

# Numerical Evaluation of Uncertainty in Water Retention Parameters and Effect on Predictive Uncertainty

Feng Pan, Ming Ye,\* Jianting Zhu, Yu-Shu Wu, Bill X. Hu, and Zhongbo Yu

Uncertainty assessment of flow and contaminant transport in the vadose zone entails probability density functions (PDFs) of soil hydraulic parameters. An unconventional maximum likelihood (ML) approach was used in this study to estimate the PDFs of water retention parameters (e.g., van Genuchten  $\alpha$  and  $m$ ) for a situation common in field-scale modeling where core samples are sparse and prior PDFs of the parameters are unknown. In this situation, the unconventional ML approach approximates the PDFs as multivariate Gaussian. This study developed a method of estimating the mean and covariance of the multivariate Gaussian PDF based on the results of least square methods that can be easily obtained in practice. The developed method was applied to and evaluated through numerical simulation of unsaturated flow and tracer transport at the proposed Yucca Mountain geologic repository. Another focus of this study was to investigate the effect of uncertainty in the water retention parameters on predictive uncertainty. By comparing the predictive uncertainty before and after incorporating random water retention parameters, it was found that the random water retention parameters had limited effects on the mean predictions of the state variables including percolation flux and tracer travel time from the potential repository to the water table. Incorporating the uncertainty in the water retention parameters, however, significantly increased the magnitude and spatial extent of predictive uncertainty of the state variables. In particular, incorporating the random water retention parameters significantly changed the 5th and 95th percentiles of the tracer travel time by tens of thousands of years.

ABBREVIATIONS: LS, least square; ML, maximum likelihood; PDF, probability density function; UZ, unsaturated zone; YM, Yucca Mountain.

**N**UMERICAL SIMULATIONS of flow and contaminant transport in unsaturated media require relationships describing the water retention characteristics. The van Genuchten (1980) equation is one of the most widely used relationships:

$$S_e(h) = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \left(1 + |\alpha h|^n\right)^{-m} \quad [1]$$

where  $S_e$  is the effective saturation,  $h$  is the pressure head,  $\theta$  is volumetric water content,  $\theta_s$  and  $\theta_r$  are saturated and residual volumetric water contents, respectively, and  $\alpha$  and  $m$  ( $n = 1 - 1/m$ ) are water retention parameters related to the water entry pressure and soil pore size distribution, respectively. The water retention parameters are usually estimated from water retention data obtained from core samples, and accurately estimating

these parameter values has been an active research field for many years (Yates et al., 1992). Due to their spatial variability, the water retention parameters are treated as random variables in stochastic subsurface hydrology. Probability density functions of the parameters are required for evaluating their uncertainty and its propagation through unsaturated flow and solute transport models (Christiaens and Feyen, 2001; Avaniadou and Paleologos, 2002; Zhou et al., 2003; Lu and Zhang, 2004; Chen et al., 2005; Boateng, 2007). The parameter estimates and the PDFs can be obtained in two ways: direct methods of fitting the water retention data (e.g., Meyer et al., 1997; Schaap and Leij, 1998; Hollenbeck and Jensen, 1998; Christiaens and Feyen, 2000, 2001; Vrugt and Bouten, 2002; Børgesen and Schaap, 2005; Ye et al., 2007a; Chirico et al., 2007) and indirect methods of calibrating the Richards equation (Yeh and Zhang, 1996; Hughson and Yeh, 2000; Wang et al., 2003; Abbaspour et al., 2004; Minasny and Field, 2005). We developed a direct method of estimating the PDFs for measuring the uncertainty of the water retention parameters and for evaluating the effect of the uncertain parameters on the predictive uncertainty of unsaturated flow and contaminant transport.

Many methods have been developed for estimating the water retention parameters and their associated estimation uncertainty. Among them, the least square (LS) method is the most widely used due to its simplicity and flexibility. The LS method has been implemented in the RETC (retention curve) software (van Genuchten et al., 1991; Yates et al., 1992), and the accuracy of the LS estimates is measured by a covariance matrix. The ML method incorporates measurement errors in a rigorous manner and can evaluate the adequacy of model fit (Hollenbeck and

F. Pan and J. Zhu, Division of Hydrologic Sciences, Desert Research Institute, Nevada System of Higher Education, Las Vegas, NV 89119; M. Ye, Dep. of Scientific Computing, Florida State Univ., Tallahassee, FL 32306; M. Ye and B.X. Hu, Dep. of Geological Sciences, Florida State Univ., Tallahassee, FL 32306; Y.-S. Wu, Dep. of Petroleum Engineering, Colorado School of Mines, Golden, CO 80401; and F. Pan and Z. Yu, Dep. of Geosciences, Univ. of Nevada, Las Vegas, NV 89154. Received 17 May 2008.

\*Corresponding author (mye@fsu.edu).

Vadose Zone J. 8:158–166  
doi:10.2136/vzj2008.0092

© Soil Science Society of America

677 S. Segoe Rd. Madison, WI 53711 USA.

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Jensen, 1998). In addition, the ML method gives the Cramer–Rao lower bound for describing the parameter estimation uncertainty. The pedotransfer method (Schaap and Leij, 1998; Christiaens and Feyen, 2000, 2001; Børgesen and Schaap, 2005; Ye et al., 2007a; Chirico et al., 2007) is another type of parameter estimation method; it uses the bootstrap method (Efron and Tibshirani, 1993) to measure the accuracy of the estimates (Schaap and Leij, 1998; Børgesen and Schaap, 2005). These methods do not explicitly yield the parameter PDFs, although normal distributions are always assumed. This renders these methods insufficient for uncertainty assessment of unsaturated flow and contaminant transport. While the Bayesian methods (e.g., Meyer et al., 1997; Vrugt and Bouten, 2002; Minasny and Field, 2005) give the parameter PDFs, they require estimating the prior PDFs from published data sets of the soil water retention parameters. Although estimating the prior PDFs is not difficult for soils, it may be difficult, if not impossible, for other types of unsaturated media such as the fractured rock in this study.

This study estimated the PDFs of the water retention parameters in a Bayesian framework based on an unconventional ML method introduced by Berger (1985, p. 223) in the statistical literature. In particular, the PDFs are estimated for a situation common in field-scale modeling where core samples are sparse and prior PDFs of the parameters are unknown. When core samples are sparse, conventional statistical methods (e.g., Carsel and Parrish, 1988; Russo and Bouton, 1992; Mallants et al., 1996; Russo et al., 2008) of estimating the PDFs based on a large database become inappropriate. When prior PDFs are unknown, regular Bayesian methods cannot be applied. The unconventional ML approach used in this study resolves the problems of sparse core sample measurements and unknown prior PDFs; it shows in a Bayesian framework that the PDFs can be approximated as multivariate Gaussian for unknown prior PDFs regardless of the number of measurements (Berger, 1985, p. 223). This is the major advantage of this approach over conventional ML methods, which give only ML parameter estimates and estimation uncertainty bounds, not the PDFs. Another feature of this approach is that it explicitly considers correlation between the water retention parameters through the multivariate Gaussian PDF, instead of ignoring the correlation (e.g., Zhou et al., 2003) or assuming a perfect correlation (e.g., Avaniadou and Paleologos, 2002). The ML approach gives only a mathematical expression of the multivariate Gaussian PDF but not the way of estimating its mean and covariance. This study showed that the mean of the multivariate normal distribution is the same as the LS parameter estimates and that the covariance can be estimated using the sensitivity matrix of the LS methods. This provides a practical way of using the unconventional ML approach, since the LS parameter estimates and the sensitivity matrix can be easily obtained.

