near the plume source spaced one horizontal  $\lambda_{fx}$  apart, with measurements at every  $2\lambda_{fz}$  depth interval. No data are sampled from an area within  $2\lambda_{fz}$  of the bottom boundary of the simulation domain. The center column intersects the source area of the solute plume, where five additional measurements are made. The total number of data points in the sparse network is 41 (53 at site 28).

A dense sampling network (Figure 2b) consists of double the data density of the sparse sampling network, that is,  $0.5\lambda_{fx}$  in the horizontal and  $1\lambda_{fz}$  in the vertical. Also in the dense sampling network, data are sampled throughout the entire



**Figure 2.** (a) Sparse measurement network and (b) dense measurement network for the RSFs  $f$  and  $a$  (large circles) and for  $h$  (small circles). The source area is indicated by the rectangular box near the top of the simulation domain.

simulation domain, resulting in a total of more than 300 data points. Only an area within  $2\lambda_{fx}$  of the vertical boundaries is excluded from data sampling. Measurements of  $K_s$  and  $\alpha$  are obtained at identical locations. The sampling grid for head measurements is shifted relative to the sampling grid of  $K_s$  and  $\alpha$  such that a head measurement point is at the center between four adjacent  $K_s$  measurements. The dense sampling network also includes measurements at all nodes or elements within and immediately adjacent to the contamination source, that is, the properties of the source area are perfectly known in the dense sampling network. Monte Carlo simulations are implemented with various combinations of  $f$ ,  $a$ , and  $h$  sampling networks as listed in Table 2.

**4. Effects of Conditioning on Unsaturated Flow**

Effects of conditioning on predicted soil water tension and flow regime are demonstrated in Figure 3 using results of two conditional simulations, A and G (Table 2), of field site 28. This field site, used as a “real world” analog, is a vertical cross section of 12.8 m depth and 24 m width. The soil is anisotropic with correlation scales for  $f$  and  $a$  of 3 m horizontally and 0.5 m vertically. Unsaturated hydraulic conductivity variance,  $\sigma_y^2$ , is 1.8. Soil water tension distribution and stream lines at this “real” site are shown in Figure 3a. In the center of the cross section a wide but relatively thin lens of near zero soil water tension overlies a dry area with tensions of about  $-2.50$  m. The dry area reflects a flow barrier which causes flow at and above the thin wet lens to be predominantly horizontal. Also illustrated in this figure is a distinct diagonal flow channel immediately beneath the contamination source (rectangular box near the top of the illustrated flow domain).

Conditional simulation A is based on a dense data network of all three variables  $f$ ,  $a$ , and  $h$ , while G is based on only 53 soil water tension data (sparse network) from four tensiometer nests. The conditional mean soil water tension fields and

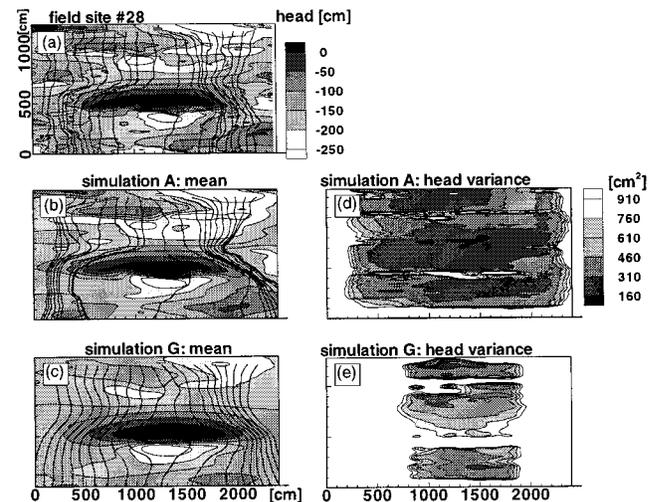
**Table 2.** Classification of the Conditional Simulation Types

Conditional Simulation Type	Sampling Density			Realizations
	$f$	$a$	$h$	
A	dense	dense	dense	150
B	dense	...	dense	150
C	dense	...	...	150
D	sparse	sparse	...	300
E	sparse	sparse	dense	300
F	...	...	dense	300
G	...	...	sparse	300
H	sparse	...	dense	300

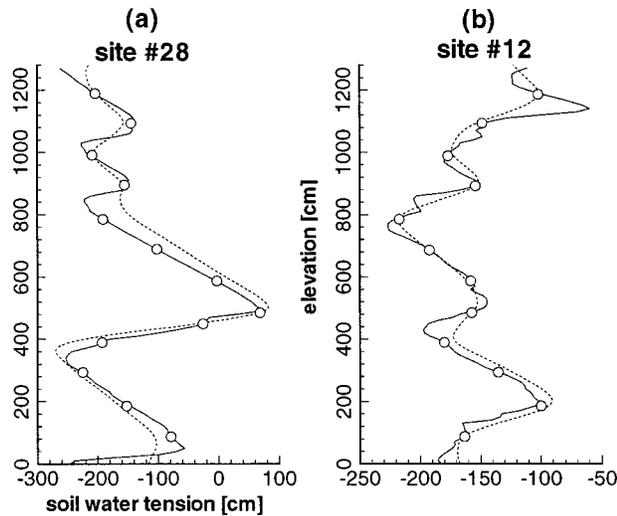
Conditional Simulation Type	Description	Realizations
I (as H)	$\sigma_f^2 = 1.8;$ $\sigma_a^2 = 0.24;$ $\langle \log \alpha \rangle = 0.6$	300
J (as H)	$\sigma_f^2 = 1.2$ $\sigma_a^2 = 0.16;$ $\langle \log \alpha \rangle = -0.6$	300
K (as H)	$\sigma_f^2 = 1.8;$ $\sigma_a^2 = 0.24;$ $F = 0.5$	300
L (as H)	$H = -1.4$ m	300

Conditional simulations A through H are based on different sampling networks for the parameters  $f$ ,  $a$ , and  $h$ . Simulations I through L are applied to field site 28 to assess the effect of erroneous statistical input parameters.

stream lines obtained from simulations A and G are depicted in Figures 3b and 3c, respectively. The main features of the flow regime at this site, as discussed above, are well predicted by these two conditional simulations. A comparison of the results shows that with decreasing number of conditioning data, conditional mean soil water tension predictions are more uniform, while the variance increases (Figures 3d and 3e). Nevertheless, the overall flow regimes in the center of the two simulation domains are quite similar, manifesting the value of



**Figure 3.** (a) Actual distribution of  $h$  and streamlines at field site 28. Mean head and flow lines resulting from mean flux for conditional simulation (b) A and (c) G. The corresponding conditional head variances are shown in Figures 3d and 3e.



