Uncertainty in dual permeability model parameters for structured soils

B. Arora,[1] B. P. Mohanty,[1] and J. T. McGuire[2]

Received 30 January 2011; revised 17 November 2011; accepted 5 December 2011; published 19 January 2012.

[1] Successful application of dual permeability models (DPM) to predict contaminant transport is contingent upon measured or inversely estimated soil hydraulic and solute transport parameters. The difficulty in unique identification of parameters for the additional macropore- and matrix-macropore interface regions, and knowledge about requisite experimental data for DPM has not been resolved to date. Therefore, this study quantifies uncertainty in dual permeability model parameters of experimental soil columns with different macropore distributions (single macropore, and low- and high-density multiple macropores). Uncertainty evaluation is conducted using adaptive Markov chain Monte Carlo (AMCMC) and conventional Metropolis-Hastings (MH) algorithms while assuming 10 out of 17 parameters to be uncertain or random. Results indicate that AMCMC resolves parameter correlations and exhibits fast convergence for all DPM parameters while MH displays large posterior correlations for various parameters. This study demonstrates that the choice of parameter sampling algorithms is paramount in obtaining unique DPM parameters when information on covariance structure is lacking, or else additional information on parameter correlations must be supplied to resolve the problem of equifinality of DPM parameters. This study also highlights the placement and significance of matrix-macropore interface in flow experiments of soil columns with different macropore densities. Histograms for certain soil hydraulic parameters display tri-modal characteristics implying that macropores are drained first followed by the interface region and then by pores of the matrix domain in drainage experiments. Results indicate that hydraulic properties and behavior of the matrix-macropore interface is not only a function of saturated hydraulic conductivity of the macroporematrix interface (K_{sa}) and macropore tortuosity (l_f) but also of other parameters of the matrix and macropore domains.

Citation: Arora, B., B. P. Mohanty, and J. T. McGuire (2012), Uncertainty in dual permeability model parameters for structured soils, *Water Resour. Res.*, *48*, W01524, doi:10.1029/2011WR010500.

1. Introduction

[2] Reliable predictions of flow and transport in the vadose zone are important to address the issue of potential contamination of groundwater and the deterioration of water quality. Various studies have reported faster transport of fertilizers, pesticides, industrial chemicals, and pathogens to groundwater through fractures and preferential flow paths [*National Research Council*, 1994; *Mohanty et al.*, 1997, 1998; *Kladivko et al.*, 2001; *Böhlke*, 2002; *Jamieson et al.*, 2002]. Preferential flow phenomenon can be described using a variety of single-, dual- or multiple-porosity/permeability models [*Gwo et al.*, 1995; *Šimùnek and van Genuchten*, 2008; *Arora et al.*, 2011]. The classical dual permeability approach assumes that the soil contains two interacting domains, one associated with the fast-flowing fracture or macropore domain and the other with the less-permeable soil matrix domain [*van Genuchten and Wierenga*, 1976; *Gerke and van Genuchten*, 1993a, 1993b]. Dual permeability model formulations differ in their description of flow through the macropore domain and in their characterization of exchange between the two regions [*Jarvis*, 1994; *Šimùnek et al.*, 2003; *Köhne et al.*, 2004]. Both types of dual permeability models (DPM) are widely applied at column, plot, and field scales [*Larsbo et al.*, 2005; *Köhne and Mohanty*, 2005; *Köhne et al.*, 2009]. The main disadvantage of DPM is the requirement of a large number of input parameters. Parameters associated with additional pore regions and matrix-macropore interface cannot be directly estimated by independent measurements or by expert judgment [e.g., *Gwo et al.*, 1995; *Schwartz et al.*, 2000; *Roulier and Jarvis*, 2003]. Since direct estimation is not feasible, an inverse procedure is applied wherein observed data are used to obtain an optimal set of model parameters [*Zachman et al.*, 1981; *Kool and Parker*, 1988]. Inverse parameter estimation is challenging with respect to obtaining a unique parameter set, nonidentifiability of the solution set, and ill-posedness of the inverse problem [*Carrera and Neumann*, 1986]. This problem is significant for the case of structured soils where interdependence and multicolinearity between dual permeability model parameters increase the risk of

[1]Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas, USA.
[2]Department of Geology and Geophysics, University of St. Thomas, St. Paul, Minnesota, USA.

0043-1397/12/2011WR010500

reaching local minima in the parameter set [*Ginn and Cushman*, 1990]. The identification of parameters is also hindered by poor measurement quality, nonoptimal experimental design, and parsimonious data sets such as omitting the near-saturated stage of an outflow experiment [*Durner et al.*, 1999; *Dubus et al.*, 2002].

[3] One response to counter the problem of parameter identification is to adopt a Bayesian viewpoint which evaluates the distribution of parameters instead of a single "best" estimate [*Vrugt et al.*, 2008]. The Bayesian approach quantifies uncertainty bands around parameter mean values and incorporates the associated uncertainty to generate better forecasts, especially for complex nonlinear systems [*Wu et al.*, 2010; *Jana and Mohanty*, 2011]. Consider a radioactive waste disposal facility, for instance, where combining the single "best" estimates for the uncertain inputs will not necessarily produce the "most probable" output estimate. Most importantly, Bayesian probabilistic modeling can prove useful in identifying additional parameters of the dual permeability model, analyzing the relationship among parameters of significant domains, and quantifying uncertainty in flow and transport predictions using the dual permeability framework.

[4] The use of Bayesian techniques in the field of preferential flow and transport is generally limited to conventional Markov chain Monte Carlo algorithms such as Metropolis-Hastings and Gibbs sampling [*Gelman et al.*, 1995; *Cowles and Carlin*, 1996; *Marshall et al.*, 2004; *Reis and Stedinger*, 2005]. The computational efficiency of sampling the parameter space can be improved by employing an adaptive Markov chain Monte Carlo (AMCMC) scheme that can cater to model parameters having a high degree of correlation and interdependence as is the case with the dual permeability framework [*Haario et al.*, 2001; *Atchadé and Rosenthal*, 2005]. The AMCMC scheme is compared to a conventional Metropolis-Hastings (MH) algorithm that uses a random walk in the parameter space while describing uncertainty based on preexisting (or prior) knowledge and experimental observations [*Metropolis et al.*, 1953; *Hastings*, 1970]. The algorithms differ in their updating mechanisms, the conventional MH algorithm uses a single-site (one parameter at a time) updating while the AMCMC approach uses the history of the process to "tune" the proposal distribution and update the parameter covariance structure [*Marshall et al.*, 2004; *Peters et al.*, 2009]. The algorithms will be compared for their predictive performance in quantifying parameter and output uncertainty.

[5] In summary, dual permeability models are paramount in predicting reliable estimates of preferential flow and contaminant transport in structured soil systems but their application is hindered by difficulties in estimating the large number of input parameters [*Šimùnek et al.*, 2001; *Jarvis et al.*, 2007]. The focus of this study is to estimate uncertainty in dual permeability model parameters and to investigate the stability of preferential flow estimates from experimental soil columns, especially when a large number of dual permeability parameters are considered unknown or random. The research is motivated by a realization that correlation and interdependence among parameters of the dual permeability framework cannot be described explicitly for any study for one of the following reasons: it may be unknown, known but extremely complex, or it may even be nonexistent, and difficult to investigate through controlled experiments alone. This leads by default to a replacement of the uncertain parameters and unknown covariance structure with probabilistic assumptions which are compatible with Bayesian statistics. Therefore, the primary objectives of this study are: (1) to quantify uncertainty in dual permeability model parameters obtained from experiments of single and multiple (low- and high-density) macropore soil columns, and (2) to compare the conventional Metropolis-Hastings and adaptive Markov chain Monte Carlo algorithms in terms of convergence rate and for quantifying uncertainty in simulating preferential flow from the experimental soil columns.

2. Theoretical Considerations

2.1. Dual-Permeability Model Formulation

[6] The dual-permeability model of *Gerke and van Genuchten* [1993a, 1993b] is used in this study. Conceptually, the model assumes the porous medium to be divided into two pore regions, with relatively fast water flow in one region (often called the interaggregate, macropore, or fracture domain) when close to full saturation, and slow in the other region (often referred to as the intra-aggregate, micropore, or matrix domain) [*Šimùnek and van Genuchten*, 2008]. Flow in both macropore (subscript f) and matrix (subscript m) domains is described using two Richards' equations primarily with different sets of water retention and hydraulic conductivity functions:

$$\frac{\partial \theta_f}{\partial t} = \frac{\partial}{\partial z}\left(K_f \frac{\partial h_f}{\partial z} + K_f\right) - \frac{\Gamma_w}{w_f}, \tag{1}$$

$$\frac{\partial \theta_m}{\partial t} = \frac{\partial}{\partial z}\left(K_m \frac{\partial h_m}{\partial z} + K_m\right) - \frac{\Gamma_w}{1 - w_f}, \tag{2}$$

where z is the vertical coordinate positive upward [L], t is time (T), h is the pressure head [L], θ is the water content [L^3L^{-3}], K is the unsaturated hydraulic conductivity [LT^{-1}], w_f is the ratio of the volumes of the macropore domain and the total soil system [dimensionless], and Γ_w is the rate of water exchange between the two domains [T^{-1}]. The soil water retention and hydraulic conductivity functions can be described using the equations [*Mualem*, 1976; *van Genuchten*, 1980]:

$$\theta_d(h) = \theta_{rd} + (\theta_{sd} - \theta_{rd})[1 + (\alpha_d h)^{n_d}]^{-m_d}, \tag{3}$$

$$K_d(h) = K_{sd} \frac{\{1 - (\alpha_d h)^{m_d n_d}[1 + (\alpha_d h)^{n_d}]^{-m_d}\}^2}{[1 + \{\alpha_d h\}^{n_d}]^{l_d m_d}}, \tag{4}$$

$$m_d = 1 - \frac{1}{n_d}, \tag{5}$$

where subscript d represents the matrix (m) or fracture (f) domains, θ_r and θ_s are the residual and saturated water contents [L^3L^{-3}], respectively, K_s is the saturated hydraulic conductivity [LT^{-1}], α [L^{-1}], n [−], m [−], and l [−] are empirical parameters determining the shape of the hydraulic conductivity functions.

[7] The mass transfer rate (Γ_w) in equations (1) and (2) is assumed to be proportional to the difference in pressure

heads between the fracture and matrix domains, given by [*Gerke and van Genuchten*, 1993a]:

$$\Gamma_w = \alpha_w(h_f - h_m), \tag{6}$$

in which α_w is a first-order mass transfer coefficient for water [$L^{-1}T^{-1}$]. For porous media with well-defined geometries, α_w can be defined as follows [*Gerke and van Genuchten*, 1993b]:

$$\alpha_w = \frac{\beta}{a^2} K_{sa}\gamma, \tag{7}$$

where β is a dimensionless geometry-dependent shape factor, a is the characteristic length of the aggregate (L) (e.g., the radius of a spherical or solid cylindrical aggregate, or the half-width of a rectangular aggregate), K_{sa} is the saturated hydraulic conductivity of the fracture/matrix interface region [LT^{-1}], and γ is a dimensionless scaling factor.