Although the unconventional ML approach was introduced decades ago, it has not received attention from vadose zone hydrologists for estimating the PDFs of the water retention parameters. This study appears to provide its first application and evaluation in the vadose zone to the best of our knowledge. We selected as a case study site the unsaturated zone (UZ) of Yucca Mountain (YM), the proposed geologic repository for spent nuclear fuel or high-level radioactive waste. The site provided a good setting for illustrating and testing the ML approach. In each hydrogeologic layer of the UZ, there are only several available

measurements of the water retention parameters, insufficient for estimating the PDFs using conventional statistical methods. On the other hand, regular Bayesian methods cannot be applied because the prior parameter PDFs are unknown for the fractured porous medium. Due to these obstacles, uncertainty in the water retention parameters has not been fully assessed, despite its importance to the unsaturated flow and radionuclide transport uncertainty, as shown in previous studies (e.g., Zhang et al., 2006; Paleologos et al., 2006).

The necessity of assessing the uncertainty in the water retention parameters at the site is illustrated in Fig. 1. The solid line plots the van Genuchten model fitted using the LS method from water retention data (symbols) of three core samples in the hydrogeologic layer TMN (details of the parameter fitting are available in Bechtel SAIC Company, 2003). Uncertainty of the parameter estimates is quantified by the 95% confidence interval of the parameters, and the corresponding van Genuchten models are plotted in the dashed lines of Fig. 1. When the PDFs of the parameters are unknown, however, using the 95% confidence interval for quantifying the uncertainty is empirical. Knowing the parameter PDFs would better quantify the parametric uncertainty. It is also expected that incorporating the parametric uncertainty into numerical modeling will better simulate the variability of the simulated state variables (e.g., saturation and concentration). The extent of improvement is yet to be examined at the site, however, which partly motivated this study.

Another focus of this study was to investigate the effect of uncertainty in the retention parameters on the predictive uncertainty of unsaturated flow and tracer transport. We were particularly interested in the effect relative to that of permeability and porosity, since understanding the relative effect is important for directing future data collection efforts for uncertainty

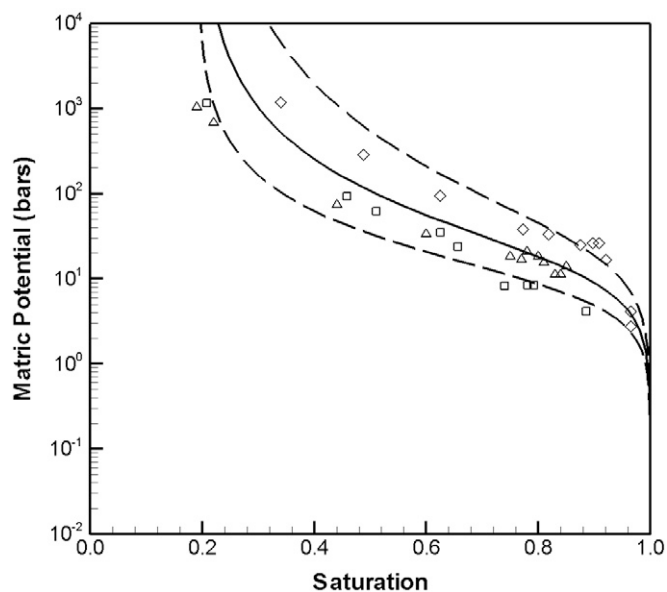


FIG. 1. The van Genuchten model fitted to the water retention data of three samples for the hydrogeologic layer TMN of the unsaturated zone model of Yucca Mountain. Symbols denote the water retention data of three samples, and the solid and dashed lines are the fitted van Genuchten model and their 95% confidence intervals, respectively. The water retention data were adopted from Bechtel SAIC Company (2003). (1 bar = 0.1 MPa.)

reduction. The relative effect has not been examined in previous uncertainty analyses (e.g., Nichols and Freshley, 1993; Avaniidou and Paleologos, 2002; Haukwa et al., 2003; Zhou et al., 2003; Illman and Hughson, 2005; Oliveira et al., 2006; Paleologos et al., 2006; Ye et al., 2007b). We investigated the relative effect by incorporating the uncertainty in the water retention parameters into the numerical modeling of Ye et al. (2007b). Since Ye et al. (2007b) already assessed the predictive uncertainty due to the uncertainty in the permeability and porosity, the relative effect was revealed by comparing the predictive uncertainty of this study with that of Ye et al. (2007b).

## Materials and Methods

### Study Site and Numerical Model

The study site (the UZ of YM) and the numerical model are briefly described here; more details of the site and the model can be found in Bechtel SAIC Company (2004) and Wu et al. (2004). The UZ is a complex geologic formation with heterogeneous layered and anisotropic fractured tuff. It consists of five major geologic units: the Tiva Canyon welded unit (TCw), the Paintbrush nonwelded unit (PTn), the Topopah Spring welded unit (TSw), the Calico Hills nonwelded unit (CHn), and the Crater Flat undifferentiated unit (CFu). Each unit is further divided into multiple hydrogeologic layers, resulting in a total of 33 layers. Table 1 lists the number of core samples used to estimate the water retention parameters for the 33 hydrogeologic layers. For 25 of these layers, the water retention parameters were

estimated from fewer than five core samples. Given that each core sample gave one set of water retention parameters, the small number of parameter estimates is insufficient for identifying the parameter PDFs using traditional statistical methods such as the Kolmogorov–Smirnov test.

The unsaturated flow module, EOS9, and radionuclide transport module, T2R3D, of the TOUGH2 family (Wu et al., 1996; Pruess et al., 1999) were used for simulating the unsaturated flow and radionuclide transport. A three-dimensional numerical grid of the UZ encompassing approximately 40 km<sup>2</sup> was developed, which consisted of 980 mesh columns and 45 numerical layers (Bechtel SAIC Company, 2004). Figure 2 is the plane view of the numerical grid, with the potential repository area highlighted using blue dots. The ground surface and water table were taken as the top and bottom model boundaries, where Dirichlet-type boundary conditions with specified pressure or saturation were applied. A no-flux boundary condition was specified for the lateral boundaries. The present-day net infiltration from precipitation at the land surface was used as a source term in the fracture grid blocks. Infiltration is the major control of the overall hydrologic and thermal conditions within the UZ (Bechtel SAIC Company, 2004). Since the dual-continuum approach was used in the numerical model for the fractured and matrix media, permeability, porosity, and the water retention parameters were required for both media. Based on the sensitivity analysis results of Zhang et al. (2006), only matrix properties were treated as random variables in this study. In Ye et al. (2007b), matrix permeability, porosity, and the adsorption coefficient were treated as

TABLE 1. The estimation of the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the van Genuchten  $\alpha$  and  $m$  parameters.