**Figure 4.** Actual pressure head profile (solid lines) for a vertical cross section at the horizontal location of the middle “borehole” of simulation G (next to the source area; see Figure 2a). Circles represent the measured data used for conditioning in simulation G. The resulting conditional head profile is shown in dashed lines. Site 28 has  $\sigma_y^2 = 1.8$ , while site 12 has  $\sigma_y^2 = 0.5$ .

soil water tension information in the conditional simulation of flow in unsaturated zones.

Note that head variance does not vanish at locations of soil water tension measurements, although head values at those locations are perfectly known. This is an artifact of the linearized conditioning procedure [Kitanidis and Vomvoris, 1983; Peck et al., 1988; Yeh et al., 1993; Gutjahr et al., 1994]. Conditional realizations  $f^c$  and  $a^c$  are obtained from head data (among others) through linear estimation (4), while conditional realizations of  $h^c$  are computed by solving the nonlinear flow equation.

Although field data of soil water tension are not preserved exactly, the conditional simulation technique generally gives satisfactory results. For example, in conditional simulation A of field site 28, head variance in most locations with tensiometer data is reduced to less than  $0.015 \text{ m}^2$  (Figure 3d) from  $0.49 \text{ m}^2$  in the unconditional simulation [Harter, 1994]. In conditional simulation G of field site 28, the minimum variance is slightly less than  $0.03 \text{ m}^2$  (Figure 3e). Head conditioning is the least accurate in areas with steep head gradients, for example, in the center of the simulation domain between the very wet and very dry areas mentioned above (illustrated in the cross section of Figure 4a). In simulations of field sites with moderate  $\sigma_y^2$  ( $\sigma_y^2 < 1$ ), differences between measured and mean conditional head decrease to almost negligible values (Figure 4b).

## 5. Impacts of Sampling Network Design on Concentration Prediction

### 5.1. Field Site Plume Movement

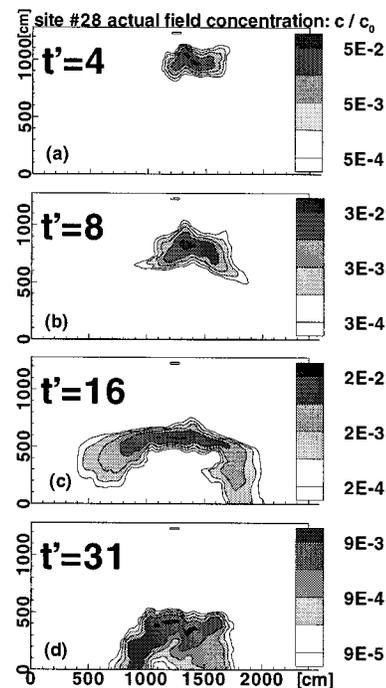
Corresponding plume dynamics at site 28 are depicted in Figures 5a–5d for dimensionless times  $t' = 4, 8, 16,$  and  $31$ . Initially, the plume moves diagonally downward along the previously mentioned flow channel (Figure 3a). The tip of the plume splits into two at or before  $t' = 8$  and spreads horizontally as it reaches a large wet area located above a relatively

dry lens in the center of the simulation domain (see previous section). At about  $t' = 31$  the center of the plume has traveled beyond the bottom of the simulation domain, and residual concentration is found primarily within and underneath the dry, low-permeability area. The plume shape is distinctly non-Gaussian, with no tendency toward a more Gaussian behavior even at late times. The following analysis will concentrate on the conditional simulation results at time  $t' = 16$ , when lateral plume spreading is relatively large. Conditional moments of plume spreading and breakthrough curves are often found to be ambiguous due to their integral character [Harter, 1994]. The analysis therefore focuses on mean concentration,  $\langle c(\mathbf{x}) \rangle$ , and concentration coefficient of variation,  $CV_c(\mathbf{x})$ . Together these two empirical functions best reflect the conditioning effects.

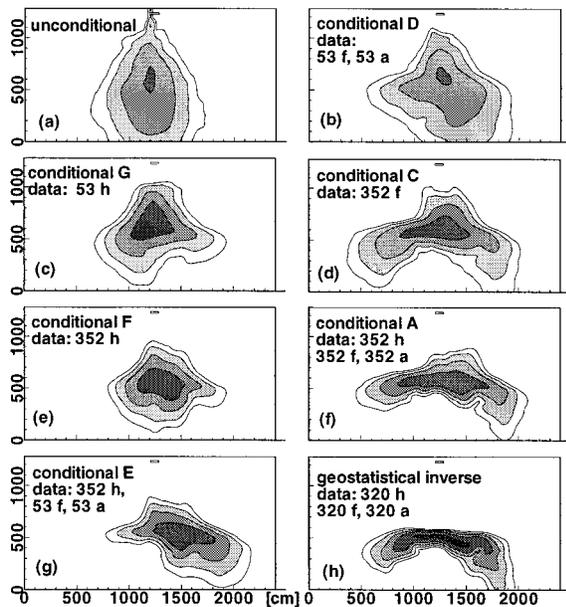
### 5.2. Sampling Soil Water Tension Only (Simulations G and F)

Sparse sampling network G provides the least information for conditional simulation. Nevertheless, 53 tensiometer data produce a significant improvement in mean concentration prediction (Figure 6c) compared to the unconditional mean concentration prediction (Figure 6a). Peak concentrations are significantly higher than in the unconditional simulation. Estimates of both plume front and plume tail are more realistic than in the unconditional simulation, albeit far from accurate.

If the number of tensiometers is doubled in the vertical and horizontal directions and extended over a larger cross section (conditional simulation F, Figure 6e), mean concentration predictions improve slightly, indicating that the amount of uncertainty reduction per additional soil water tension measurement point is not proportional to the number of measurements. The higher the sampling grid density, the smaller is the conditional effect of additional data within the sampling grid. Note that perfect knowledge of soil water tension at the source has little



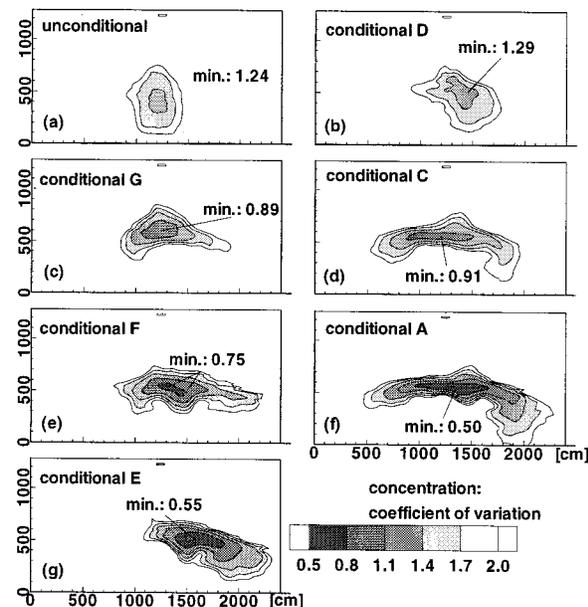
**Figure 5.** Actual field concentration at site 28 for four different dimensionless times,  $t'$ .