2.2. Description of Bayesian Methods

[8] Bayesian methods provide a statistical framework for obtaining an improved estimate of parameter distributions by mathematically combining specific prior knowledge with what is known about those parameters through observations. To facilitate the description of the Bayesian analysis, we represent the soil hydrologic system in a Bayes' framework:

$$\frac{p(\mathbf{\Theta}|\,\mathbf{D}) = f(\mathbf{D}|\mathbf{\Theta})\pi(\mathbf{\Theta})}{\pi(\mathbf{D}),} \tag{8}$$

where $\mathbf{D}$ is the observed data, $p(\mathbf{\Theta}|\mathbf{D})$ is the conditional posterior probability of the soil hydraulic parameters given the data, $f(\mathbf{D}|\mathbf{\Theta})$ is the likelihood function summarizing the model for the data given the parameters, $\pi(\mathbf{D})$ is a normalizing constant, $\pi(\mathbf{\Theta})$ is the prior joint probability for the soil hydraulic parameters, and $\mathbf{\Theta}$ is the vector of van Genuchten soil hydraulic parameters given by:

$$\mathbf{\Theta} = \{(\theta_{rd}, \theta_{sd}, \alpha_d, n_d, K_{sd}, l_d); (w_f, \beta, \gamma, a, K_{sa})\},\ d = m \text{ or } f, \tag{9}$$

where subscripts m and f represent the matrix and macropore domain parameters, respectively, and (w_f, β, γ, a, K_{sa}) constitute the interface region (*int*) parameters. The prior joint probability can be further broken down as the joint probability for the matrix, macropore, and interface components of the dual-permeability model:

$$\pi(\mathbf{\Theta}) = \Pi_{npar_m}\boldsymbol{\varphi}_m \times \Pi_{npar_f}\boldsymbol{\varphi}_f \times \Pi_{npar_{\text{int}}}\boldsymbol{\varphi}_{\text{int}}, \tag{10}$$

where *npar* is the number of parameters of a particular region that are considered random and $\boldsymbol{\varphi}$ is the set containing the random soil hydraulic parameters for that particular region. Once the conditional posterior probability is known, the marginal posterior distribution $p(.|\mathbf{D})$ that retains the dependence on one parameter exclusively (e.g., residual soil water content for the matrix domain, θ_{rm}) is given by:

$$p(\theta_{rm}|\mathbf{D}) = \frac{\int\int\int_{\theta_2,\ldots,\theta_{\text{tot}}} f(\mathbf{D}|\mathbf{\Theta}) \times \pi(\mathbf{\Theta})d\theta_2,\ \ldots, d\theta_{\text{tot}}}{\pi(\mathbf{D})}, \tag{11}$$

where θ_2, $\theta_3, \ldots, \theta_{\text{tot}}$ represent the soil hydraulic parameters contained in the set $\mathbf{\Theta}$ apart from θ_1 ($= \theta_{rm}$). The main complication is the intractability of the multidimensional integration of the target density including the computation of the normalizing constant $\pi(\mathbf{D})$. A possible solution is to use any MCMC algorithm that ignores $\pi(\mathbf{D})$ and generates a sequence of parameter sets, $\{\mathbf{\Theta}(0), \mathbf{\Theta}(1), \ldots, \mathbf{\Theta}(t)\}$ that converge to the stationary proposal distribution $p(\mathbf{\Theta}|\mathbf{D})$ for large number of iterations t [*Gelman et al.*, 1995]. Diagnostic measures of central tendency and dispersion can then be calculated from the mean and variance of the large sample generated using MCMC simulations. The MCMC algorithms used in this study are described below.

2.2.1. Metropolis-Hastings Algorithm

[9] One of the widely used MCMC techniques is the Metropolis-Hastings algorithm proposed by *Hastings* [1970]. It samples the posterior distribution of the parameters as described below:

[10] 1. Choose a starting point randomly within the feasible parameter space, $\mathbf{\Theta}(i) = \mathbf{\Theta}(0)$ with a covariance matrix $\mathbf{\Sigma}_0$.

[11] 2. Draw a candidate vector $\mathbf{\Theta}(i+1)$ from the previous vector $\mathbf{\Theta}(i)$ using a proposal distribution $q(\mathbf{\Theta}(i+1)|\mathbf{\Theta}(i)) \sim N(\mathbf{\Theta}(i), \mathbf{\Sigma}_0)$, where $\mathbf{\Theta}(i)$ is the current state of the chain, and the proposal density is a normal distribution (for this study).

[12] 3. Compute the odds ratio: $r = q(\mathbf{\Theta}(i+1))/q(\mathbf{\Theta}(i))$.

[13] 4. If $r \geq 1$, accept the new candidate vector $\mathbf{\Theta}(i+1)$ as it leads to a higher value of the proposal distribution.

[14] 5. If $r < 1$, draw a number at random from a uniform distribution U[0,1]. If the random number is less than r, accept "$\mathbf{\Theta}(i+1)$" else remain at the current position "$\mathbf{\Theta}(i)$."

[15] 6. Repeat steps 2–5 for the given number of iterations (t).

[16] A single parameter updating is usually done in this algorithm which may be problematic with high-dimensional $\mathbf{\Theta}$. If two or more parameters are highly correlated, a larger number of simulations are required and block or simultaneous updating is necessitated for correlated parameters [*Marshall et al.*, 2004].

2.2.2. Adaptive Metropolis Algorithm

[17] We employ the AMCMC scheme of *Harrio et al.* [2001], which corresponds to our need for resolving a large number of dual-permeability parameters and understanding the correlation among these parameters. *Harrio et al.* [2001] chose a multivariate normal distribution as the proposal density which is centered on the current state and obtains empirical covariance from a fixed number of previous states. A fixed value of the covariance matrix $\mathbf{\Sigma}$ is employed for a finite number of initial iterations (t_0) after which it is updated as a function of all the previous iterations:

$$\sum\nolimits_i = \begin{cases} \sum_0, i \leq t_0 \\ s_d \text{Cov}(\mathbf{\Theta}_1, \mathbf{\Theta}_2, \ldots, \mathbf{\Theta}_{\text{iter}-}1) + s_d\varepsilon \boldsymbol{I}_d, i > t_0, \end{cases} \tag{12}$$

where $\mathbf{\Sigma}_0$ is the initial covariance based on prior knowledge, d is the dimension of $\mathbf{\Theta}$, ε is a small parameter chosen to ensure $\mathbf{\Sigma}_i$ does not become singular, $\boldsymbol{I}_d$ is the d-dimensional identity matrix, and s_d is a scaling parameter that depends only on d. A basic choice for the scaling parameter can be $s_d = (2.4)^2/d$ for Gaussian targets and Gaussian proposals [*Gelman et al.*, 1995]. The covariance at iteration $(i + 1)$ can be obtained without much computational cost using the recursive formula:

$$\sum_{i+1} = \frac{i-1}{i}\sum_{i} + \frac{s_d}{i}(i\bar{\mathbf{\Theta}}_{i-1}\,\bar{\mathbf{\Theta}}_i^T - (i+1)\bar{\mathbf{\Theta}}_i\bar{\mathbf{\Theta}}_i^T + \varepsilon \boldsymbol{I}_d). \quad (13)$$

[18] The important steps of the AMCMC algorithm can be described as follows:

[19] 1. Choose a starting point randomly within the feasible parameter space, $\mathbf{\Theta}(i) = \mathbf{\Theta}(0)$ with a covariance matrix $\mathbf{\Sigma}_i = \Sigma_0$.

[20] 2. Draw a candidate vector $\mathbf{\Theta}(i + 1)$ from the previous vector $\mathbf{\Theta}(i)$ using a proposal distribution $q(\mathbf{\Theta}(i+1)|\mathbf{\Theta}(i)) \sim \mathrm{N}(\mathbf{\Theta}(i),\mathbf{\Sigma}_i)$, where $\mathbf{\Theta}(i)$ and $\mathbf{\Sigma}_i$ define the current state of the chain, and the proposal density is a normal distribution (for this study). $\mathbf{\Sigma}_i$ depends on the iteration number i according to equation (12).

[21] 3. Compute the odds ratio: $r = q(\mathbf{\Theta}(i + 1))/q(\mathbf{\Theta}(i))$.

[22] 4. If $r \geq 1$, accept the new candidate vector $\mathbf{\Theta}(i + 1)$ as it leads to a higher value of the proposal distribution.

[23] 5. If $r < 1$, draw a number at random from a uniform distribution U[0,1]. If the random number is $<r$, accept "$\mathbf{\Theta}(i + 1)$" or else remain at the current position "$\mathbf{\Theta}(i)$."

[24] 6. Repeat steps 2–5 for the given number of iterations (t).

[25] The distinguishing feature of adaptive MCMC algorithms, compared to the MH algorithm, is that it updates all elements of $\mathbf{\Theta}$ simultaneously due to the description of the covariance structure. This also helps in adapting the simulation at an early stage and reducing computation time. Both adaptive MCMC and AMCMC terms are used interchangeably throughout the paper.

3. Case Study

3.1. Soil Column Data

[26] This work uses soil column experiments with well-defined boundary conditions [*Arora et al.*, 2011] to fully understand the prospects and limitations of employing adaptive MCMC versus the conventional MH algorithm to quantify uncertainty in 10 out of 17 dual-permeability model parameters. Three large acrylic cylinders (75 cm long and 24 cm wide) are packed with sandy loam soil to a dry bulk density of 1.56 g cm^{-3}. The central macropore column is provided with a single macropore of 1 mm diameter all along the vertical axis of the column, open to both the soil surface and to the bottom outflow plate. In the low- and high-density multiple macropore columns, 3 and 19 artificial holes (1 mm diameter each) are created with a stainless steel rod in one-half of the column cross-section, respectively (Figure 1). Basic outflow curves from the three columns are also displayed in Figure 1. Tensiometers and time domain reflectometry (TDR) probes are installed at various depths in both macropore and nonmacropore halves of the soil columns to monitor pressure head profiles and water/tracer concentrations, respectively (Figure 2). At the bottom of the column, outflow rates and flux-averaged bromide (Br^-) concentrations are measured separately in six effluent chambers: three for each soil region with and without macropores. The top boundary condition is maintained using a tension infiltrometer and the bottom boundary is open to the atmosphere with provision for pressure control. A detailed description of the soil columns and collection of data are provided elsewhere [*Arora et al.*, 2011].

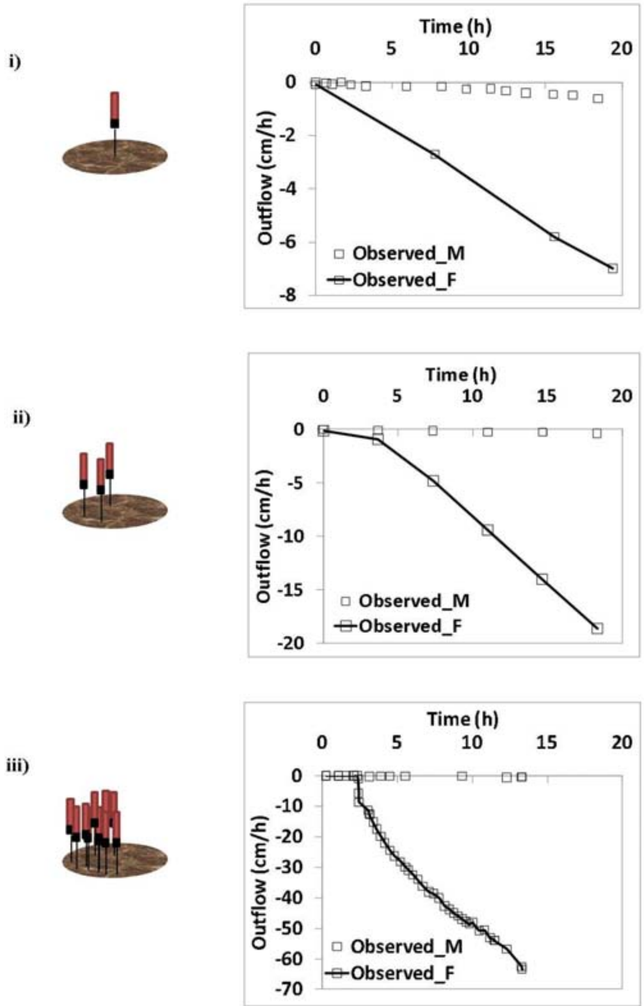

Figure 1. Experimental design and outflow from infiltration experiments of the (i) single macropore, (ii) low-density, and (iii) high-density multiple macropore columns. Symbol *M* represents soil matrix and *F* represents fracture or macropore domain.