Layer	Core sample no.	$\mu_{\log(\alpha)}$	$\sigma_{\log(\alpha)}$	$\mu_m$	$\sigma_m$
CCR & CUC	3	0.004	0.244	0.388	0.081
CUL & CW	10	-0.509	0.199	0.280	0.046
CMW	6	-0.488	0.192	0.259	0.044
CNW	8	1.207	0.269	0.245	0.038
BT4	8	1.164	0.169	0.219	0.019
TPY	2	0.391	0.728	0.247	0.104
BT3	3	1.897	0.375	0.182	0.028
TPP	3	1.015	0.189	0.300	0.039
BT2	11	1.992	0.335	0.126	0.017
TC	4	0.939	0.544	0.218	0.068
TR	5	0.055	0.118	0.290	0.025
TUL	4	-0.210	0.114	0.283	0.025
TMN	3	-0.074	0.776	0.317	0.122
TLL	5	0.032	0.447	0.216	0.058
TM2 & TM1	3	-0.081	0.934	0.442	0.173
PV3	5	-0.206	0.446	0.286	0.092
PV2a	1	-0.337	0.156	0.059	0.007
PV2v	1	0.686	0.043	0.293	0.011
BT1a	3	-1.678	0.183	0.349	0.073
BT1v	3	0.940	0.050	0.240	0.008
CHV	5	1.413	0.092	0.158	0.008
CHZ	4	-0.648	0.094	0.257	0.022
BTa	1	-1.807	0.043	0.499	0.036
BTv	1	0.196	0.253	0.147	0.025
PP4	3	-1.349	0.513	0.474	0.200
PP3	5	-0.055	0.094	0.407	0.033
PP2	3	-0.622	0.168	0.309	0.044
PP1	3	-1.036	0.442	0.272	0.116
BF3	2	0.098	0.940	0.193	0.077
BF2	1	-1.921	0.032	0.617	0.070

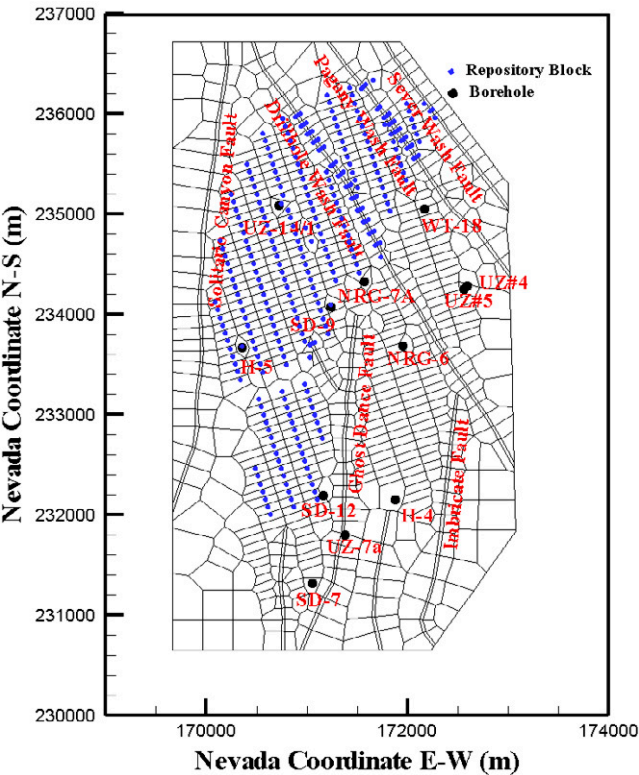


FIG. 2. Plan view of the three-dimensional unsaturated zone numerical model grid showing the model domain, faults, proposed high-level radioactive waste repository layout, and locations of several boreholes (modified from Bechtel SAIC Company, 2004).



random variables; this study further incorporates the matrix water retention parameters as random variables.

### Maximum Likelihood Method of Estimating Probability Density Functions

This study determined the PDFs of the water retention parameters based on the ML theory of Berger (1985, p. 224): "Suppose that  $X_1, X_2, \dots, X_N$  are i.i.d. from the density  $f_0(x_i|\beta)$ ,  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$  being an unknown vector of parameters. (We will write  $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$  and  $f(\mathbf{x}|\beta) = \prod_{i=1}^N f(x_i|\beta)$ , as usual.) Suppose  $\pi(\beta)$  is a prior density, and that  $\pi(\beta)$  and  $f(x_i|\beta)$  are positive and twice differential near  $\hat{\beta}$ , the (assumed to exist) maximum likelihood estimate (MLE) of  $\beta$ . Based on Bayes' theorem, the posterior density of  $\beta$

$$p(\beta | \mathbf{x}) = f(\mathbf{x} | \beta) \pi(\beta) / m(\mathbf{x}) \quad [2]$$

[ $m(\mathbf{x})$  being a normalizing factor], can be approximated by a multivariate normal distribution,  $N_p(\hat{\beta}, [\hat{\mathbf{I}}(\mathbf{x})]^{-1})$ , where  $\hat{\mathbf{I}}$  is the observed (or conditional) Fisher information matrix, having  $(i,j)$  elements

$$\begin{aligned} \hat{I}_{ij}(\mathbf{x}) &= - \left[ \frac{\partial^2}{\partial \beta_i \partial \beta_j} \ln f(\mathbf{x} | \beta) \right]_{\beta=\hat{\beta}} \\ &= - \sum_{i=1}^N \left[ \frac{\partial^2}{\partial \beta_i \partial \beta_j} \ln f(x_i | \beta) \right]_{\beta=\hat{\beta}}. \end{aligned} \quad [3]$$

Taking  $\mathbf{x}$  as the retention data and  $\beta$  as the water retention parameters (or their transforms such as logarithmic), this ML approach provides a method of estimating the PDFs of the water retention parameters. Without having large number of measurements of the water retention parameters and knowing the prior PDF, the posterior PDF is approximated as multivariate Gaussian. This feature renders the ML theory the only way of identifying the PDF of the retention parameters for the UZ and other sites in a similar situation.

The ML approach only gives the expression of the Gaussian PDF,  $N_p(\hat{\beta}, [\hat{\mathbf{I}}(\mathbf{x})]^{-1})$ ; this study shows that its mean (the MLE) is the same as the least square estimate and that its covariance can be estimated from the sensitivity matrix also obtained from the LS method. Assuming that residuals,  $\mathbf{r} = \theta - \hat{\theta}(\beta)$ , between the observed water saturation data ( $\theta$ ) and the estimated data ( $\hat{\theta}$ ) using the van Genuchten model follow a normal distribution with a mean of zero and covariance matrix of  $\sigma^2 \omega^{-1}$  (where  $\sigma^2$  is unknown and the same for all  $x_i$  and  $\omega$  is a weight matrix of the residuals related to measurement error and model quality) (Carrera and Neuman, 1986), the likelihood function is

$$f(\mathbf{x} | \beta, \sigma^2) = \frac{1}{\sqrt{(2\pi)^N |\sigma^2 \omega^{-1}|}} \exp \left( -\frac{\mathbf{r}^T \omega \mathbf{r}}{2\sigma^2} \right) \quad [4]$$