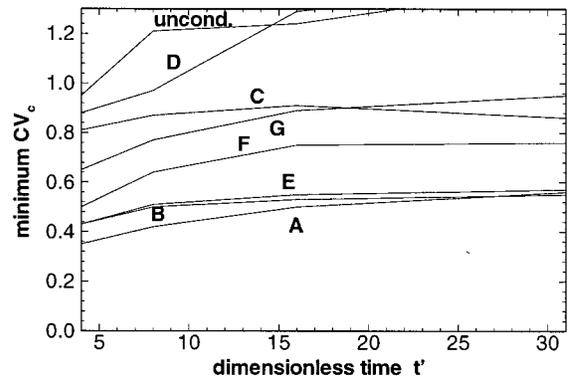
**Figure 6.** (Figure 6a) Unconditional and (Figures 6b–6g) conditional mean concentration, and concentration based on an inverse solution for site 28 at dimensionless time  $t' = 16$ . Letters refer to the type of conditional simulation (see Table 2).

impact on predicting the later behavior of the plume when compared to simulation G.

Figures 7c and 7e show the distribution of the concentration coefficient of variation  $CV_c$  associated with simulations G and F. Contour line,  $CV_c = 2$ , is similar in shape to that of the mean concentration plume, with almost equal extent in the horizontal direction, but vertically it is narrower than the mean



**Figure 7.** (Figure 7a) Unconditional and (Figures 7b–7g) conditional concentration coefficient of variation for field site 28 at dimensionless time  $t' = 16$ . Also indicated are the minimum values of the concentration coefficient of variation and its location at/near the center of the plume. Letters refer to the type of conditional simulation (see Table 2).



**Figure 8.** Dynamical behavior of the conditional minimum concentration coefficient of variation for various conditional simulations of site 28 (see Table 2).

plume. The size of the mean plume is arbitrarily defined by the  $\langle c(t) \rangle / c_{\max}(t) = 0.01$  contour line, where  $c_{\max}(t)$  is the maximum concentration of the actual plume at time  $t$ . Differences in the shape of the mean plume and the  $CV_c$  plume can be explained by considering the directional effects of conditioning. Head correlation scale is larger in the horizontal direction than in the vertical direction. Also, relative horizontal sampling density (sampling points per correlation scale of  $f$ ) is higher than relative vertical sampling density. Conditioning therefore reduces flow and concentration uncertainty more in the horizontal direction than in the vertical direction. Note that the area of low uncertainty outlined by contour line  $CV_c = 2$  increases with increasing mean plume size and time.

Minimum  $CV_c$  locations coincide approximately with locations of peak mean concentration. At  $t' = 4$ , the minimum  $CV_c$  of simulations F and G are 0.50 and 0.65, respectively, compared to 0.95 in the unconditional simulation. At  $t' \leq 20$ , minimum  $CV_c$  generally increases with time, but the ratios between minimum  $CV_c$ 's of the two simulations are almost constant (Figure 8), underscoring conclusions drawn from observations of mean concentration.

By calculating the variance of the mass balance versus time function, the numerical mass balance error is found to contribute approximately 0.1 to the  $CV_c$ . Initially, the numerical mass balance error,  $CV_{\text{bal}}(t) = \text{std}_{\text{bal}}(t) / \text{mass}_{\text{tot}}$  (standard deviation of the total mass balance in the domain divided by the total initial mass) is 0, rapidly increases at early time and reaches a peak of 0.1. Mass balance errors are inherent to the MMOC and must be attributed to heterogeneous velocity fields, for which fourth-order Runge-Kutta travel path integration is known to be inaccurate [see Yeh *et al.*, 1993]. Similar mass balance errors are found in all MCSs of site 28.

**5.3. Sampling Saturated Hydraulic Conductivity Only (Simulation C)**

The variance of  $a$  at field site 28 is moderate, and the soil is relatively wet. Observed spatial variability in unsaturated hydraulic conductivities is therefore much like that of saturated hydraulic conductivities. Particularly, the random spatial patterns of  $y$  should be similar to those of  $f$ . Consequently, if  $f$  is sampled only on the sparse network (not shown), results are almost identical to those shown in Figure 6b for conditional simulation D, with  $f$  and  $a$  data on the sparse network. This implies that at site 28,  $f$  data by themselves should be helpful in discriminating the most probable fast flow paths from likely

slow flow areas. Compared with simulation C (Figure 6d, dense sampling network for  $f$ ), conditional solute plume D is considerably more disperse, particularly at later times. Although the minimum  $CV_c$  of 0.88 at  $t' = 4$  shows an increase of less than 10% over simulation C,  $CV_c = 1.29$  at  $t' = 16$  gives a 30% increase over simulation C (Figure 8). These values approach those for the unconditional simulation, even exceed them at  $t' = 16$ . Away from the plume center, however,  $CV_c$  in conditional simulation D is always less than in the unconditional simulation.

Most interestingly, the minimum  $CV_c$  in simulation C (direct conditioning) is significantly higher than in both simulations F and G (indirect conditioning) (Figure 8). In terms of uncertainty it appears that soil water tension data by themselves improve the concentration prediction more than saturated hydraulic conductivity data by themselves. Only at  $t' = 31$  is the minimum  $CV_c$  in simulation G higher than in simulation C, while the minimum  $CV_c$  in simulation F (dense  $h$  data, indirect conditioning) remains below that in simulation C (Figure 8). This is in partial contrast to the mean concentration prediction, which, particularly at  $t' = 16$  and  $t' = 31$ , visually appears to be significantly better in simulation C than in either F or G. Lower minimum  $CV_c$  in simulations F and G is probably caused by smaller horizontal spreading of mean concentration plumes, which results in higher peak concentration. It is therefore difficult to generalize results from these findings. The results merely point out the possible effects of soil water tension measurements. These examples suggest that  $CV_c$  (or variances) as the only measure of uncertainty may yield ambiguous results.