3.2. Model Parameters, Initial and Boundary Conditions

[27] We present results for infiltration and drainage experiments performed on the single macropore, and low- and high-density multiple macropore columns. Simulations of the experimental soil columns are implemented using the HYDRUS-1D software package [*Šimùnek et al.*, 2001, 2003]. Initial conditions for the simulations are described in terms of vertical pressure head distribution using tensiometric data at different depths of the soil column (5 cm intervals from the top). Upper and lower boundary conditions are derived from observed tensiometric data at the top (close to 0 cm) and bottom (close to 75 cm) of the soil profile, respectively. A spatial discretization (Δz = 0.5 cm)

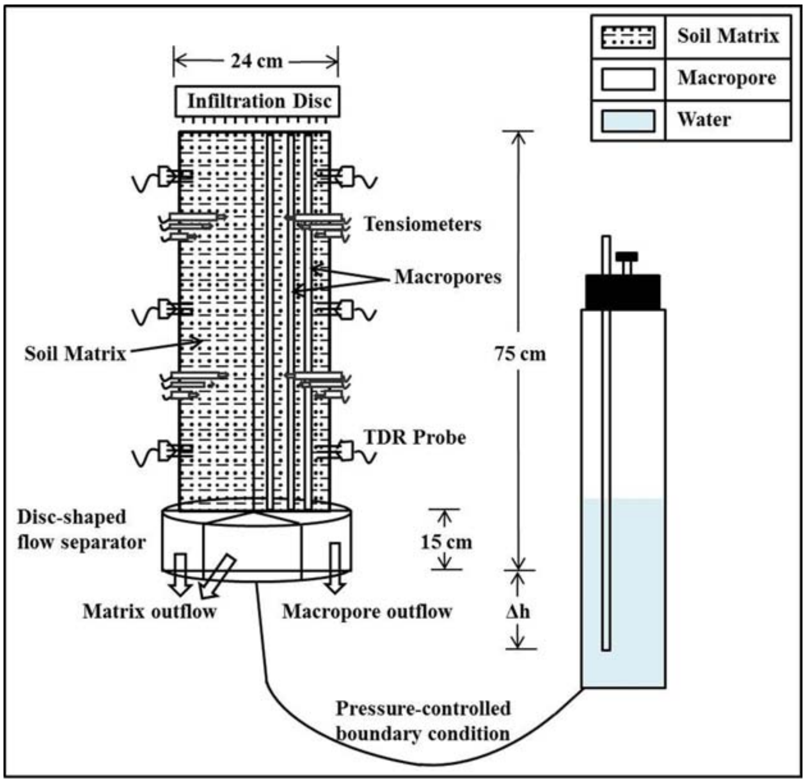

Figure 2. Schematic of the soil column with instrumentation.

uniformly distributed over the length of the column is used for all experiments. The initial time step is $\Delta t = 10^{-5}$ h, and minimum and maximum time steps are $\Delta t_{\min} = 10^{-6}$ h and $\Delta t_{\max} = 10^{-1}$ h, respectively. Space and time discretization are kept identical for all soil columns. The simulation periods for the different experiments vary according to the respective duration of each experiment.

[28] In the dual-permeability framework, any water flow simulation requires the following 17 parameters: van Genuchten-Mualem parameters (θ_r, θ_s, α, n, K_s, and l) for both matrix and macropore domains, and interface parameters (w_f, β, γ, a, and K_{sa}). The parameters of the matrix-macropore interface, except K_{sa}, are either based on their geometry (w_f, β, and a) or obtained by empirical estimation (γ) for the single and multiple macropore columns [*Castiglione et al.*, 2003; *Arora et al.*, 2011]. As these parameters are kept as constants, one may argue that α_w is a function of K_{sa} only (equations (6) and (7)), which is regarded as a calibration parameter in HYDRUS. This suggests that there are only 13 independent parameters based on degrees of freedom. These constant interface parameters along with l_m, θ_{rm}, and θ_{rf} are not included in the uncertainty analysis because they are not considered to be sensitive (see section 4.1). However, correlations with respect to θ_{rm} and θ_{rf} are taken into account.

3.3. Markov Chain Monte Carlo Sampling

[29] The MCMC algorithms are applied to the experimental soil columns to investigate the effect of parameter correlations and uncertain model parameters on model outputs. The first step is to establish prior density and parameter uncertainty limits for each of the random parameters. As discussed in section 3.2, the 10 dual permeability parameters that will be analyzed using MCMC algorithms are $\boldsymbol{\varphi}_m = \{\theta_{sm}, \alpha_m, n_m, \text{and } K_{sm}\}$, $\boldsymbol{\varphi}_f = \{\theta_{sf}, \alpha_f, n_f, K_{sf}, \text{and } l_f\}$, and $\boldsymbol{\varphi}_{\text{int}} = \{K_{sa}\}$. A log-transformation is used for the saturated hydraulic conductivity parameter (K_s) of matrix, macropore, and interface regions as suggested by *de Rooij et al.* [2004]. A uniform distribution is assigned to parameters whose literature references are unavailable except for their ranges. Therefore, the prior for l_f is U[0,1]. A normal distribution is assigned as a prior to the rest of the soil hydraulic parameters for both matrix and macropore domains, e.g., $\theta_{sm} \sim \mathrm{N}(\mu_{\theta sm}, \sigma_{\theta sm})$. Nonnormal priors can be used as well but they will increase the computational complexity considering the number of parameters involved in this problem. The means of the prior densities for the matrix and macropore domains are set at the optimized values obtained using inverse modeling of the various flow experiments as they reflect the least squares estimate from HYDRUS. Table 1 summarizes the inverse modeling technique used in this study and further details are given elsewhere [*Arora et al.*, 2011]. The variances for the normal densities are obtained from *Vrugt et al.* [2003] using the van Genuchten model for the loam and coarse sand textures reflecting the parameters of the matrix and macropore domains, respectively. The uncertainty limits for these parameters are based on ranges obtained from the UNSODA database [*Nemes et al.*, 1999, 2001] again using the loam and sand textures. To avoid MCMC algorithms from progressively sampling outside realistic parameter ranges, the variances and applicable uncertainty limits are further refined by prior experiences with the model. Table 2 enlists the optimized parameter values used as means for the prior density

Table 1. Experimental Observations Used for Parameter Estimation and Likelihood Calculations

Group of Observations	Method	Resolution for Data Collection	Minimum Resolution for Likelihood Calculation
Pressure head (cm)	13 tensiometers	5-cm depth intervals starting from top until bottom of the soil column	Three depths
Soil water content (cm^3 cm^{-3})	12 TDR probes	10-cm depth intervals starting from 5 cm until 55 cm on both matrix and macropore halves of the columns	Two depths in both matrix and macropore halves
Outflow (cm)	Six pie-shaped chambers, intermittent use of fraction collector	75 cm depth	One depth in both matrix and macropore halves

and the uncertainty bounds approximately reflect the values at $\pm 3\sigma$ (standard deviation) for parameters with normal priors.

[30] The second step is to consider an appropriate likelihood function and create a proposal distribution that is close to the posterior target distribution. Sampling from proposal distributions should be consistent with expected model responses to changes in parameter values [*Larsbo et al.*, 2005]. Therefore, the proposal distribution is taken to be a multivariate normal distribution for each region/domain, and a Gaussian jump function is used to move around the parameter space. HYDRUS-1D is run for each "new" vector in the dual-permeability framework and the likelihood is based on the weighted least squares estimate between observed (D) and predicted values (E) [*Šimùnek et al.*, 2001, 2003]:

$$f(\mathbf{D}|\mathbf{\Theta}_r, \sigma_r) \propto \sigma_r^{-N} \exp\left(-\frac{1}{2\sigma_r^2}\sum_{i=1}^{N} w_i[r(\mathbf{\Theta})]^2\right), \tag{14}$$

$$r(\mathbf{\Theta}) = D(x,t) - E(x,t,\mathbf{\Theta}), \tag{15}$$

where N is the number of observations, w_i are weights associated with a particular observation, $r(\mathbf{\Theta})$ are model residuals calculated using the observation data $D(x, t)$ at time t and location x (see Table 1), and the corresponding model predictions $E(x, t, \mathbf{\Theta})$ for the vector $\mathbf{\Theta}$ of dual-permeability model parameters. We assumed w_i's to be equal to 1 for this study to represent similar error variances for all observations. A problem with equation (14) is that the standard deviation of model residuals (σ_r), which is not known a priori, is also included in the likelihood function. Typically, σ_r can be integrated out of the equation using a Jeffreys prior, and the likelihood therefore becomes [*Scharnagl et al.*, 2011]:

$$f(\mathbf{D}|\mathbf{\Theta}_r) \propto \left(\sum_{i=1}^{N} w_i[r(\mathbf{\Theta})]^2\right)^{-N/2}. \tag{16}$$

[31] The Bayesian technique can thus produce full probability distributions for each parameter that is obtained after integrating all possible combinations of the dual permeability parameters using equation (11). This multidimensional integration is performed using the MH and AMCMC algorithms which differ primarily in their dealings with the covariance matrix.

3.4. Implementation of the MCMC Algorithms

3.4.1. Convergence Criteria

[32] A variety of graphical techniques such as trace plots, running mean plots, posterior means, variances, and standard errors are used to assess convergence of MCMC chains. Apart from these, convergence diagnostics of

Table 2. Initial Uncertainty Range and Optimal Parameter Values Obtained From HYDRUS for MCMC Simulations

Dual Permeability Parameters		Initial Uncertainty Range	Parameter Value for Best HYDRUS Simulation: Single Macropore Column	Low-Density Macropore Column	High-Density Macropore Column
Matrix or immobile region	θ_{rm} (−)	Fixed	0.2	0.2	0.2
	θ_{sm} (−)	0.35–0.41	0.38	0.38	0.38
	α_m (cm^{-1})	0–0.14	0.004	0.004	0.004
	n_m (−)	1.38–2.22	1.8	1.8	1.8
	K_{sm} (cm h^{-1})	0.003–5.53	0.13	0.13	0.13
	l_m (−)	Fixed	0.5	0.5	0.5
Macropore or mobile region	θ_{rf} (−)	Fixed	0.08	0.08	0.08
	θ_{sf} (−)	0.36–0.42	0.39	0.39	0.39
	α_f (cm^{-1})	0–0.14	0.01	0.01	0.01
	n_f (−)	1.1–2.9	2	2	2
	K_{sf} (cm h^{-1})	1.85–37	8.27	8.27	8.27
	l_f (−)	0–1	0.5	0.5	0.5
Interface region	w_f (−)	Fixed	1.7×10^{-5}	5.2×10^{-5}	3.3×10^{-4}
	β (−)	Fixed	0.45	0.54	0.67
	γ (−)	Fixed	0.001	0.001	0.001
	a (cm)	Fixed	11.95	4.85	1.89
	K_{sa} (cm h^{-1})	0.07–4.15[a] 0.25–13.87[b]	0.26	4.17	4.17

[a]Value is best suited for the single macropore column.
[b]Value is best suited for the multiple macropore columns.

MCMC are also based on the Geweke test statistic [*Geweke*, 1992]. The Geweke test splits the MCMC chain into two "windows": the first window containing the beginning 20% of the chain, and the second usually containing the last 50% of the chain. If the samples are drawn from the stationary distribution of the chain, the mean of the two windows are equal. A *Z*-test of the hypothesis of equality of these two means is carried out and the chi-squared marginal significance is reported. A value of <0.01 for the chi-squared estimate indicates that the mean of the series is still drifting.

3.4.2. Number of Simulations

[33] *Raftery and Lewis*'s [1992] method is intended to detect convergence to the stationary distribution as well as to provide the total number of iterations required to estimate quantiles of any MCMC output with desired accuracy. The estimation of quantiles is very useful as they provide robust estimates of the mean and variability of the parameter. If the number of iterations is insufficient, the diagnostic process can be repeated to verify the minimum number of samples ($N_{\min}$) that should be run. One can determine the increment required in the number of simulations because of the dependence (I) in the sequence:

$$I = \frac{B + T}{N_{\min}}, \tag{17}$$

where B is the number of initial iterations to be discarded and commonly referred to as the burn-in length, and T is the total number of simulations. Values of I larger than 5 indicate strong autocorrelation which may be due to a poor choice of starting value, high posterior correlations, or stickiness of the MCMC algorithm.