Taking its natural logarithm and multiplying it by  $-1$  on both sides gives

$$\begin{aligned} -\ln f(\mathbf{x} | \beta, \sigma^2) \\ = \frac{N}{2} \ln(2\pi) + \frac{N}{2} \ln(\sigma^2) + \frac{1}{2} \sum_{i=1}^N \ln |\omega^{-1}| + \frac{\mathbf{r}^T \omega \mathbf{r}}{2\sigma^2} \end{aligned} \quad [5]$$

One of the differences between the ML and LS methods is that the ML estimates both  $\beta$  and  $\sigma^2$ , while the LS only estimates  $\hat{\beta}$ . Considering that  $\beta$  and  $\sigma^2$  are independent, the ML estimate  $\hat{\beta}$  of  $\beta$  can be obtained by setting  $-\partial \ln[f(\mathbf{x}|\beta, \sigma^2)]/\partial \beta = 0$  without knowing  $\sigma^2$ . Since  $N_z \ln(2\pi)$ ,  $N_z \ln \sigma^2$ , and  $\ln |\omega^{-1}|$  in Eq. [5] are independent of  $\beta$ , this is equivalent to minimizing the LS objective function:

$$O(\beta) = \mathbf{r}^T \omega \mathbf{r} = [\theta - \hat{\theta}(\beta)]^T \omega [\theta - \hat{\theta}(\beta)] \quad [6]$$

Therefore, the ML estimate  $\hat{\beta}$  is the same as the LS estimates. The equivalence between the MLE and LS estimate is achieved based on the assumption that the residuals,  $\mathbf{r}$ , are Gaussian, a reasonable assumption according to Press et al. (1992) and Carrera and Neuman (1986). A general comparison between the ML and LS methods can be found in Hollenbeck and Jensen (1998), Hill and Tiedeman (2007), and Ye et al. (2008). One can then estimate  $\sigma^2$ , a posteriori, by setting  $-\partial \ln[f(\hat{\beta}, \sigma^2|\mathbf{x})]/\partial \sigma^2 = 0$ , which results in the MLE (Carrera and Neuman, 1986; Seber and Wild, 1989; Seber and Lee, 2003):

$$\hat{\sigma}^2 = \frac{\mathbf{r}^T \omega \mathbf{r}}{N} \Big|_{\beta=\hat{\beta}} \quad [7]$$

To estimate the Fisher information matrix in Eq. [3], taking the second-order derivative of Eq. [5] with respect to the water retention parameters gives

$$\begin{aligned} \hat{I}_{ij}(\mathbf{x}) &= \frac{1}{2\sigma^2} \frac{\partial^2 (\mathbf{r}^T \omega \mathbf{r})}{\partial \beta_i \partial \beta_j} \\ &= \frac{1}{2\sigma^2} \frac{\partial^2 [(\mathbf{x} - \hat{\mathbf{x}})^T \omega (\mathbf{x} - \hat{\mathbf{x}})]}{\partial \beta_i \partial \beta_j} \end{aligned} \quad [8]$$

which can be approximated by (Nelles, 2001)

$$\hat{\mathbf{I}}(\mathbf{x}) = \frac{1}{\hat{\sigma}^2} \mathbf{J}^T \omega \mathbf{J} \quad [9]$$

where  $\mathbf{J}$  is the Jacobian matrix with element  $J_{ij} = \partial \hat{x}_i / \partial \beta_j$  evaluated at  $\hat{\beta}$ . The covariance matrix explicitly measures the correlation between the water retention parameters. The expression of Eq. [9] can be also be found in Carrera and Neuman (1986), Hill and Tiedeman (2007), and Ye et al. (2008). The ML approach was applied to the hydrogeologic layers of the UZ, and the approximated Gaussian PDFs were evaluated in two ways described below.

## Results

In addition to the numerical evaluation of the approximated Gaussian PDF, we also discuss the effect of the uncertainty in the water retention parameters on the predictive uncertainty of the unsaturated flow and tracer transport. Random parameters in this study included not only the water retention parameters but also the matrix permeability, porosity, and sorption coefficient. Uncertainty of the latter three parameters was addressed in Ye et al. (2007b). By comparing the statistics in this study with those of Ye et al. (2007b), the relative (to permeability and porosity) effect of the uncertainty in the water retention parameters to the

predictive uncertainty of unsaturated flow and tracer transport at the UZ of YM was investigated.

### Uncertainty of Matrix van Genuchten $\alpha$ and $m$

Following the tradition of fitting water retention data, the  $\log \alpha$  and  $m$  were fitted from water retention data for each hydrogeologic layer of the UZ, and the fitted mean and variance of the two parameters are listed in Table 1. Values of the mean and variance are significantly different for different layers, reflecting the layering structure of the UZ. Uncertainty in  $\log \alpha$  is particularly large, resulting in an uncertain flow path in the matrix and between the matrix and the fracture. Figure 3 plots the cumulative distribution function (CDF) of the two parameters together with the five parameters fitted from core samples using the RETC software for the TLL layer. The CDF was estimated based on 200 random numbers of the retention parameters generated using the Latin hypercube sampling (LHS) method (McKay et al., 1979). It is well known that the LHS is more efficient for sampling the parameter space than random sampling methods. The parameter correlation was measured using the Spearman rank correlation coefficient, which can measure nonlinear correlation and is thus superior to the commonly used Pearson correlation coefficient (Iman and Conover, 1982; Helton and Davis, 2003). To obtain the rank correlation from the covariance matrix, the statistical software Minitab (Minitab Inc., State College, PA) was used to generate 2000 realizations based on the multivariate Gaussian PDF, and the Spearman rank correlation was estimated based on the 2000 realizations. Figure 3 shows that the fitted parameter values are within the range of their respective CDFs, indicating that the approximated Gaussian distribution was able to describe the uncertainty in the water retention parameters.

### Predictive Uncertainty of Unsaturated Flow

Figure 4 shows the mean and uncertainty bounds of the simulated matrix saturation and corresponding observations at Borehole SD-12 (its location is shown in Fig. 2). The uncertainty bounds are the 5th and 95th percentiles of the simulated state variables (e.g., saturation and percolation fluxes) based on 200 Monte Carlo realizations. Both the variance and uncertainty bounds were used to measure the predictive uncertainty. Since the uncertainty bounds correspond to the 5th and 95th percentiles and directly reveal the variability of the simulated variables, they were considered more informative than

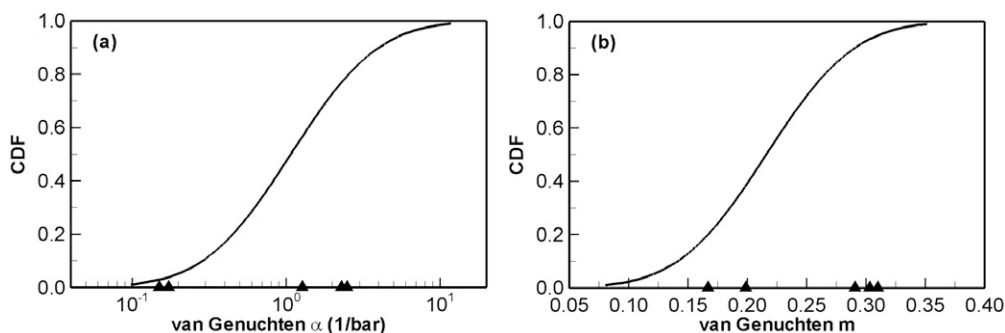


FIG. 3. Cumulative distribution functions (CDFs) of the matrix van Genuchten  $\alpha$  and  $m$  in the TLL layer. Fitted parameter values of five core samples in the layer are also plotted as solid triangles on the x axis. (1 bar = 0.1 MPa.)

the variance. The mean predictions capture the observed variation trend reasonably well, and the uncertainty bounds bracket a large portion of the observations. This suggests that the approximated Gaussian PDFs of the water retention parameters resulted in reasonable simulations of the observed state variables.