#### 5.4. Other Sampling Network Combinations for $f$ , $a$ , and $h$

Figure 6g shows results of conditional simulation E (sparse network of  $f$  and  $a$ , dense network of  $h$  measurements). These results are again almost identical to those with sparse  $f$  data and a dense  $h$  measurement network, but without  $a$  data (conditional simulation H, not shown). Spatial concentration distribution is significantly better predicted than in either simulation D (sparse  $f$  and  $a$ ) or F (dense  $h$ ). The mean plume is much less dispersed, resulting in higher concentrations at the center of the plume. The improvement can be seen at  $t' = 16$ , when both the actual and the mean plume exhibit the strongest horizontal spreading. The  $CV_c$  contour map corresponding to this simulation is illustrated in Figure 7g. It has a shape similar to that of the predicted mean concentration plume. The minimum  $CV_c$  for simulation E at  $t' = 4$  is 0.43 compared to 0.88 and 0.50 in simulations D and F, respectively (Figure 8).

Simulation E results are better than those obtained from conditioning on a dense  $f$  sampling network (simulation C). On the other hand, compared with simulation F ( $h$  only), additional saturated hydraulic conductivity information in simulation E helps outline the extremely high- and extremely low-permeability areas, since the spatial variability of  $\alpha$  is not very strong. Results of simulation E are very similar to simulation B, which is based on a dense network for both  $f$  and  $h$  (not shown). This again points to the fact that not much is gained by increasing the number of  $f$  measurements from 53 to over 300, if head data are already available for conditioning. Saturated hydraulic conductivity measurements are more difficult and more expensive to implement in situ than head measurements. Thus a combination of in situ  $h$  and  $f$  data, with more  $h$  measurements than  $f$  measurements, may be the most economical approach to design a monitoring or sampling network.

#### 5.5. "Dense" Sampling Network for All Parameters ( $f$ , $a$ , $h$ ): Simulation A

From a practical point of view, the high sampling density for simulation A (dense network for  $f$ ,  $a$ , and  $h$ ) cannot be achieved without partially removing or destroying the actual site (e.g., the trench site experiment by *Wierenga et al.* [1989]). But this type of conditional simulation serves as a benchmark test to illustrate uncertainty reduction in an optimally sampled field site.

Owing to high data density, the conditional mean concentration distribution is very similar, albeit not identical, to the actual concentration distribution (Figure 6f). The conditional solute plume shows many of the particular distribution patterns of the actual field plume but does not reproduce every local detail. A perfect match is not expected because conditional mean concentrations are obtained not from concentration measurements but from field measurements of variables nonlinearly related to concentration.

The shape of the  $CV_c$  plume again follows the predicted plume (Figure 7f). As expected, the  $CV_c$  in simulation A is by far the lowest in all simulations. As in other conditional simulations, the minimum coefficient of variation in the center of the  $CV_c$  plume increases over the simulated time period from 0.35 at  $t' = 4$  to 0.56 at  $t' = 31$ , indicating increased uncertainty near the center of the plume at later time (Figure 8).

### 6. Unsaturated Hydraulic Conductivity Variance and the Effect of Conditioning Data

#### 6.1. Comparison of Wet and Dry, Anisotropic Field Sites Having Equivalent Variability in $y$

Field site 15 has a much smaller textural variability than field site 28: The variance of  $f$  and  $a$  are only 1 and 0.01, respectively, instead of 2.25 and 0.04 at site 28. However, the increased dryness (mean head  $H = -10.0$  m) leads to a strong increase in the unsaturated hydraulic conductivity and head unconditional variance, such that these are approximately comparable (1.5 versus 1.8, and 0.44 versus 0.49 m<sup>2</sup>, respectively, for sites 15 and 28). Note that in both soils  $f$  and  $a$  are known to be uncorrelated.

Figure 9e shows the actual plume at field site 15 at time  $t' = 10$ . Also shown are the mean concentration results from conditional simulations A, C, and G and from the unconditional simulation. Agreement between conditional mean plumes A and G and actual plume is comparable to findings at field site 28. At similar variances of unsaturated hydraulic conductivity and soil water tension and for identical mean  $\alpha$  and correlation structure, effects of conditioning on a dense set of  $f$ ,  $a$ , and  $h$  data (simulation A) or on  $h$  data alone (simulation G) are similar, regardless of mean soil water tension and variability in saturated hydraulic conductivity.

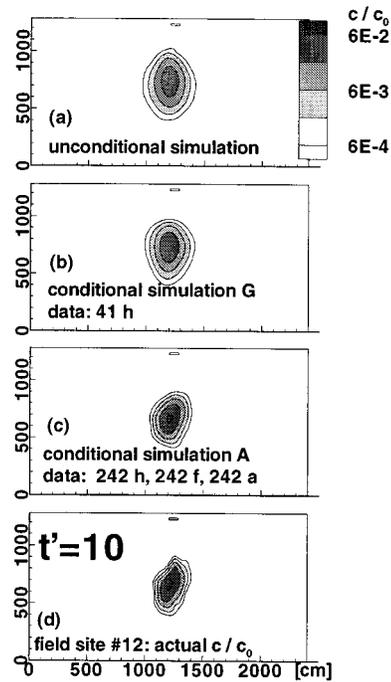
At site 15, however, conditioning on  $f$  alone (conditional simulation C, Figure 9c) neither improves mean concentration predictions nor reduces minimum  $CV_c$  values as much as in wet soil 28 when compared to the unconditional simulation (Figure 9a). Relative to simulation G (sparse  $h$ ), the mean concentration in simulation C (dense  $f$ ) extends more in longitudinal and transverse direction with more uncertainty about the actual travel velocity and travel path. Higher uncertainty is caused by the variability in  $a$ , which here is uncorrelated to  $f$ , and therefore reduces the correlation between saturated and unsaturated hydraulic conductivity as the soil becomes drier. In

contrast, soil water tension data provide information not so much on unsaturated hydraulic conductivity but on the gradient field and hence the approximate travel path of the plume. Particularly in dry soils, soil water tension data should be considered an important source of information for more accurate transport predictions.