4. Results

4.1. Sensitivity Analysis

[34] The objective of sensitivity analysis is to evaluate appropriate range of parameters and identify critical values that may lead to suboptimal or local solutions. In this study, sensitivity analysis is carried out by individually varying each parameter by $\pm 30\%$ and keeping the rest of the parameters constant at their inversely estimated values. Table 3 lists the top three parameters that produced the most sensitivity to modeled preferential flow results when compared with the optimal HYDRUS simulation. This ranking suggests that variations in matrix parameters cause larger sensitivity than macropore parameters for preferential flow through experimental soil columns. Tortuosity of the matrix domain (l_m), and residual water content (θ_r) for the matrix and macropore domains are not considered sensitive parameters as they result in small changes to the optimal HYDRUS simulation. Therefore, these parameters are disregarded for uncertainty evaluation using MCMC simulations essentially to curtail the dimensionality of the problem.

Table 3. Sensitive Parameters for Different Types of Experiments of the Single and Multiple Macropore Columns

Column Type	Type of Experiment	Sensitive Parameters		
Single macropore	Infiltration (0 cm head)	θ_{sm}	α_m	n_m
	Drainage	θ_{sm}	α_m	n_m
Low-density multiple macropore	Infiltration (6 cm head)	θ_{sm}	n_m	–
	Drainage	n_f	θ_{sm}	–
High-density multiple macropore	Infiltration (0 cm head)	θ_{sm}	α_m	–
	Infiltration (4 cm head)	θ_{sm}	n_f	n_m
	Drainage	θ_{sm}	n_m	–

4.2. Comparison of Adaptive and Conventional MH Algorithms

[35] MCMC iterations are run for developing an initial covariance structure among the soil hydraulic parameters for the experimental soil columns. Although more than 50% acceptance ratio is observed for all experiments, the initial 4000 MCMC samples do not show convergence for certain parameters (not shown here). Specifically, the pore size distribution index for the matrix domain (n_m), saturated water content for the matrix (θ_{sm}), and fracture (θ_{sf}) domains do not converge for any of the soil columns. Among these, θ_{sm} and n_m are found to be sensitive parameters for most of the experiments (Table 3). However, another common sensitive parameter α_m seems to converge efficiently. We argue that it is not the information in the measurements that is lacking but in extracting information about the interactions of the parameters which restricts us from obtaining a unique parameter set. By simultaneously using a number of correlated parameters, the identification of unique dual-permeability parameters is at stake. This result is confirmed by posterior cross-correlation plots, which show high correlation between parameters such as $\theta_{sm}{-}n_m$, $\theta_{sm}{-}\alpha_m$, and $\theta_{sf}{-}n_f$ for different experiments of the soil columns. A correlation among soil hydraulic parameters is not uncommon, however, prior information about correlation between the soil properties is nonexistent for most soils [*Vrugt et al.*, 2003; *Pollacco et al.*, 2008]. Therefore, the initial covariance structure ($\boldsymbol{\Sigma}_0$) of the parameters for both MCMC techniques is obtained from the initial 4000 MCMC simulations for all types of flow experiments as follows:

$$\begin{aligned} \textstyle\sum_0(\boldsymbol{\Theta}_m^a) &= E(\{\boldsymbol{\Theta}_m^a - E[\boldsymbol{\Theta}_m^a]\}^T\{\boldsymbol{\Theta}_m^a - E[\boldsymbol{\Theta}_m^a]\}), \\ \boldsymbol{\Theta}_m^a &= \{(\theta_{rm}^a, \theta_{sm}^a, \alpha_m^a, n_m^a, K_{sm}^a)\}, \end{aligned} \tag{18}$$

$$\begin{aligned} \textstyle\sum_0(\boldsymbol{\Theta}_f^a) &= E(\{\boldsymbol{\Theta}_f^a - E[\boldsymbol{\Theta}_f^a]\}^T\{\boldsymbol{\Theta}_f^a - E[\boldsymbol{\Theta}_f^a]\}), \\ \boldsymbol{\Theta}_f^a &= \{(\theta_{rf}^a, \theta_{sf}^a, \alpha_f^a, n_f^a, K_{sf}^a, l_f^a)\}, \end{aligned} \tag{19}$$

where, E is the mathematical expectation, a is the number of accepted samples from the initial 4000 MCMC simulations after 10% burn-in and thinning, and $\boldsymbol{\Theta}_m^a$ ($\boldsymbol{\Theta}_f^a$) is the set of random matrix (macropore) parameters as well as θ_{rm} (θ_{rf}) as suggested in section 3.2. The covariance with respect to interface parameters is limited to the variance of K_{sa} as the rest of the parameters are constant (section 3.2). Our goal here is to compare the traditional Metropolis-Hastings (MH) and the adaptive (AMCMC) techniques in estimating soil hydraulic parameters and in producing meaningful outputs that mimic the properties of our preferential flow system.

[36] After initializing the covariance structure, the MH and AMCMC techniques were used to determine uncertainty in the random parameter set $\{\boldsymbol{\varphi}_m, \boldsymbol{\varphi}_f, \boldsymbol{\varphi}_i\}$ for an

infiltration experiment of the single macropore column. Although both algorithms share the HYDRUS-optimized starting values and parameter priors, *Raftery and Lewis*'s [1992] diagnostic indicates 3295 additional iterations for the MH algorithm as opposed to 235 additional iterations for AMCMC to estimate 0.975 quantile of the parameters to the specified accuracy (equal to 0.02). Figure 3 presents contrasting posterior parameter distributions for the two algorithms. Since the truth about parameter distributions is unknown, there is no way to ascertain which algorithm predicts the correct posterior. However, the prediction of a unimodal distribution for K_{sf} by the MH algorithm implies that the chain takes a long time to move away from a local mode because of the single-update mechanism of the MH algorithm. On the other hand, the identification of a multimodal distribution for K_{sf} and l_f in the vicinity of local maxima is suggestive of desirable convergence and mixing characteristics of the AMCMC algorithm. The mean acceptance rate of AMCMC (34%) as compared to the MH algorithm (43%) is also suggestive of the comparatively slow convergence of the MH algorithm.

[37] The MCMC procedure is also carried out for infiltration experiments with constant pressure head boundary conditions for the low- and high-density multiple macropore columns. Parameter trace plots for 5000 and 3000 simulations for these experimental soil columns are shown in Figures 4 and 5, respectively. Figure 4 indicates that the sequence of draws converged quickly, within 5000 iterations, using the AMCMC technique. The performance of both algorithms is similar except for parameters such as θ_{sm}, n_m, θ_{sf}, K_{sf}, and K_{sa}. Many more iterations are required to obtain convergence and/or better mixing with the MH approach. Since smoothness of the running mean plots is an indicator of good mixing of the MCMC chain, Figure 6 compares the running mean plots of n_m and n_f parameters of the low-density macropore column for the two algorithms. This plot suggests slow mixing of the MH chain as compared to the AMCMC chain for both of the parameters. Geweke's diagnostic is also used to assess chain convergence and rejects convergence of θ_{sf} and K_{sa} at 90% level of significance using the MH algorithm (column 5 of Table 4). On the other hand, Geweke's statistic indicates satisfactory convergence (chi-squared probability >0.01) for all dual-permeability parameters using the AMCMC algorithm (column 6 of Table 4). The higher acceptance rate of 33% for the MH algorithm again confirms the slow mixing and convergence characteristics of this algorithm as compared to the lower mean acceptance (26%) of the AMCMC technique.

[38] Consistent with findings from the single macropore and low-density multiple macropore columns, the AMCMC algorithm provides better mixing and convergence with a 36% acceptance rate for the dual-permeability parameters of the high-density macropore column (Figure 5). This time series plot shows poor mixing (θ_{sm}) and trends in data (α_m, n_m, θ_{sf}, and α_f) at the 45% acceptance rate for the conventional MH algorithm. The results of the Geweke test confirm the lack of convergence for some of these dual-permeability parameters (n_m and θ_{sf}) using the MH algorithm (last two columns of Table 4). Raftery and Lewis's convergence diagnostic also indicates high autocorrelation ($I > 5$) in all parameters except K_{sm} and l_f for the MH algorithm and in θ_{sm} and n_m for the AMCMC algorithm for the high-density macropore column (Table 5). Since the statistic is calculated before thinning of the chains, autocorrelation observed in θ_{sm} and n_m using AMCMC, and n_f, K_{sf}, and K_{sa} using MH is expected as the chain is not independent and identically distributed (i.i.d.) as yet. The burn-in length (B) and additional number of samples (not shown here) obtained from the Raftery-Lewis statistic are not unreasonable even for the MH algorithm, however, this problem may worsen with addition of parameters, changes to correlation structure, and increment in desired accuracy.

[39] The nonconvergent parameters across the different experiments using the conventional Metropolis-Hastings

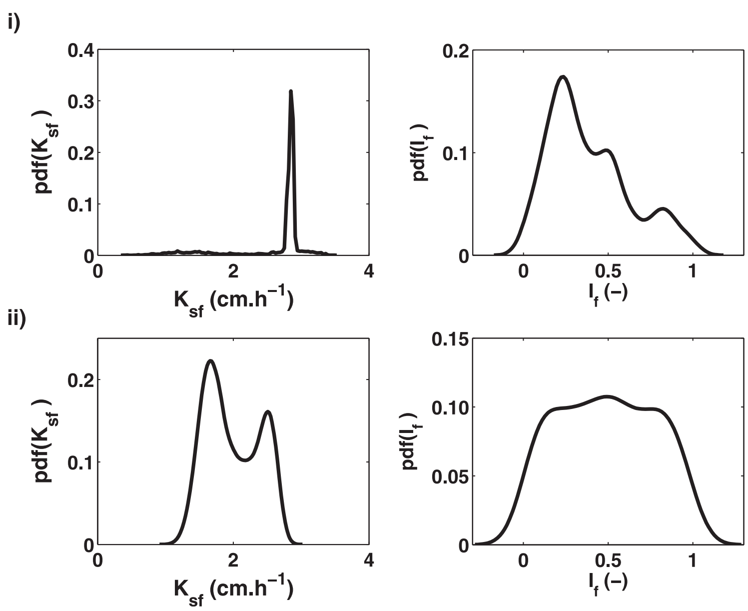

Figure 3. Posterior distributions of K_{sf} and l_f using (i) MH and (ii) AMCMC algorithms for an infiltration experiment of the single macropore column.