Figure 4 also includes the same statistics obtained in Ye et al. (2007b), in which the uncertainty in the water retention parameters was not considered. The mean predictions of both cases (with and without considering uncertainty in the water retention parameters) captured the observed variation trend reasonably well. In Topopah Spring welded unit (TSw) where the potential repository will be located, 75% of the observations are covered by the uncertainty bounds (solid lines) of this study, while the uncertainty bounds (dashed lines) of Ye et al. (2007b) cover only 65% of the observations. This is attributed to the fact that uncertainty in the water retention parameters was not incorporated in Ye et al. (2007b).

Percolation flux and its variations through the UZ can significantly affect the migration of radionuclide released from the

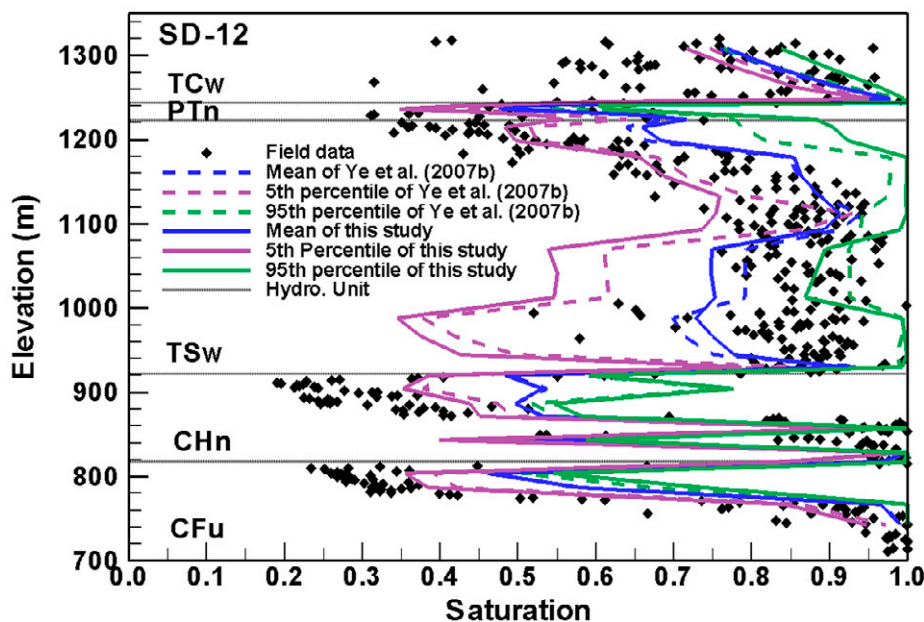


FIG. 4. Comparison of the observed and simulated matrix liquid saturation with (solid line) and without (dashed line) considering the water retention parameter uncertainty for Borehole SD-12.

potential repository. Figures 5a and 5b plot the mean and variance of the simulated percolation fluxes at the water table, and Fig. 5c and 5d are those of Ye et al. (2007b) in which the water retention parameters were treated as deterministic. Comparison of the mean values (Fig. 5a and 5c) shows that the magnitude and spatial pattern are similar across the entire domain, suggesting a limited effect of the uncertainty in the water retention parameters on the mean predictions. Comparing Fig. 5b and 5d, however, reveals that the variance of the percolation flux increased significantly after the uncertainty in the water retention parameters was incorporated. On average across the simulation domain, the variance increased by about 38%; the number of grid blocks at the water table with a variance  $>10 \text{ mm}^2 \text{ yr}^{-2}$  is almost doubled.

### Predictive Uncertainty of Unsaturated Tracer Transport

Transport of a conservative tracer, technetium ( $^{99}\text{Tc}$ ), and a reactive tracer, neptunium ( $^{237}\text{Np}$ ) was simulated for a scenario in which a constant-concentration source was released instantaneously from the fracture continuum grid blocks representing the potential repository (Fig. 2). Predictive uncertainty of the tracer transport was quantified in terms of the plume and breakthrough of the tracers at the water table. Spatial distribution of the normalized cumulative mass arrival at the water table is an important variable in investigating transport patterns and in estimating the potential locations of high radionuclide concentrations. The cumulative mass arrival is the cumulative mass arriving at each cell of the water table with time, normalized by the total mass of the released tracers from the repository. Figures 6a and 6b show the mean and variance of the normalized cumulative mass arrival contours of  $^{237}\text{Np}$  at the water table after 1,000,000 yr. The mean and variance are large in the area directly below the footprint of the proposed repository. The spatial pattern of the variance (Fig. 6b) is similar to that of the flow variance contour shown in

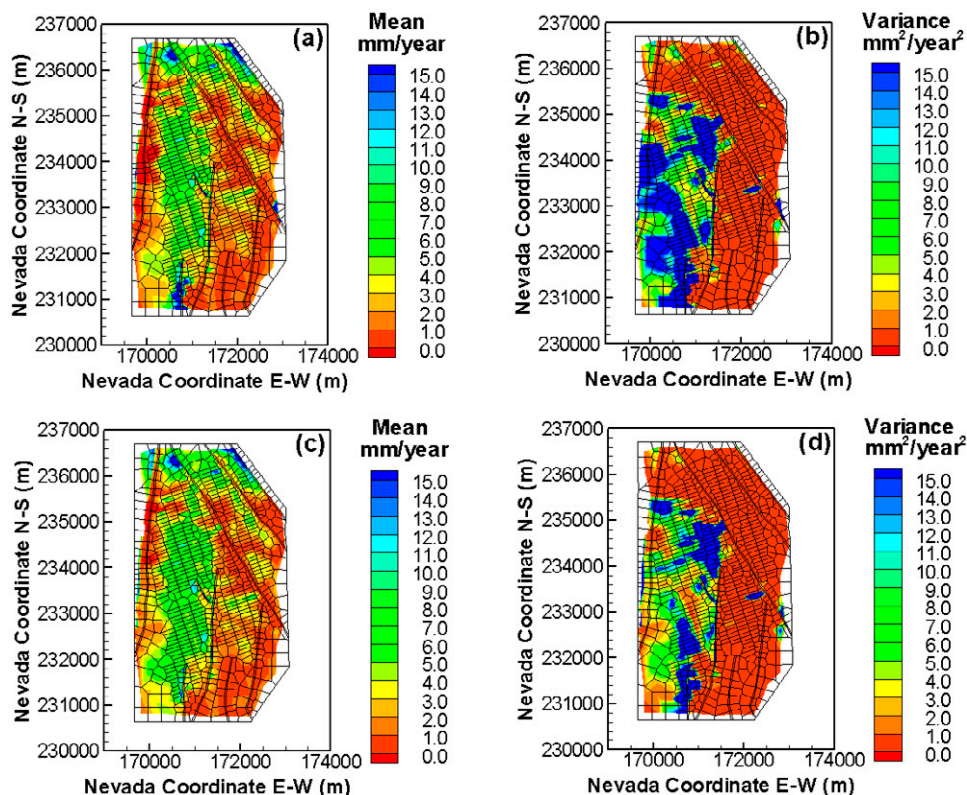


FIG. 5. Mean and variance of the simulated percolation fluxes at the water table with (a and b) and without (c and d) considering the water retention parameter uncertainty.