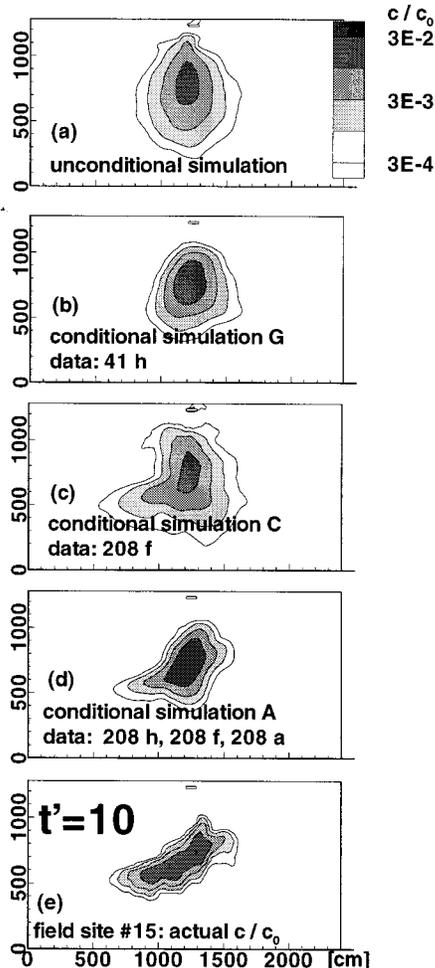
**6.2. Comparison of Mildly Heterogeneous and Strongly Heterogeneous Soils**

Here an example is given of the effect of soil flux variability on the conditional concentration moment prediction. The concentration moments of field sites 12 and 21 are compared with those at field site 28. Field site 12 is a moderately heterogeneous soil with correlated  $f$  and  $a$ , unsaturated hydraulic conductivity variance  $\sigma_y^2 = 0.53$ , and head variance  $\sigma_h^2 = 0.19$  m<sup>2</sup>. Mean head is  $H = -1.50$  m. Field site 21 is the same as field site 12, but in a very dry condition ( $H = -30.0$  m), resulting in  $\sigma_y^2 = 3.2$  and head variance  $0.76$  m<sup>2</sup>. In terms of  $\sigma_y^2$ , field site 12 ranks lowest and field site 21 highest among the sites tested. Note that  $\rho_{af} = 1$  at these two sites, which means that  $f$  data perfectly predict  $a$  at the same location.

The unconditional plume for field site 12 spreads less than those at other field sites (Figure 10a). Consequently, the unconditional simulation itself is a fairly good description of the



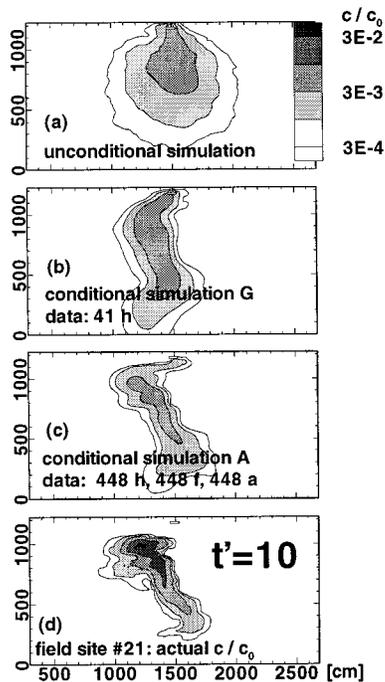
**Figure 10.** Unconditional mean, conditional mean, and actual concentration distribution for site 12 at dimensionless time  $t' = 10$ .



**Figure 9.** Unconditional mean, conditional mean, and actual concentration distribution for site 15 at dimensionless time  $t' = 10$ .

actual plume (at least compared to the conditional simulation results at field site 28), although ergodicity (zero concentration variance) is not achieved within the simulated time period, even for this moderately heterogeneous soil. The most obvious difference between unconditional mean plume and actual plume is the rate of displacement. Conditional simulation A (Figure 10c, high data density) captures the actual rate of displacement of the field plume as well as its particular shape. In contrast, conditional simulation G (Figure 10b, sparse head data only) offers little improvement over the unconditional simulation. At  $t' = 20$ , the minimum  $CV_c$  in the unconditional simulation is 0.79, which decreases by less than 25% to 0.63 in conditional simulation G, but by more than 75% to 0.19 in conditional simulation A. Similar observations can be made at other output times. In soil 12, flow is almost parallel, a situation which has been conceptualized in many stochastic unsaturated flow and transport models as the “parallel column model” [cf. Destouni, 1993]. Therefore uncertainty is reduced to predicting the rate of solute displacement while the travel path is well known. Vertical velocity is primarily controlled by saturated hydraulic conductivity and  $\alpha$ ; hence such data are most important in reducing conditional prediction uncertainty.

Head data, which give information on the overall potential distribution and hence the travel path are even more important in simulating site 21, with its widely fluctuating flow paths, than in simulating site 28. Indeed, the characteristic features of the actual plume are well captured even by conditional simulation G (compare Figures 11a and 11b). The effect of conditioning with head data (when compared to unconditional predictions) is far more important at this site than at any of the other simulated sites. In contrast, even simulation A of site 21 (Figure 11c) cannot exactly predict the actual plume (Figure 11d). At  $t' = 5$  the unconditional minimum  $CV_c$  is 1.51, which



**Figure 11.** Unconditional mean, conditional mean, and actual concentration distribution for site 21 at dimensionless time  $t' = 10$ .

improves by over 50% in simulation G to 0.73, and by almost 70% to 0.47 in conditional simulation A.

Note that sites 12 and 21 represent the same type of soil under different mean soil water tension conditions. Depending on H, the same amount of on-site field data yields distinctly different improvements in conditional plume prediction relative to unconditional stochastic plume predictions. Conditioning on tensiometer measurements is particularly useful in soils with highly heterogeneous flow paths, that is, in soils with a high degree of textural heterogeneity, in very dry soils, or in soils with a steep average slope  $\alpha$  of the  $\log K(h)$  function. In soils with almost exclusively parallel vertical flow, and therefore only mildly heterogeneous unsaturated hydraulic conductivity fields, the same tensiometer measurements have almost negligible effects. Saturated hydraulic conductivity data and data defining  $\alpha$  are important for reducing uncertainty in soils with approximately vertical parallel flow but are less useful (measured in terms of minimum  $CV_c$  reduction relative to the unconditional minimum  $CV_c$ ) in soils with very tortuous flow paths, that is, very heterogeneous or dry soils.