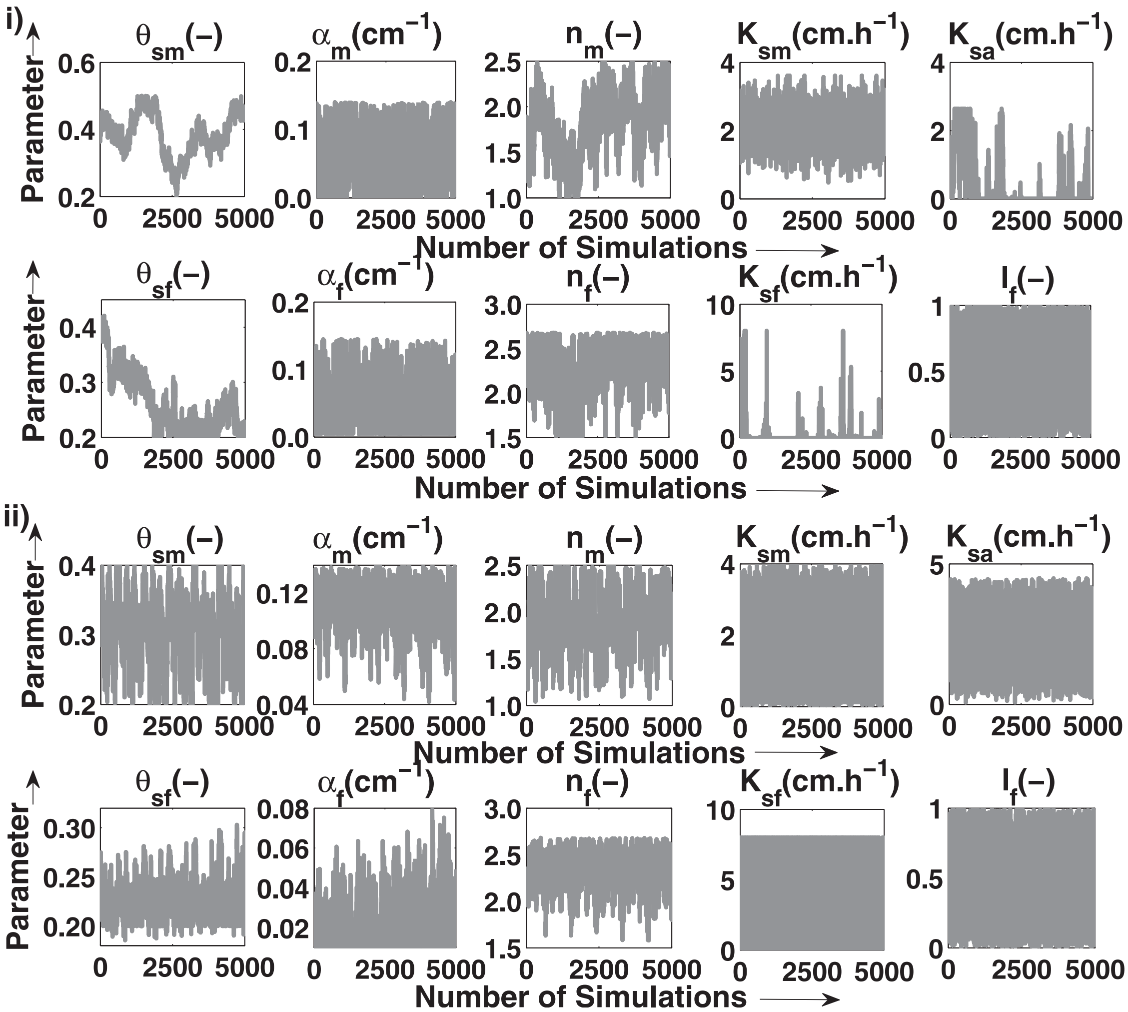

Figure 4. Parameter trace plots using (i) MH and (ii) AMCMC algorithms for an infiltration experiment (6 cm head) of the low-density multiple-macropore column.

algorithm do not have a direct relationship with the listed sensitive parameters for the different soil columns (Table 3). We argue that the MH algorithm was analyzing the trade-offs in the fitting of these highly correlated parameters due to its one-parameter-at-a-time updating approach. This argument is further strengthened by investigations into posterior cross-correlations among the simulated matrix and macropore domain parameters. Figures 7 and 8 present a scatterplot of parameters generated by the MH and AMCMC algorithms after convergence has been achieved for infiltration experiments of the low- and high-density multiple macropore columns, respectively. Specifically, parameter correlations ($|r| > 0.5$) are evident for θ_{sm} with α_m and n_m for the low-density macropore column, and θ_{sf} with α_f for the high-density macropore column using the MH algorithm. For the high-density macropore column, the scatterplots developed using the AMCMC algorithm are patchy only at the ends with respect to K_{sf}, while the MH algorithm produces scatterplots that are patchy within the parameter space for almost all of the macropore parameters. This suggests that the MH algorithm has been unable to cover the entire parameter space and explore the full posterior distribution of the parameters in the given number of iterations due to evident correlations between the parameters. On the other hand, the simultaneous updating of the parameters within the AMCMC algorithm enables it to provide better posterior estimates in lesser iterations. We conclude that carefully formulated AMCMC yields sufficient information to estimate parameter uncertainty with a faster convergence rate when a large number of parameters (as in a dual-permeability model) are considered random and prior information with respect to their interdependence and correlation is lacking.

4.3. Ouput Uncertainty

[40] To verify whether improved predictions of preferential flow can be made by either algorithm, we compare AMCMC and MH simulation results for a constant head (0 cm) infiltration experiment of the high-density multiple macropore column. Figure 9 illustrates pressure head profiles at 10 cm depth and soil water retention curves for the matrix domain for the two algorithms. The MH algorithm displays a wider range of uncertainty in predicting the entire pressure head profile as compared to the AMCMC algorithm. This is also true for water content profiles at all depths and for all experiments of the different soil columns (not

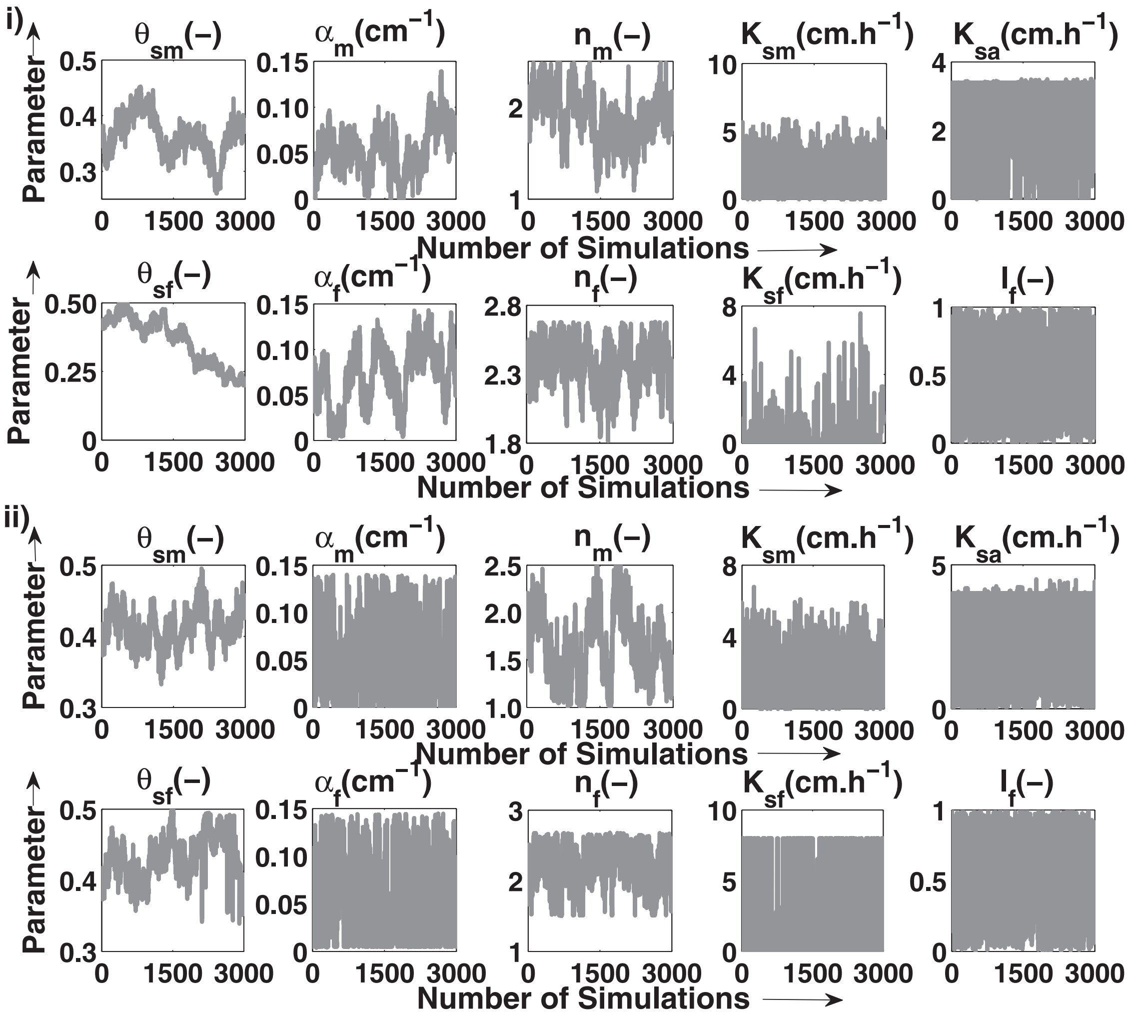

Figure 5. Parameter trace plots using (i) MH and (ii) AMCMC algorithms for an infiltration experiment (4 cm head) of the high-density multiple-macropore column.

shown here). This can be explained with the reasoning that the AMCMC algorithm has a narrow range of the highest-posterior density region in the physically plausible space for each of the dual-permeability parameters. The AMCMC algorithm is able to resolve parameter correlations and consequently, has a lower uncertainty associated with the dual-permeability parameters. On the other hand, the MH algorithm relies on the inverse procedure, which minimizes the squared residuals between model predictions and measurements, and fails to provide a single, relatively unique set of hydraulic parameters from experimental observations. This is also reflected in the 99% prediction uncertainty bounds where the most optimal hydraulic properties, obtained from the inverse procedure and indicated with the dotted line (Figure 9), are at the center of the bounds for the pressure head curve. On the contrary, the observations,

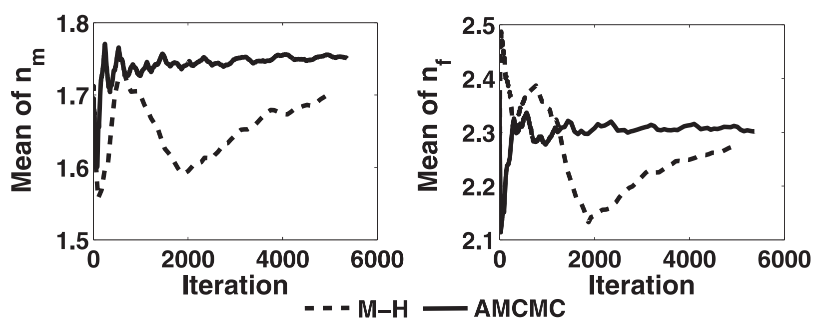

Figure 6. Moving average plots for n_m (−) and n_f (−) for an infiltration experiment of the low-density multiple-macropore column.

Table 4. Geweke Convergence Diagnostics Following 10% Burn-In for Dual-Permeability Parameters of Single and Multiple Macropore Columns

Dual-Permeability Parameters		Chi-Squared Probability[a]					
		Single Macropore Column		Low-Density Macropore Column		High-Density Macropore Column	
		MH	AMCMC	MH	AMCMC	MH	AMCMC
Matrix or immobile region	θ_{sm} (−)	$\underline{0.003}$	0.728	0.963	0.330	0.807	0.992
	α_m (cm^{-1})	0.610	0.164	0.355	0.205	0.060	0.127
	n_m (−)	0.960	0.180	0.057	0.934	$\underline{0.001}$	0.147
	K_{sm} (cm h^{-1})	0.632	0.209	0.190	0.809	0.157	0.163
Macropore or mobile region	θ_{sf} (−)	0.898	0.246	$\underline{0.001}$	0.161	$\underline{0.001}$	0.182
	α_f (cm^{-1})	0.363	0.507	0.155	0.155	0.023	0.870
	n_f (−)	0.234	0.260	0.147	0.579	0.758	0.698
	K_{sf} (cm h^{-1})	$\underline{0.001}$	0.448	0.056	0.294	0.268	0.134
	l_f (−)	0.413	0.944	0.798	0.691	0.105	0.680
Interface region	K_{sa} (cm h^{-1})	$\underline{0.008}$	0.336	$\underline{0.001}$	0.439	0.342	0.777

[a]Underline indicates chi-squared probability <0.01.