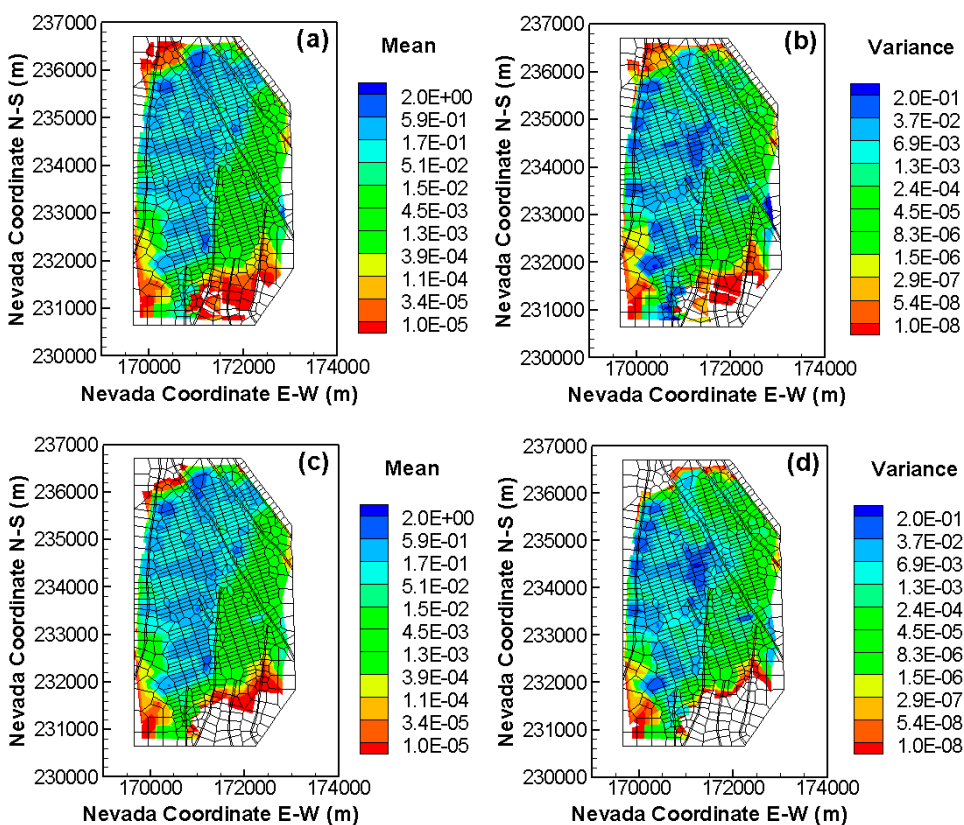


FIG. 6. Mean and variance of the normalized cumulative mass arrival contours of the reactive tracer ( $^{237}\text{Np}$ ) at the water table after 1,000,000 yr with (a and b) and without (c and d) considering the water retention parameter uncertainty.



Fig. 5b, indicative of correlation between the uncertainty of flow and tracer transport. Figures 6c and 6d depict the same mean and variance of the normalized cumulative mass without considering the uncertainty in the water retention parameters (Ye et al., 2007b). Comparing contours of the mean predictions in Fig. 6a and 6c suggests a limited effect of the uncertainty in the water retention parameters on the mean predictions of the tracer transport. The variance shown in Fig. 6b is significantly larger than that of Fig. 6d, however—almost doubled on average across the whole domain. In addition, the area with a variance  $>0.01$  in Fig. 6b also increased by about 3% relative to that shown in Fig. 6d.

Tracer travel time from the potential repository to the water table is another important variable for performance assessment of the proposed repository. Different from calculating the normalized cumulative mass arrival, the tracer travel time is obtained by summing the cumulative mass arriving at the water table for all blocks at a given time. Figure 7 plots the simulated breakthrough curves as the fractional cumulative mass arriving at the water table for the  $^{99}\text{Tc}$  and  $^{237}\text{Np}$ . The uncertainty bounds of the breakthrough curves in Fig. 7 show that fractional mass arrival is significantly uncertain. Figure 7 also includes the same statistics without considering the uncertainty in the water retention parameters (Ye et al., 2007b). Due to the large time scale used in Fig. 7, for better evaluation of the travel time uncertainty, Table 2 lists the 5th and 95th percentiles at the 10, 25, 50, 75, and 90% mass fractional breakthrough for both cases. Similar to what has been observed from the contours, the mean breakthrough was affected only slightly by considering the uncertain water retention parameters, while the uncertainty bounds increased more significantly. For example, with the random water retention parameters, the 5th and 95th percentiles of the simulated travel time of  $^{99}\text{Tc}$  are  $8.05 \times 10^3$  and  $9.43 \times 10^2$  yr until 50% of the radionuclide would have arrived at the water table. With the deterministic water retention parameters, the corresponding travel times are  $7.17 \times 10^3$  and  $8.22 \times 10^2$  yr. The uncertainty range increases from 6348 to 7107 yr if the uncertainty in water retention parameters is considered. Similarly, for 50% of the reactive tracer ( $^{237}\text{Np}$ ) arriving at the water table, the uncertainty range increases from 255,000 to 278,100 yr.

## Conclusions

This study addressed two problems in numerical simulations of unsaturated flow and contaminant transport. The first one is how to estimate the PDFs of the water retention parameters when measurements of the parameters are sparse and the prior PDFs are unknown; the other question is whether the uncertainty in the water retention parameters is important in the predictive uncertainty of unsaturated flow and contaminant transport. The first problem was resolved using the unconventional ML approach (Berger, 1985), which approximates the PDFs as multivariate Gaussian without requiring the