## 7. Anisotropy Ratio and the Effect of Conditioning Data

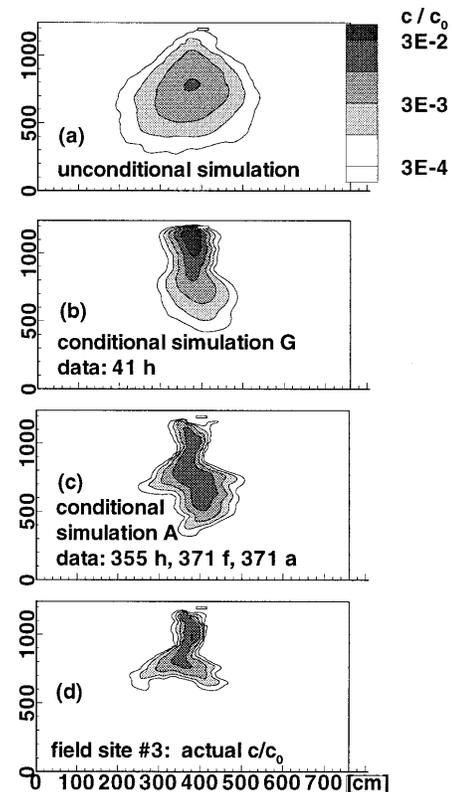
Statistically isotropic soil site 3 is chosen for comparison with conditional simulation results of anisotropic soils 12, 15, and 28. Relative to its horizontal correlation scale, the horizontal plume spreading in site 3 is much larger than in the anisotropic soil 28, even though  $\sigma_y^2$  at site 3 is only half of that at site 28 (Table 1). This is due to increased flow path tortuosity as the transverse-oriented anisotropy of the soil decreases. With larger flow path tortuosity, in situ head measurements themselves significantly reduce prediction uncertainty (Figure 12b), as discussed above. Again, simulation A gives the

best prediction relative to other conditional simulations (Figure 12c). Nevertheless, even in simulation A significant differences to the actual plume (Figure 12d) exist. Minimum  $CV_c$  reduction at  $t' = 5$  is more than 50% from 1.07 to 0.51 in simulation G, and more than 85% from 1.07 to 0.14 in simulation A.

## 8. Conditional Simulation Under Parameter Uncertainty

In all of the previous simulations it is assumed that stochastic parameters describing first and second moments of input parameters  $f$  and  $a$  are known with certainty. In reality, sample populations are small, and estimated mean and covariance values are associated with a degree of uncertainty that can be estimated by, for example, the theoretical sampling error [Harter, 1994]. Parameter uncertainty in a conditional stochastic framework has been addressed by Smith and Schwartz [1981b], who implemented a specific type of conditional Monte Carlo analysis of saturated flow and transport to assess the additional uncertainty introduced by sample estimation errors.

Here an alternative method is implemented to qualitatively describe the effects of parameter uncertainty in conditional simulations of solute transport in unsaturated porous media. We compare mean concentrations predicted by using true (ensemble) parameters with those computed with erroneous statistical parameters that are at the 95% confidence level of the sample moment distribution of  $H$ ,  $F$ ,  $A$ ,  $\sigma_f^2$ , and  $\sigma_a^2$  [see Harter, 1994]. In other words, there is only a 5% chance that parameters obtained from field data differ even more from the



**Figure 12.** Unconditional mean, conditional mean, and actual concentration distribution for site 3 at dimensionless time  $t' = 10$ .

ensemble parameters than do the erroneous ones chosen for these examples. As in work by *Smith and Schwartz* [1981b], the exercise here will be restricted to parameter uncertainty about the mean and the variance of a RSF. Functional forms of the covariance and pdfs of  $f$  and  $a$  are assumed to be known.

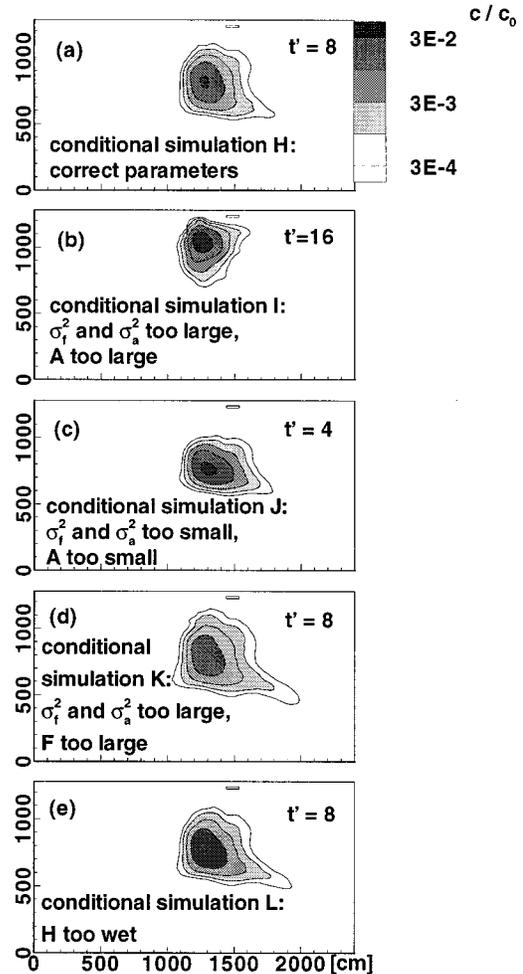
Data for the conditional simulation H of soil site 28 (Figure 13a, sparse  $f$ , dense  $h$ ) are chosen to demonstrate the effect of unwittingly using erroneous statistical parameters. To compute the sample estimation error for  $f$  and  $a$ , one can reasonably assume independence due to the sampling distance (one correlation scale horizontally and two correlation scales vertically). Using standard Gaussian statistics [Harter, 1994], the theoretical sampling errors are determined to be  $\varepsilon_F = 0.24$ ,  $\varepsilon_A = 0.032$ ,  $\varepsilon_{s,f} = 0.17$ ,  $\varepsilon_{s,a} = 0.022$ . Hence with a 95% probability, sample mean estimates of  $f$  and  $a$  are within the intervals  $[-0.5, 0.5]$  and  $[-0.1, 0.1]$ , respectively (two standard deviations about the mean). With the same probability, sample standard deviations of  $f$  and  $a$  should be within the intervals  $[1.2, 1.8]$  and  $[0.16, 0.24]$ , respectively. Soil water tension data are strongly correlated and are available in a dense grid. It is conservatively assumed that sample moments computed from the 352 measured head data have a sampling error equivalent to that of 50 independent head data. The unconditional head standard deviation, computed from unconditional MCS, is 0.7 m [Harter, 1994]. Then  $\varepsilon_H \approx 0.1$  m.