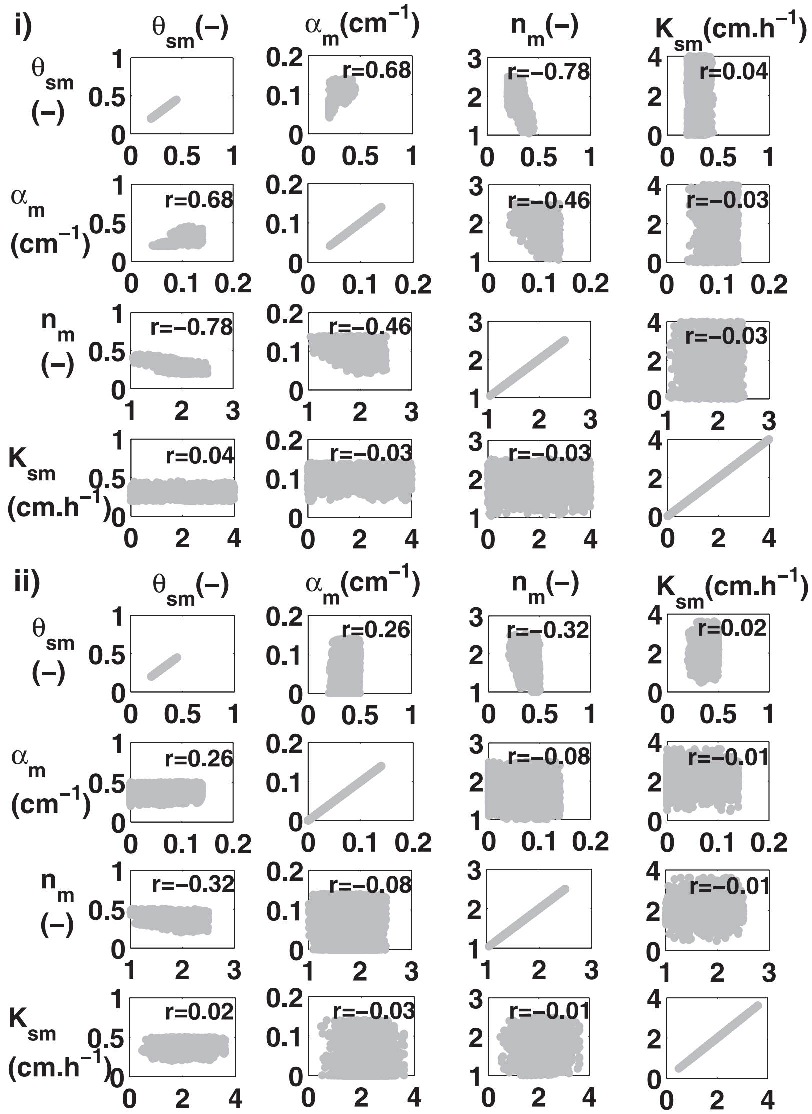

Figure 7. Scatterplots of 5000 combinations of different matrix parameters for the low-density macropore column using (i) MH and (ii) AMCMC algorithms.

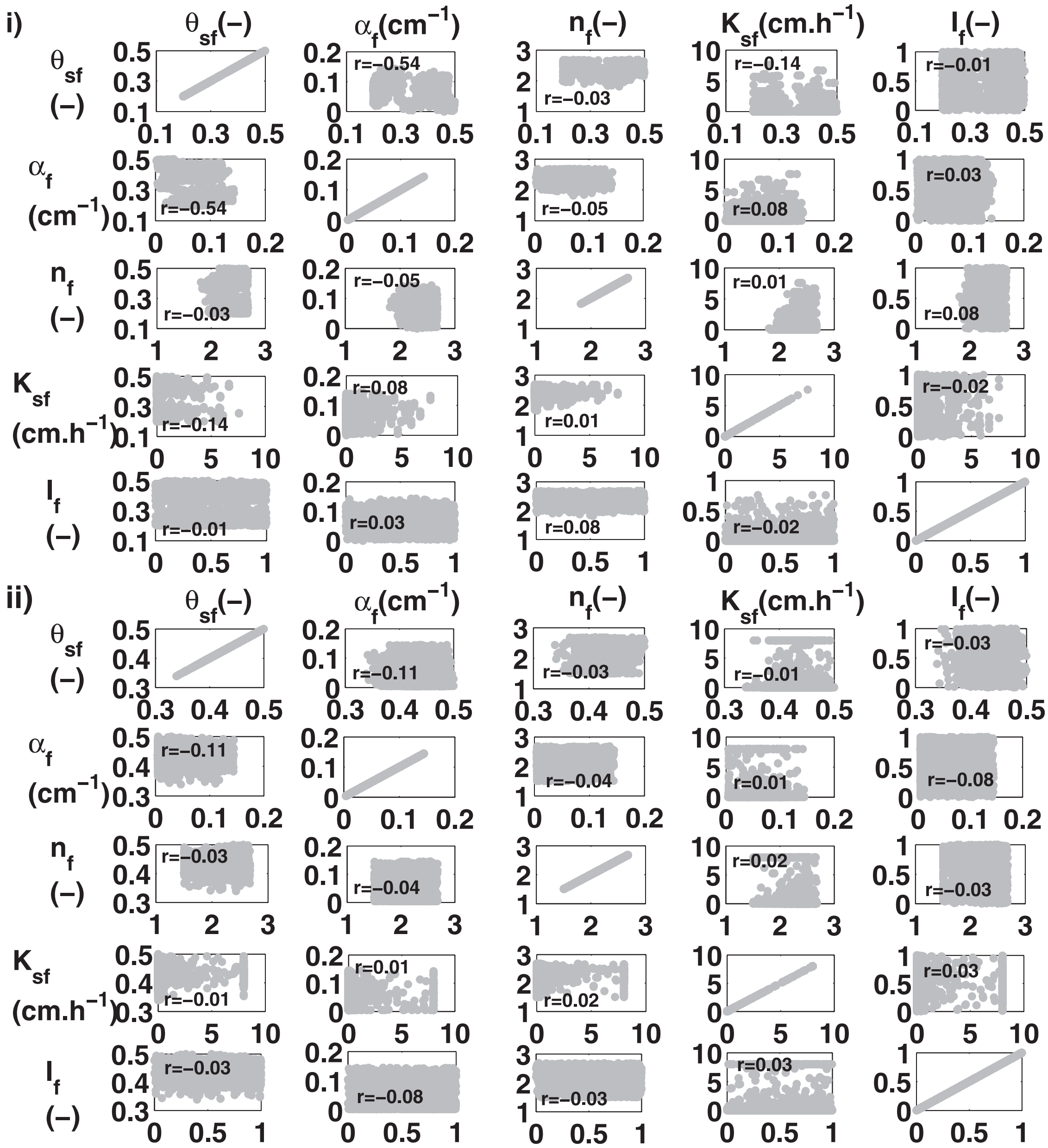

Figure 8. Scatterplots of 3000 combinations of different macropore parameters for the high-density macropore column using (i) MH and (ii) AMCMC algorithms.

indicated with squares, are at the center of the prediction bounds for the AMCMC algorithm, especially during perturbations of the pressure head potential between 12 and 18 h. There is also considerable uncertainty associated with the MH algorithm where the soil moisture potential is at saturation. This is in agreement with θ_{sm} being highly correlated with other parameters (Figure 7) and the high sensitivity of preferential flow output associated with θ_{sm} for all experiments (Table 3). It is important to note that AMCMC is not deemed better due to the smaller uncertainty range in output predictions as true uncertainty bounds are unknown for the experimental soil columns. However, we believe that significant uncertainly associated with the fitted soil water retention functions is due to unresolved parameter correlations using the MH algorithm. It is therefore recommended that additional water content measurements at a lower pressure potential be included to condense parameter correlations and reduce uncertainty associated with such parameter sampling algorithms. For the dual permeability modeling framework, the comparison between MH and AMCMC algorithms clearly demonstrates that correlations among dual-permeability parameters are present, and the output uncertainty range suggests that these correlations must be accounted for by the parameter sampling algorithms (either by including additional information on the correlation structure or through an efficient sampling scheme).

4.4. Uncertainty in Soil Hydraulic Parameters

[41] The estimation of marginal posterior distribution is obtained assuming homoscedatic, uncorrelated error terms using the adaptive MCMC technique. Histograms of the dual-permeability parameters generated after convergence

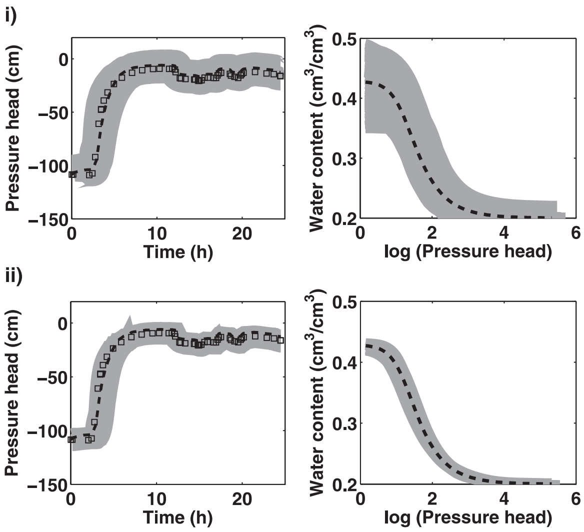

Figure 9. Uncertainty in predicting pressure head profiles and θ-h curves of the high-density multiple-macropore column for an infiltration experiment using (i) MH and (ii) AMCMC algorithms. The dashed lines define the HYDRUS simulation for the most likely parameter set, the gray shaded area denotes the 99% prediction uncertainty range, and the squares correspond to experimental observations at 10 cm depth.

to the stationary posterior distribution for drainage experiments of the single macropore and high-density multiple macropore columns are shown in Figures 10 and 11, respectively. The posterior distributions show evidence of the bi- and multimodal nature for certain soil hydraulic parameters. An explanation for the occurrence of multiple modes in the posterior is the inherent structure of the prior distribution. Multivariate normal priors can result in multimodal or Student's *t*-type of posterior distributions [*Escobar and West*, 1995]. For the soil column data, the different modes suggest that the experimental data are coming from two (or three) sets of population, which represent the different retention and hydraulic conductivity functions. *de Rooij et al.* [2004] obtained different modes with the same parametric distribution for soil hydraulic parameters of the plow layer and the subsoil thereby reflecting different soil depths and different retention functions. This result can be transferred here to suggest that these modes are related to the different domains of the dual-permeability system. The relative dominance of the matrix, macropore, and interface regions is easy to discern in the histograms. Specifically, in the drainage experiments of the single macropore and high-density multiple macropore columns, the macropores are drained first, followed by the matrix-macropore interface, and then by pores of the matrix domain. Therefore, the existence of three modes in K_{sm}, K_{sa}, θ_{sf}, and l_f for the single macropore column, and in K_{sa}, θ_{sm}, θ_{sf}, n_f, and l_f for the high-density multiple macropore column suggest the participation of these parameters in controlling flow processes through the matrix, macropore, and interface regions. Note that the parameters showing bimodality such as α_m, α_f, and n_f for the single macropore column, and K_{sm} and α_f for the high-density multiple macropore column belong to matrix and macropore domains only. This suggests that apart from the conductivity parameter of the matrix-macropore interface (K_{sa}) and the tortuosity of the macropores (l_f), soil hydraulic parameters of matrix and macropore domains also play an important role in regulating the flow through the interface region.