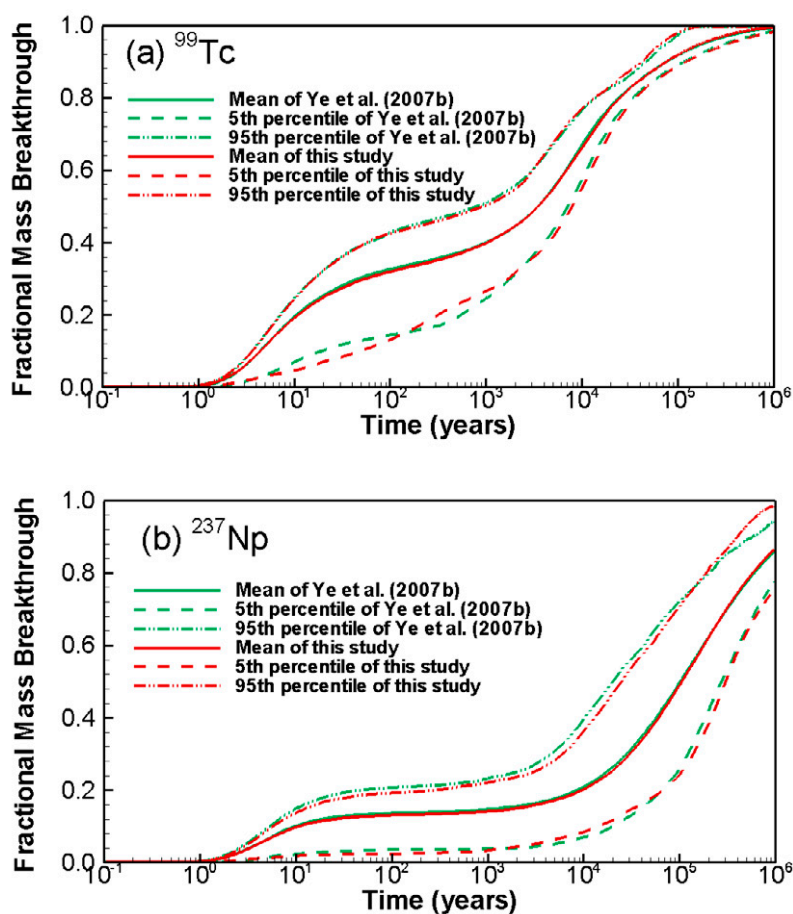


FIG. 7. Simulated breakthrough curves of the cumulative mass arriving at the water table for (a) the conservative tracer ( $^{99}\text{Tc}$ ) and (b) the reactive tracer ( $^{237}\text{Np}$ ) with (this study) and without (Ye et al., 2007b) considering the water retention parameter uncertainty.

TABLE 2. Comparison of the 5th and 95th percentiles of the simulated travel time of the conservative ( $^{99}\text{Tc}$ ) and reactive ( $^{237}\text{Np}$ ) tracers arriving at the water table at 10, 25, 50, 75, and 90% mass fraction breakthrough with (this study) and without (Ye et al., 2007b) considering the water retention parameter uncertainty.

Breakthrough curve	Mass fraction	Travel time of this study		Travel time of Ye et al. (2007b)	
		$^{99}\text{Tc}$	$^{237}\text{Np}$	$^{99}\text{Tc}$	$^{237}\text{Np}$
	%				
5th percentile	10	$4.97 \times 10^1$	$1.43 \times 10^4$	$1.87 \times 10^1$	$1.99 \times 10^4$
	25	$7.53 \times 10^2$	$1.05 \times 10^5$	$1.08 \times 10^3$	$9.40 \times 10^4$
	50	$8.05 \times 10^3$	$3.03 \times 10^5$	$7.17 \times 10^3$	$2.75 \times 10^5$
	75	$2.55 \times 10^4$	$9.42 \times 10^5$	$2.32 \times 10^4$	$8.38 \times 10^5$
	90	$1.23 \times 10^5$	$>1.00 \times 10^6$	$1.17 \times 10^5$	$>1.00 \times 10^6$
95th percentile	10	3.72	5.91	3.86	5.34
	25	$1.08 \times 10^1$	$3.00 \times 10^3$	$1.03 \times 10^1$	$1.98 \times 10^3$
	50	$9.43 \times 10^2$	$2.49 \times 10^4$	$8.22 \times 10^2$	$2.00 \times 10^4$
	75	$8.66 \times 10^3$	$1.33 \times 10^5$	$9.00 \times 10^3$	$1.29 \times 10^5$
	90	$4.19 \times 10^4$	$4.01 \times 10^5$	$4.70 \times 10^4$	$5.80 \times 10^5$

prior PDFs and a large number of parameter measurements. This study developed the method of estimating the mean and covariance of PDFs based on the LS fitting results, which can be easily estimated from existing software such as RETC. For the case study of the YM UZ, water retention parameter ranges obtained from the Gaussian distributions encompassed the parameter values of individual samples, and were significantly larger than the ranges

of the measured parameter values. This indicates that uncertainty in the water retention parameters should not be ignored.

The relative effect of the uncertainty in the water retention parameters on the predictive uncertainty of flow and transport was evaluated using the Monte Carlo method. After the random water retention parameters were taken into account, the variability of the observed matrix saturations was better represented in that 10% more observations were bracketed by the uncertainty bounds. The predictive variance of the percolation flux increased when the random water retention parameters were taken into account, while the uncertain water retention parameters had a limited effect on the mean prediction of percolation fluxes. A similar conclusion was also drawn for the magnitude and spatial pattern of the simulated plume of both conservative and reactive tracers. The travel time of the two types of tracers also became more uncertain after incorporating the uncertain water retention parameters, signified by the result that the uncertainty bounds of the tracer travel time increased by tens of thousands of years.

#### ACKNOWLEDGMENTS

This work was conducted, in part, under the cooperative agreement between the U.S. Department of Energy (DOE) and the Nevada System of Higher Education. F. Pan was partly supported by the Maki Fellowship of the Desert Research Institute (DRI); M. Ye conducted part of this research at the DRI.