In simulation I (Figure 13b), parameters for the variances of  $f$  and  $a$  are overestimated ( $\sigma_f = 1.8$ ;  $\sigma_a = 0.24$ ). The mean of  $a$  is strongly overestimated: Although  $A$  can be determined very accurately under the above assumptions, it is the most difficult one to estimate in the field, since it is generally derived from fitting theoretical equations to measurements of unsaturated hydraulic conductivity or the soil water retention curve. In simulation I, mean  $\log \alpha$  is arbitrarily set to 0.6, simulating a measurement error of half an order of magnitude.

The erroneously large  $\alpha$  values cause mean conductivity, and hence mean vertical flux, to be lower than in simulation H. Plume movement is therefore slower but along the same path as in the perfect parameter case. Higher variances in  $f$  and  $a$  contribute relatively little to the overall spreading of the mean concentration. Their erroneous effects are offset by the conditioning effect of the actual field data (which are exactly the same as in the perfect parameter case). Note that the dimensionless time  $t' = V_z t / \lambda_{fz}$ , indicated in the individual panels of Figure 13, are all based on the same actual ensemble mean velocity  $V_z$  obtained from an unconditional MCS with correct parameters. In Figure 13, plumes are compared at different times, but at approximately equal displacements from the source.

If variances are underestimated ( $\sigma_f^2 = 1.2$ ,  $\sigma_a^2 = 0.16$ ) and the mean of  $\alpha$  is found lower than in the ensemble such that mean of  $\log \alpha = -0.6$  (simulation J, Figure 13c), the plume moves much faster than the actual plume. Due to smaller  $\alpha$  values, the unconditional mean vertical velocity is 6.8 times faster than in the previous case I, and almost twice as large as that at the actual site. Yet the plume moves along a similar travel path and with only a small decrease in plume spreading.

Figure 13d shows the mean plume prediction from a simulation that again overestimates the variances of  $f$  and  $a$  but has the correct  $A$  (mean of  $a$ ) and an overestimate of  $F$ , the mean of  $f$  (conditional simulation K). The plume moves only slightly faster than in the correct parameter case (Figure 13a) and with only slightly more spreading. Similar results are found if the  $f$  and  $a$  parameters are estimated correctly, but the mean soil



**Figure 13.** Results for conditional mean concentration under parameter uncertainty.

water tension is too wet (conditional simulation L), resulting in a higher average conductivity (Figure 13e).

These results show that conditioning reduces not only the uncertainty attributable to spatial heterogeneity but also the unknown errors arising from a limited knowledge of overall soil properties. Conditioning data tend to neutralize the parameter estimation error. With a high amount of tension data and some conductivity data, mean and variance estimation becomes a relatively minor source of uncertainty compared to the uncertainty arising from the spatial variability of the parameters. Uncertainty in the correlation function needs to be explored in a future study.

## 9. Deterministic Geostatistical Inverse Approach: Comparison

By assuming that all pertinent statistical moments are known and by setting the unconditional random input fields equal to their mean, the conditional simulation algorithm of section 2.1 (Figure 1) reduces to the geostatistical inverse modeling approach described similarly for saturated groundwater flow by *Neuman and Yakowitz* [1979] and by *Kitanidis and Vomvoris* [1983]. Measurements of  $f$ ,  $a$ , and  $h$  are used to estimate the remaining unknown  $f$  and  $a$  data in the simulation grid through the linear, unbiased, cokriging estimator introduced earlier.

Steady state head and solute transport solutions are then computed for the cokriged  $f$  and  $a$  fields. The geostatistical inverse modeling technique is only one of several other indirect inverse modeling techniques [Schweppe, 1973; Neuman and Yakowitz, 1979; Carrera and Neuman, 1986; Peck et al., 1988].

The inverse approach (cokriging) is applied to field site 28 with the same data as those used for conditional simulation A of that site. Since the measured data density is relatively exhaustive,  $f$  and  $a$  parameter estimation is associated with only small errors. Like any random realization of conditional simulation A, the concentration distribution predicted from the geostatistical inverse model is a very good approximation of the overall plume movement (compare Figures 6h and 6a). The solute plume predicted by the inverse model is less dispersed than the conditional mean solute plume since it is not an average concentration. It is also less dispersed than the actual plume, since the underlying parameter fields for  $f$ ,  $a$ , and  $h$  are subject to minimal perturbation given the conditional data. A less tortuous travel path and a mass balance error in the transport simulation of up to +13% lead to higher predicted peak concentrations in the inverse model than observed at the field site.

## 10. Conclusions

Conditioning effects of  $f$  data by themselves decrease not only with increasing heterogeneity but also as soils dry out, particularly if mean and variability of  $a$  are large and if  $a$  is not strongly correlated with  $f$ . There the information content of soil water tension becomes important for two reasons. First, the spatial distribution of head values carries information about the head gradient field in the soil and therefore about the travel path of a solute plume. Second, soil water tension data help to better estimate the unsaturated hydraulic conductivity, which controls both travel velocity and travel path. Hence in soils with highly variable flow fields, conditioning with head data significantly reduces transport prediction uncertainty. Our simulations suggest that a sampling or monitoring network with a dense interval for soil water tension measurements and a sparse interval for saturated hydraulic conductivity significantly reduces prediction uncertainty of concentration. Even at high sampling densities, substantial uncertainty remains about actual concentration levels. It was also found that uncertainty of solute transport predictions arising from soil heterogeneity is much more significant than uncertainty arising from parameter uncertainty, if simulations are conditioned.

From a practical point of view the results are both encouraging and disappointing. They are encouraging in that they show that with less computational effort than in the classic unconditional approach, and with data that are relatively simple to obtain in situ (soil water tension), uncertainty about predicted plume movement in space can be reduced, particularly for applications to highly heterogeneous soils. It is encouraging also in that conditional mean concentration predictions at the very least pinpoint areas where plume displacement significantly differs from the typical downward movement. This information could be used, for example, to identify locations from which additional data may be taken. If the unsaturated flow field is very heterogeneous, conditioning on a few indirect or direct data will greatly improve unconditional stochastic predictions from MCS or macrodispersion analysis. However, results are discouraging in that the simula-

tions have shown how difficult and expensive (in terms of field-sampling cost) it is to accurately predict solute plume movement even under idealized conditions.

In this study, several simplifications are made not only to be able to compare numerical with analytical solutions [Harter and Yeh, this issue], but also to be able to establish some fundamental relationships between monitoring/sampling network and the heterogeneity of the soil. Future work must address effects of variable moisture content and transient flow conditions. Measurement errors, parameter estimation errors, particularly about the correlation structure, and error in assuming wrong models for describing  $K(h)$  and  $h(\theta)$  may further increase prediction uncertainty. Thus it is expected that actual prediction improvements due to conditioning are smaller than shown in these hypothetical examples. These issues await further research.