[42] Table 6 summarizes the posterior mean and variance of the various dual-permeability parameters for drainage experiments of the single and multiple macropore columns using the AMCMC algorithm. Since the MH algorithm produces incorrect posterior means and large variances for certain highly correlated variables, these results are not presented here. It is important to note that same initial parameters were employed for all of the soil columns and the only difference between them was in the number of macropores and therefore, in the geometry-based interface parameters (Table 2). The results presented in Table 6 illustrate that we end up with different parameter means for the different experimental columns. Most importantly, the posterior means of the K_s parameter for the matrix, macropore, and interface regions show a similarity between the low- and

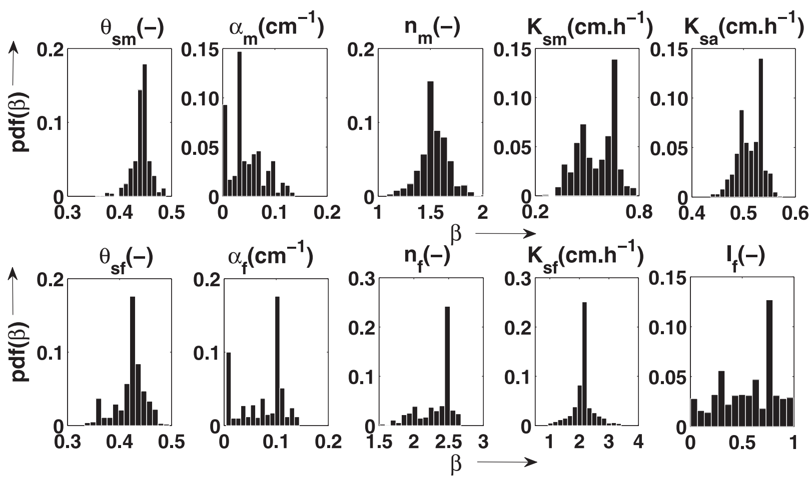

Figure 10. Posterior probability distributions of the parameters using observed data for drainage experiment of the single macropore column.

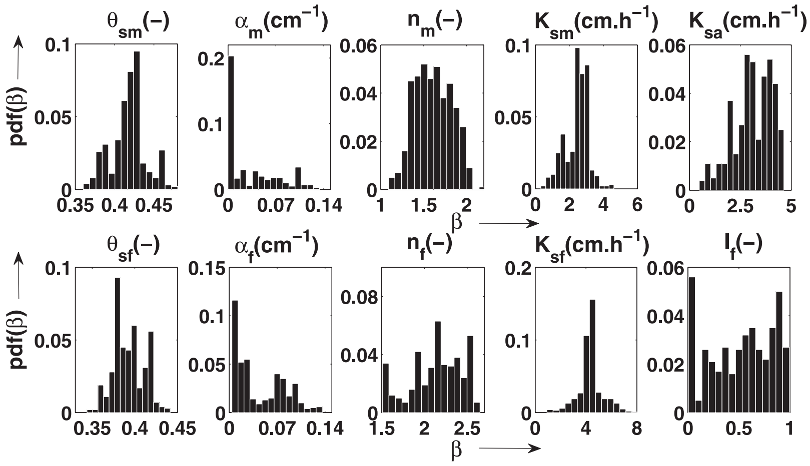

Figure 11. Posterior probability distributions of the parameters using observed data for drainage experiment of the high-density multiple macropore column.

high-density macropore columns, and are consistently lower for the single macropore column. Also, saturated hydraulic conductivity for the macropore domain (K_{sf}) is found to have the highest posterior variance for all the soil columns. This suggests that saturated hydraulic conductivity parameter is influenced by macropore density. Mild nonequilibrium conditions observed in the single macropore column are reflected through low posterior mean of K_s parameters for all three regions. Results from our previous study [*Arora et al.*, 2011] also indicate the need to adjust saturated hydraulic conductivity parameter (K_{sm}) to account for an increase in macropore density and to correctly predict flow through the structured soil system.

4.5. Comparison with Deterministic Approach

[43] For the sake of comparison with the stochastic/Bayesian approach, a deterministic framework is applied using a similar weighted least squares approach as described in equation (14):

$$\psi(\mathbf{\Theta}) = \sum\nolimits_{j=1}^{m} v_j \sum\nolimits_{i=1}^{n_j} w_{i,j}[r(\mathbf{\Theta})]^2, \tag{20}$$

Table 5. Evaluation of the Raftery-Lewis Statistic for Dual-Permeability Parameters of the High-Density Multiple Macropore Column

	MH		AMCMC	
Parameters	I	B	I	B
θ_{sm} (−)	34.54	138	10.17	41
α_m (cm^{-1})	17.83	99	1.81	4
n_m (−)	10.36	42	11.62	49
K_{sm} (cm h^{-1})	0.96	2	0.96	2
θ_{sf} (−)	67.71	195	3.82	20
α_f (cm^{-1})	8.29	46	2.63	5
n_f (−)	18.78	78	3.32	12
K_{sf} (cm h^{-1})	6.03	21	2.42	5
l_f (−)	0.71	3	0.71	3
K_{sa} (cm h^{-1})	38.69	103	5.89	21

where w_{ij}'s are equal to 1 (as in the stochastic approach), m is the number of different sets of measurements, n_j is the number of observations in a particular measurement set such that the total number of observations N (in equation (14)) is a summation of n_j (for $j = 1, 2, \ldots, m$). An additional set of weights (v_j) associated with each measurement set is used in the deterministic approach. The weighting elements v_j are inversely related to measurement variances (σ_j^2) and number of data (n_j) [*Clausnitzer and Hopmans*, 1995]:

$$v_j = \frac{1}{n_j \sigma_j^2}. \tag{21}$$

An advantage of the Bayesian approach is that it integrates out the error related to measurement variances (equation (16)). As mentioned in section 4.4, the deterministic approach resulted in similar parameters for all the soil columns except the interface parameters (Table 2) and suggested changes to K_{sm} for incorporating the effect of macropore density [*Arora et al.*, 2011]. On the other hand, the Bayesian framework resulted in consistently lower posterior means for K_s parameters for all regions of the single macropore column as compared to the multiple macropore columns. Thus, AMCMC suggests that the impact of macropore density be incorporated by calibrating saturated hydraulic conductivity parameters for all three regions. Another difference between the two approaches is highlighted through hydrologic outputs from the soil columns. The Bayesian framework provides a comprehensive evaluation of multiple realizations of preferential flow output from the columns using uncertain parameters, while the deterministic approach provides a single realization of the output (Figure 9). This single realization does not even lie at the center of the 99% uncertainty bounds obtained through AMCMC because the deterministic approach also analyzes parameter tradeoffs due to correlation among DPM parameters. We must mention that the Bayesian technique does not consider error related to the model structure. The use of a probabilistic framework in this study was solely to emphasize the correlation structure of DPM

Table 6. Summary of Posterior Distributions for the Soil Hydraulic Parameters Using the AMCMC Algorithm

Dual-Permeability Parameters		Single Macropore Column		Low-Density Macropore Column		High-Density Macropore Column	
		Mean	Variance	Mean	Variance	Mean	Variance
Matrix or immobile region	θ_{sm} (−)	0.457	0.024	0.304	0.054	0.413	0.024
	α_m (cm^{-1})	0.070	0.027	0.107	0.020	0.060	0.026
	n_m (−)	1.725	0.323	1.904	0.320	1.663	0.342
	K_{sm} (cm h^{-1})	0.434	0.091	2.097	1.002	2.603	1.020
Macropore or mobile region	θ_{sf} (−)	0.256	0.046	0.226	0.020	0.433	0.029
	α_f (cm^{-1})	0.058	0.026	0.021	0.015	0.061	0.034
	n_f (−)	2.220	0.237	2.302	0.223	2.258	0.285
	K_{sf} (cm h^{-1})	2.518	1.092	3.871	1.518	3.530	1.326
	l_f (−)	0.494	0.028	0.530	0.029	0.510	0.028
Interface region	K_{sa} (cm h^{-1})	0.524	0.034	2.508	1.029	2.311	1.001

parameters and its effect on posterior parameter values, uncertainty limits, and hydrological output.

5. Summary and Conclusions

[44] The applicability of dual-permeability models for structured soils is hindered by the large number of input parameters, some of which cannot be measured directly [*Šimùnek et al.*, 2003]. This study depicts the usefulness of Bayesian methods in evaluating parameter uncertainty and its effect on model predictions in a preferential flow system that considers 10 out of 17 (or 13 based on degrees of freedom) DPM parameters to be random. Bayesian modeling framework is applied using an adaptive MCMC scheme and the conventional Metropolis-Hastings algorithm on experimental soil columns with different macropore distributions (single macropore, low-, and high-density multiple macropores). The distinguishing feature of the AMCMC algorithm is its simultaneous parameter update due to the description of the parameter covariance matrix as opposed to the single-site update of the MH algorithm. The results indicate that AMCMC accelerates convergence of the multidimensional dual-permeability model for all experimental soil columns and identifies marginal posterior distributions even in the vicinity of local maxima due to its online updating mechanism. On the other hand, the MH algorithm reveals high-posterior correlations obtained with respect to θ_{sm} with n_m and α_m, and θ_{sf} with α_f for different experiments of the soil columns. In terms of predicting preferential flow, this study shows that the MH algorithm produces larger uncertainties than AMCMC in pressure head and water content profiles at different depths of the soil columns. The larger variability near the saturation end of the water retention curve using the MH algorithm is related to high correlations with θ_{sf} and high sensitivity of preferential flow estimates to the saturated water content parameter (θ_{sm}). It seems that the MH algorithm requires additional experimental data sets or supplemental information on parameter covariance structure to resolve these correlations efficiently while AMCMC has faster convergence in estimating unique parameters using just the information contained in experimental observations. For the dual-permeability framework, the comparison between the two algorithms highlights the existence of a correlation structure among DPM parameters and indicates that the selection of parameter sampling algorithms, whether deterministic or stochastic, is paramount in obtaining unique DPM parameters. When correlation structure of dual-permeability parameters is unknown or complex, the parameter sampling schemes should either have efficient update mechanisms (e.g., AMCMC) or be supplied with supplemental information (e.g., MH) to improve identification of DPM parameters. Other studies have also reported that prior knowledge about correlation structure significantly improves equifinality of parameter estimates [*Flores et al.*, 2010; *Scharnagl et al.*, 2011].

[45] In terms of parameter uncertainty, both order and value of parameters are well-estimated and within credible limits according to the UNSODA database using the AMCMC algorithm [*Nemes et al.*, 1999, 2001]. The effect of macropore density is evident in a saturated hydraulic conductivity parameter for matrix (K_{sm}), macropore (K_{sf}), and interface regions (K_{sa}) as their posterior means are consistently lower for the single macropore column as compared to the multiple macropore columns. A high posterior variance found in K_{sf} also reflects higher uncertainty in the consistency of this parameter across soil columns with changing macropore density. Our previous study also emphasizes the need to account for changes in macropore density through some parameters of the dual permeability model [*Arora et al.*, 2011]. Histograms of certain parameters are found to display bi- or tri-modal characteristics. We believe that this is not a peculiarity of the posterior distribution but reflects the sequence of flow processes of the matrix, macropore, and/or the interface region. This is similar to observations in natural systems, where macropores are predominantly active at and near saturation; the micropores get active at a relatively lower pressure; and the interface at a variety of pressure heads in between the extremes. Results indicate that the degree of local nonequilibrium in the matrix-macropore interface is controlled not only by the transfer term parameter (K_{sa}) and macropore tortuosity (l_f) but also by other parameters governing the shape of water retention curves for the matrix and macropore domains. This result is important from the perspective of understanding the physical meaning and effect of dual permeability parameters, and incorporating uncertainty in certain parameters to better account for lateral flow processes through the matrix-macropore interface region.

[46] We must note that theoretical concepts derived from this one-dimensional column study are applicable to multidimensional settings of structured soils. This is because preferential flow causes the majority of the flow (disregarding

macropore tortuosity and dead ends) to be carried through macropores and fractures, making the flow essentially one-dimensional [*Flury et al.*, 1994; *Mohanty et al.*, 1998]. Therefore, specific results like the existence of correlation among DPM parameters, the need for requisite changes to K_s to account for increase in macropore density, and the dominance of interface region in any flow process are all transferrable to the field scale. A recent study by *Kodêsová et al.* [2010] also demonstrates correlations with respect to K_{sf} with K_{sa}, and K_{sf} with shape parameters of the macropore domain for an experimental field setting. In addition, three-dimensional field settings can only enhance the problem of correlated parameters by introducing spatial correlation in the added dimension [*Mallants et al.*, 1997; *Coppola et al.*, 2009].

[47] **Acknowledgments.** This project was supported by the National Science Foundation (grant EAR 0635961). This research was partly supported by award 5R01ES015634 from the National Institute of Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health.