#### References

- Abbaspour, K.C., C.A. Johnson, and M.Th. van Genuchten. 2004. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone J.* 3:1340–1352.
- Avanidou, T., and E.K. Paleologos. 2002. Infiltration in stratified, heterogeneous soils: Relative importance of parameters and model variations. *Water Resour. Res.* 38(11):1232, doi:10.1029/2001WR000725.
- Berger, J.O. 1985. *Statistical decision theory and Bayesian analysis*. 2nd ed. Springer-Verlag, New York.
- Boateng, D. 2007. Probabilistic unsaturated flow along the textural interface in three capillary barrier models. *J. Environ. Eng.* 133:1024–1031.
- Børgesen, C.D., and M.G. Schaap. 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. *Geoderma* 127:154–167.
- Bechtel SAIC Company. 2003. Analysis of hydrologic properties data. Rep. MDL-NBS-HS-000014 REV00. Lawrence Berkeley Natl. Lab., Berkeley, CA.
- Bechtel SAIC Company. 2004. UZ flow models and submodels. Rep. MDL-NBS-HS-000006 REV02. Lawrence Berkeley Natl. Lab., Berkeley, CA.
- Carrera, J., and S.P. Neuman. 1986. Estimation of aquifer parameters under transient and steady-state conditions: 1. Maximum-likelihood method incorporating prior information. *Water Resour. Res.* 22:199–210.
- Carsel, R.F., and R.S. Parrish. 1988. Developing joint probability distributions of soil water retention characteristics. *Water Resour. Res.* 24:755–769.
- Chen, M., D. Zhang, A.A. Keller, and Z. Lu. 2005. A stochastic analysis of steady state two-phase flow in heterogeneous media. *Water Resour. Res.* 41:W01006, doi:10.1029/2004WR003412.
- Chirico, G.B., H. Medina, and N. Romano. 2007. Uncertainty in predicting soil hydraulic properties at the hillslope scale with indirect methods. *J. Hydrol.* 344:405–422.
- Christiaens, K., and J. Feyen. 2000. The influence of different methods to derive soil hydraulic properties on the uncertainty of various model outputs of a distributed hydrological model. *Phys. Chem. Earth B* 25:679–683.
- Christiaens, K., and J. Feyen. 2001. Analysis of uncertainties associated with different methods to determine soil hydraulic properties and their propagation in the distributed hydrological MIKE SHE model. *J. Hydrol.* 246:63–81.
- Efron, B., and R.J. Tibshirani. 1993. *An introduction to the bootstrap*. Monogr. Stat. Appl. Probab. 57. Chapman & Hall, New York.
- Haukwa, C.B., Y.W. Tsang, Y.-S. Wu, and G.S. Bodvarsson. 2003. Effect of heterogeneity in fracture permeability on the potential for liquid seepage into a heated emplacement drift of the potential repository. *J. Contam. Hydrol.* 62–63:509–527.
- Helton, J.C., and F.J. Davis. 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Safety* 81:23–69.
- Hill, M.C., and C.R. Tiedeman. 2007. Effective methods and guidelines for groundwater model calibration, including analysis of data, sensitivities, predictions, and uncertainty. John Wiley & Sons, New York.
- Hollenbeck, K.J., and K.H. Jensen. 1998. Maximum-likelihood estimation of unsaturated hydraulic parameters. *J. Hydrol.* 210:192–205.
- Hughson, D.L., and T.-C. Yeh. 2000. An inverse model for three-dimensional flow in variably saturated porous media. *Water Resour. Res.* 36:829–840.
- Illman, W.A., and D.L. Hughson. 2005. Stochastic simulations of steady state unsaturated flow in a three-layer, heterogeneous, dual continuum model of fractured rock. *J. Hydrol.* 307:17–37.
- Iman, R.L., and W.J. Conover. 1982. A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. Simul. Comput.* 11:311–334.
- Lu, Z., and D. Zhang. 2004. Analytical solutions to steady state unsaturated flow in layered, randomly heterogeneous soils via Kirchhoff transformation. *Adv. Water Resour.* 27:775–784.
- Mallants, D., D. Jacques, M. Vanclooster, J. Diels, and J. Feyen. 1996. A stochastic approach to simulate water flow in a macroporous soil. *Geoderma* 10:299–324.
- McKay, M.D., R.J. Beckman, and W.J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245.
- Meyer, P.D., M.L. Rockhold, and G.W. Gee. 1997. Uncertainty analysis of infiltration and subsurface flow and transport for SDMP sites. NUREG/CR-6565, PNNL-11705. U.S. Nucl. Regul. Commiss., Office of Nucl. Regul. Res., Washington, DC.
- Minasny, B., and D.J. Field. 2005. Estimating soil hydraulic properties and their uncertainty: The use of stochastic simulation in the inverse modeling of the evaporation method. *Geoderma* 126:277–290.
- Nelles, O. 2001. *Nonlinear system identification: From classical approaches to neural networks and fuzzy models*. Springer-Verlag, Berlin.
- Nichols, W.E., and M.D. Freshley. 1993. Uncertainty analyses of unsaturated zone travel time at Yucca Mountain. *Ground Water* 31:293–301.
- Oliveira, L.I., A.H. Demond, L.M. Abriola, and P. Goovaerts. 2006. Simulation of solute transport in a heterogeneous vadose zone describing the hydraulic properties using a multistep stochastic approach. *Water Resour. Res.* 42:W05420, doi:10.1029/2005WR004580.
- Paleologos, E.K., T. Avanidou, and N. Mylopoulos. 2006. Stochastic analysis and prioritization of the influence of parameter uncertainty on the predicted pressure profile in heterogeneous, unsaturated soils. *J. Hazard. Mater.* 136:137–143.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical recipe in Fortran 77*. 2nd ed. Cambridge Univ. Press, New York.
- Pruess, K., C. Oldenburg, and G. Moridis. 1999. TOUGH2 user's guide, version 2.0. LBL-43134. Lawrence Berkeley Natl. Lab., Berkeley, CA.
- Russo, D., and M. Bouton. 1992. Statistical analysis of spatial variability in unsaturated flow parameters. *Water Resour. Res.* 28:1911–1925.
- Russo, D., J. Zaidel, and A. Laufer. 2008. Numerical analysis of solute transport from trickle sources in a combined desert soil–imported soil flow system. *Vadose Zone J.* 7:53–66 10.2136/vzj2007.0050.
- Schaap, M.G., and F.J. Leij. 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil Tillage Res.* 47:37–42.
- Seber, G.A.F., and A.J. Lee. 2003. *Linear regression analysis*. 2nd ed. John Wiley & Sons, New York.
- Seber, G.A.F., and C.J. Wild. 1989. *Nonlinear regression*. John Wiley & Sons, New York.
- van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44:892–898.
- van Genuchten, M.Th., F.J. Leij, and S.R. Yates. 1991. The RETC code for quantifying the hydraulic functions of unsaturated soils. EPA/600/2-91-065. USEPA, Office of Res. and Dev., Washington, DC.
- Vrugt, J.A., and W. Bouton. 2002. Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. *Soil Sci. Soc. Am. J.* 66:1740–1751.
- Wang, W., S.P. Neuman, T. Yao, and P.J. Wierenga. 2003. Simulation of large-



- scale field infiltration experiments using a hierarchy of models based on public, general, and site data. *Vadose Zone J.* 2:297–312.
- Wu, Y.-S., C.F. Ahlers, P. Fraser, A. Simmons, and K. Pruess. 1996. Software qualification of selected TOUGH2 modules. LBNL-39490. Lawrence Berkeley Natl. Lab., Berkeley, CA.
- Wu, Y.S., G. Lu, K. Zhang, and G.S. Bodvarsson. 2004. A mountain-scale model for characterizing unsaturated flow and transport in fractured tuffs of Yucca Mountain. *Vadose Zone J.* 3:796–805.
- Yates, S.R., M.Th. van Genuchten, A.W. Warwick, and F.J. Leij. 1992. Analysis of measured, predicted, and estimated hydraulic conductivity using the RETC computer program. *Soil Sci. Soc. Am. J.* 56:347–354.
- Ye, M., R. Khaleel, M.G. Schaap, and J. Zhu. 2007a. Simulation of field injection experiments in heterogeneous unsaturated media using cokriging and artificial neural network. *Water Resour. Res.* 43:W07413, doi:10.1029/2006WR005030.
- Ye, M., P.D. Meyer, and S.P. Neuman. 2008. On model selection criteria in multimodel analysis. *Water Resour. Res.* 44:W03428, doi:10.1029/2008WR006803.
- Ye, M., F. Pan, Y.-S. Wu, B.X. Hu, C. Shirley, and Z. Yu. 2007b. Assessment of radionuclide transport uncertainty in the unsaturated zone of Yucca Mountain. *Adv. Water Resour.* 30:118–134.
- Yeh, T.-C.J., and J. Zhang. 1996. A geostatistical inverse method for variably saturated flow in the vadose zone. *Water Resour. Res.* 32:2757–2766.
- Zhang, K., Y.-S. Wu, and J.E. Houseworth. 2006. Sensitivity analysis of hydrological parameters in modeling flow and transport in the unsaturated zone of Yucca Mountain, Nevada, USA. *Hydrogeol. J.* 14:1599–1619.
- Zhou, Q., H.H. Liu, G.S. Bodvarsson, and C.M. Oldenburg. 2003. Flow and transport in unsaturated fractured rock: Effects of multiscale heterogeneity of hydrogeologic properties. *J. Contam. Hydrol.* 60:1–30.