**Acknowledgments.** We thank Alan Gutjahr for his support in developing the conditional simulation technique. We also acknowledge the anonymous reviewers for their helpful editorial comments. This study was funded by the United States Geological Survey contract 14-08-0001-G2090.

## References

- Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum likelihood method incorporating prior information, *Water Resour. Res.*, 22, 199–210, 1986.
- Clifton, P. M., and S. P. Neuman, Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona, *Water Resour. Res.*, 18, 1215–1234, 1982.
- Dagan, G., Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 2, The solute transport, *Water Resour. Res.*, 18, 835–848, 1982.
- Dagan, G., Solute transport in heterogeneous porous formations, *J. Fluid Mech.*, 145, 151–177, 1984.
- Delhomme, J. P., Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach, *Water Resour. Res.*, 15, 269–280, 1979.
- Destouni, G., Stochastic modelling of solute flux in the unsaturated zone at the field scale, *J. Hydrol.*, 143, 45–61, 1993.
- Graham, W., and D. McLaughlin, Stochastic analysis of nonstationary subsurface solute transport, 1, Unconditional moments, *Water Resour. Res.*, 25, 215–232, 1989a.
- Graham, W., and D. McLaughlin, Stochastic analysis of nonstationary subsurface solute transport, 2, Conditional moments, *Water Resour. Res.*, 25, 2331–2355, 1989b.
- Gutjahr, A. L., Fast Fourier transforms for random field generation, Project report for Los Alamos grant to New Mexico Tech, contract 4-R58-2690R, Dep. of Math., New Mexico Inst. of Mining and Technol., Socorro, 1989.
- Gutjahr, A. L., Q. Bai, and S. Hatch, Conditional simulation applied to contaminant flow modeling, Technical completion report, Dep. of Math., New Mexico Inst. of Mining and Technol., Socorro, 1992.
- Gutjahr, A. L., B. Bullard, S. Hatch, and L. Hughson, Joint conditional simulations and the spectral approach for flow modeling, *J. Stochastic Hydrol. Hydraul.*, 8, 79–108, 1994.
- Harter, T., Unconditional and conditional simulation of flow and transport in heterogeneous, variably saturated porous media, Ph.D. dissertation, Univ. of Ariz., Tucson, 1994.
- Harter, T., and T.-C. J. Yeh, An efficient method for simulating steady unsaturated flow in random porous media: Using an analytical perturbation solution as initial guess to a numerical model, *Water Resour. Res.*, 29, 4139–4149, 1993.
- Harter, T., and T.-C. J. Yeh, Stochastic analysis of solute transport in heterogeneous, variably saturated soils, *Water Resour. Res.*, this issue.
- Isaaks, E. H., and R. M. Srivastava, *An Introduction to Applied Geostatistics*, 559 pp., Oxford Univ. Press, New York, 1989.
- Journal, A. G., Geostatistics for conditional simulations of ore bodies, *Econ. Geol.*, 69, 673–687, 1974.

- Journel, A. G., and C. J. Huijbregts, *Mining Geostatistic*, 600 pp., Academic, San Diego, Calif., 1978.
- Kitanidis, P. K., and E. G. Vomvoris, A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resour. Res.*, 19, 677–690, 1983.
- Matheron, G., The intrinsic random functions and their applications, *Adv. Appl. Probab.*, 5, 438–468, 1973.
- Myers, D. E., Matrix formulation of co-kriging, *J. Math. Geol.*, 14, 249–257, 1982.
- Neuman, S. P., and S. Yakowitz, A statistical approach to the inverse problem of aquifer hydrology, 1, Theory, *Water Resour. Res.*, 15, 845–860, 1979.
- Peck, A., S. Gorelick, G. de Marsily, S. Foster, and V. Kovalevsky, Consequences of spatial variability in aquifer properties and data limitations for groundwater modelling practice, *IAHS Publ.* 175, 272 pp., 1988.
- Rubin, Y., Prediction of tracer plume migration in disordered porous media by the method of conditional probabilities, *Water Resour. Res.*, 27, 1291–1308, 1991.
- Scheweppe, F. C., *Uncertain Dynamic Systems*, 563 pp., Prentice-Hall, Englewood Cliffs, N. J., 1973.
- Smith, L., and F. W. Schwartz, Mass transport, 2, Analysis of uncertainty in prediction, *Water Resour. Res.*, 17, 351–369, 1981a.
- Smith, L., and F. W. Schwartz, Mass transport, 3, Role of hydraulic conductivity data in prediction, *Water Resour. Res.*, 17, 1463–1479, 1981b.
- Wierenga, P. J., A. F. Toorman, D. B. Hudson, J. Vinson, M. Nash, and R. G. Hills, Soil physical properties at the Las Cruces trench site, *NUREG/CR-5441*, Nucl. Regul. Comm., Washington, D. C., 1989.
- Yeh, T.-C. J., R. Srivastava, A. Guzman, and T. Harter, A numerical model for two-dimensional water flow and chemical transport, *Ground Water*, 32, 2–11, 1993.
- Zhang, D., and S. P. Neuman, Eulerian-Lagrangian analysis of transport conditioned on hydraulic data, 1, Analytical-numerical approach, *Water Resour. Res.*, 31, 39–51, 1995a.
- Zhang, D., and S. P. Neuman, Eulerian-Lagrangian analysis of transport conditioned on hydraulic data, 2, Effects of log transmissivity and hydraulic head measurements, *Water Resour. Res.*, 31, 53–63, 1995b.
- Zhang, D., and S. P. Neuman, Eulerian-Lagrangian analysis of transport conditioned on hydraulic data, 3, Spatial moments, travel time distribution, mass flow rate, and cumulative release across a compliance surface, *Water Resour. Res.*, 31, 65–75, 1995c.
- Zhang, D., and S. P. Neuman, Eulerian-Lagrangian analysis of transport conditioned on hydraulic data, 4, Uncertain initial plume state and non-Gaussian velocities, *Water Resour. Res.*, 31, 77–88, 1995d.

---

T. Harter, Department of Land, Air, and Water Resources, University of California, 9240 South Riverbend Ave., Parlier, CA 93648. (e-mail: thharter@ucdavis.edu)

T.-C. J. Yeh, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721.

(Received August 17, 1995; revised January 15, 1996; accepted February 12, 1996.)