References

Arora, B., B. P. Mohanty, and J. T. McGuire (2011), Inverse estimation of parameters for multi-domain flow models in soil columns with different macropore densities, *Water Resour. Res.*, *47*, W04512, doi:10.1029/2010WR009451.

Atchadé, Y. F., and J. S. Rosenthal (2005), On adaptive Markov chain Monte Carlo algorithms, *Bernoulli*, *11*(5), 815–828.

Böhlke, J.-K. (2002), Groundwater recharge and agricultural contamination, *Hydrogeol. J.*, *10*, 153–179.

Carrera, J., and S. P. Neumann (1986), Estimation of aquifer parameters under transient and steady-state conditions: 2. Uniqueness, stability and solution algorithms, *Water Resour. Res.*, *22*, 211–227.

Castiglione, P., B. P. Mohanty, P. J. Shouse, J. Simunek, M. T. Van Genuchten, and A. Santini (2003), Lateral water diffusion in an artificial macroporous system: Modeling and experimental evidence, *Vadose Zone J.*, *2*, 212–221.

Clausnitzer, V., and J. W. Hopmans (1995), Non-linear parameter estimation: LM_OPT, General-purpose optimization code based on the Levenberg-Marquardt algorithm, paper presented at Land, Air and Water Resour. Pap. 100032, University of Calif., Davis.

Coppola, A., A. Basile, A. Comegna, and N. Lamaddalena (2009), Monte Carlo analysis of field water flow comparing uni- and bimodal effective hydraulic parameters for structured soil, *J. Contam. Hydrol.*, *104*(1–4), 153–165.

Cowles, M. K., and B. P. Carlin (1996), Markov Chain Monte Carlo convergence diagnostics: A comparative review, *J. Am. Statistics Assoc.*, *91*, 883–904.

de Rooij, G. H., R. T. A. Kasteel, A. Papritz, and H. Flühler (2004), Joint distributions of the unsaturated soil hydraulic parameters and their effect on other variates, *Vadose Zone J.*, *3*, 947–955.

Dubus, I. G., S. Beulke, and C. D. Brown (2002), Calibration of pesticide leaching models: Critical review and guidance for reporting, *Pest. Manage. Sci.*, *58*, 745–758.

Durner, W., E. B. Schultze, and T. Zurmühl (1999), State-of-the-art in inverse modeling of inflow/outflow experiments, in *Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, Proceedings of international workshop, Riverside, Calif., 22–24*, edited by M. Th. van Genuchten et al., pp. 661–681, U.S. Salinity Laboratory, Riverside, CA.

Escobar, M. D., and M. West (1995), Bayesian density estimation and inference using mixtures, *J. Am. Statistics Assoc.*, *90*, 577–588.

Flores, A. N., D. Entekhabi, and R. L. Bras (2010), Reproducibility of soil moisture ensembles when representing soil parameter uncertainty using a Latin hypercube-based approach with correlation control, *Water Resour. Res.*, *46*, W04506, doi:10.1029/2009WR008155.

Flury, M., H. Fluhler, W. A. Jury, and J. Leuenberger (1994), Susceptibility of soils to preferential flow of water: A field study, *Water Resour. Res.*, *30*(7), 1945–1954.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, 526 pp., Chapman and Hall, London, U. K.

Gerke, H. H., and M. T. van Genuchten (1993a), A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media, *Water Resour. Res.*, *29*, 305–319.

Gerke, H. H., and M. Th. van Genuchten (1993b), Evaluation of a first-order water transfer term for variably saturated dual-porosity models, *Water Resour. Res.*, *29*, 1225–1238.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to thecalculation of posterior moments, in *Bayesian Statistics 4*, edited by J. M. Bernado et al., pp. 169–193, Oxford Press, Oxford, U. K.

Ginn, T. R., and J. H. Cushman (1990), Inverse methods for subsurface flow: A critical review of stochastic techniques, *Stochastic Hydrol. Hydraul.*, *4*, 1–26.

Gwo, J. P., P. M. Jardine, G. V. Wilson, and G. T. Yeh (1995), A multiple-pore-region concept to modeling mass transfer in subsurface media, *J. Hydrol.*, *164*, 217–237.

Harrio, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli 7*(2), 223–242.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109.

Jamieson, R. C., R. J. Gordon, K. E. Sharples, G. W. Stratton, and A. Madani (2002), Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water—A review, *Can. Biosys. Eng.*, *44*, 1.1–1.9.

Jana, R. B., and B. P. Mohanty (2011), Enhancing PTFs with remotely sensed data for multi-scale soil water retention estimation, *J. Hydrol.*, *399*(3–4), 201–211.

Jarvis, N. J. (1994), The MACRO Model (version 3.1): Technical description and sample simulations, *Rep. and Dissertations 19*, Dept. of Soil Sci., Swedish Univ. of Agric. Sci. Uppsala, Sweden, p. 51.

Jarvis, N., M. Larsbo, S. Roulier, A. Lindahl, and L. Persson (2007), The role of soil properties in regulating non-equilibrium macropore flow and solute transport in agricultural topsoils, *Eur. J. Soil Sci.*, *58*(1), 282–292.

Kladivko, E. J., L. C. Brown, and J. L. Baker (2001), Pesticide transport to subsurface tile drains in humid regions of North America, *Crit. Rev. Env. Sci. Tec.*, *31*, 1–62.

Kodêsová, R., J. Šimùnek, A. Nikodem, and V. Jirku (2010), Estimation of the dual-permeability model parameters using tension disk infiltrometer and Guelph permeameter, *Vadose Zone J.*, *9*(2), 213–225.

Köhne, J. M., and B. P. Mohanty (2005), Water flow processes in a soil column with a cylindrical macropore: Experiment and hierarchical modeling, *Water Resour. Res.*, *41*, W03010, doi:10.1029/2004WR003303.

Köhne, J. M., B. P. Mohanty, J. Šimùnek, and H. H. Gerke (2004), Numerical evaluation of a second-order water transfer term for variably saturated dual-permeability models, *Water Resour. Res.*, *40*, W07409, doi:10.1029/2004WR003285.

Köhne, J. M., S. Köhne, and J. Šimùnek (2009), A review of model applications for structured soils: a) Water flow and solute transport, *J. Contam. Hydrol.*, *104*, 4–35, doi:10.1016/j.jconhyd.2008.10.002.

Kool, J. B., and J. C. Parker (1988), Analysis of the inverse problem for transient unsaturated flow, *Water Resour. Res.*, *24*, 817–830.

Larsbo, M., S. Roulier, F. Stenemo, R. Kasteel, and N. Jarvis (2005), An improved dual-permeability model of water flow and solute transport in the vadose zone, *Vadose Zone J.*, *4*, 398–406.

Mallants, D., B. P. Mohanty, A. Vervoort, and J. Feyen (1997), Spatial analysis of saturated hydraulic conductivity in a soil with macropores, *Soil Tec.*, *10*(2), 115–131.

Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, *40*, W02501, doi:10.1029/2003WR002378.

Mualem, Y. (1976), A new model for predicting the hydraulic conductivity of unsaturated soils, *Water Resour. Res.*, *12*, 513–522.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1091.

Mohanty, B. P., R. S. Bowman, J. M. H. Hendrickx, and M. Th. van Genuchten (1997), New piecewise-continuous hydraulic functions for modeling preferential flow in an intermittent-flood-irrigated field, *Water Resour. Res.*, *33*(9), 2049–2063.

Mohanty, B. P., R. S. Bowman, J. M. H. Hendrickx, J. Šimùnek, and M. Th. van Genuchten (1998), Preferential transport of nitrate to a tile drain in an intermittent-flood-irrigated field: Model development and experimental evaluation, *Water Resour. Res.*, *34*(5), 1061–1076.

National Research Council (1994), *Ground water recharge using waters of impaired quality*, 283 pp., Natl. Academy Press, Washington, D. C.

Nemes, A., M. G. Schaap, and F. J. Leij (1999), The UNSODA unsaturated soil hydraulic database, version 2.0, U. S. Salinity Laboratory, U. S. Dept. of Agriculture, Agricultural Research Service, Riverside, Calif.

Nemes, A., M. G. Schaap, F. J. Leij, and J. H. M. Wosten (2001), Description of the unsaturated soil hydraulic database UNSODA version 2.0, *J. Hydrol.*, *251*(3–4), 151–162.

Peters, G. W., B. Kannan, B. Lasscock, and C. Mellen (2009), Model selection and adaptive Markov chain Monte Carlo for Bayesian cointegrated VAR model, paper presented at 15th Interntl. Conf. on Computing in Economics and Finance, Sydney, Australia.

Pollacco, J. A. P., J. M. S. Ugalde, R. Angulo-Jaramillo, I. Braud, and B. Saugier (2008), A linking test to reduce the number of hydraulic parameters necessary to simulate groundwater recharge in unsaturated soils, *Adv. Water Resour.*, *31*(2), 355–369.

Raftery, A. E., and S. Lewis (1992), "How Many Iterations in the Gibbs Sampler?", in *Bayesian Statistics 4*, edited by J. M. Bernardo et al., pp. 763–773, Oxford Univ. Press, Oxford, U.K.

Reis, D. S., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, *J. Hydrol.*, *313*, 97–116.

Roulier, S., and N. Jarvis (2003), Analysis of inverse procedures for estimating parameters controlling macropore flow and solute transport in the dual-permeability model MACRO, *Vadose Zone J.*, *2*, 349–357.

Scharnagl, B., J. A. Vrugt, H. Vereecken, and M. Herbst (2011), Inverse modelling of in situ soil water dynamics: Investigating the effect of different prior distributions of the soil hydraulic parameters, *Hydrol. Earth Syst. Sci.*, *8*(1), 2019–2063.

Schwartz, R. C., A. S. R. Juo, and K. J. McInnes (2000), Estimating parameters for a dual-porosity model to describe non-equilibrium, reactive transport in a fine textured soil, *J. Hydrol.*, *229*, 149–167.

Šimùnek, J., and M. T. van Genuchten (2008), Modeling nonequilibrium flow and transport processes using HYDRUS, *Vadose Zone J.*, *7*(2), 782–797.

Šimùnek, J., O. Wendroth, N. Wypler, and M. Th. van Genuchten (2001), Non-equilibrium water flow characterized by means of upward infiltration experiments, *Eur. J. Soil Sci.*, *52*(1), 13–24.

Šimùnek, J., N. J. Jarvis, M. T. van Genuchten, and A. Gärdenäs (2003), Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone, *J. Hydrol.*, *272*, 14–35, doi:10.1016/S0022-1694(02)00252-4.

van Genuchten, M. T. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, *44*, 892–898.

van Genuchten, M. T., and P. J. Wierenga (1976), Mass transfer studies in sorbing porous media, I, analytical solutions, *Soil Sci. Soc. Am. J.*, *40*, 473–481.

Vrugt, J. A., W. Bouten, H. V. Gupta, and J. W. Hopmans (2003), Toward improved identifiability of soil hydraulic parameters: On the selection of a suitable parametric model, *Vadose Zone J.*, *2*, 98–113.

Vrugt, J. A., P. H. Stauffer, Th. Wöhling, B. A. Robinson, and V. V. Vesselinov (2008), Inverse modeling of subsurface flow and transport properties: A review with new developments, *Vadose Zone J.*, *7*, 843–864.

Wu, W., J. S. Clark, and J. M. Vose (2010), Assimilating multi-source uncertainties of a parsimonious conceptual hydrological model using hierarchical Bayesian modeling, *J. Hydrol.*, *394*(3–4), 436–446.

Zachmann, D. W., P. C. Duchateau, and A. Klute (1981), The calibration of the Richards' flow equation for a draining column by parameter identification, *Soil Sci. Soc. Am. J.*, *45*, 1012–1015.

B. Arora, and B. P. Mohanty, Department of Biological and Agricultural Engineering, Texas A&M University, MS 2117, College Station, TX 77843–2117, USA. (bmohanty@tamu.edu)

J. T. McGuire, Department of Geology and Geophysics, University of St. Thomas, St. Paul, MN 55105, USA.