Boise State University
ScholarWorks

Geosciences Faculty Publications and Presentations | Department of Geosciences

4-17-2010

Reproducibility of Soil Moisture Ensembles When Representing Soil Parameter Uncertainty Using a Latin Hypercube-Based Approach with Correlation Control

Alejandro N. Flores
Boise State University

Dara Entekhabi
Massachusetts Institute of Technology

Rafael L. Bras
University of California

Copyright 2010 by the American Geophysical Union. DOI: 10.1029/2009wr008155

WATER RESOURCES RESEARCH, VOL. 46, W04506, doi:10.1029/2009WR008155, 2010

Click Here for Full Article

Reproducibility of soil moisture ensembles when representing soil parameter uncertainty using a Latin Hypercube–based approach with correlation control

Alejandro N. Flores,[1] Dara Entekhabi,[2] and Rafael L. Bras[3]

Received 28 April 2009; revised 3 October 2009; accepted 4 November 2009; published 17 April 2010.

[1] Representation of model input uncertainty is critical in ensemble-based data assimilation. Monte Carlo sampling of model inputs produces uncertainty in the hydrologic state through the model dynamics. Small Monte Carlo ensemble sizes are desirable because of model complexity and dimensionality but potentially lead to sampling errors and correspondingly poor representation of probabilistic structure of the hydrologic state. We compare two techniques to sample soil hydraulic and thermal properties (SHTPs): (1) Latin Hypercube (LH) based sampling with correlation control and (2) random sampling from SHTP marginal distributions. A hydrology model is used to project SHTP uncertainty onto the soil moisture state for given forcings. For statistical comparison, we generate 20 ensembles for 7 ensemble sizes. Variance in ensemble moment estimates decreases with increasing ensemble size. The LH-based approach yields less variance in the estimate of ensemble moments at all ensemble sizes, an advantage greatest with small ensembles. Implications for hydrologic uncertainty assessment, data assimilation, and parameter estimation are discussed.

Citation: 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.

1. Introduction

[2] The objective of hydrologic data assimilation is to produce estimates of variables characterizing hydrologic conditions of a study area through the fusion of uncertain model estimates and noisy observations. Estimation techniques such as the Kalman filter provide mathematical frameworks to derive optimal estimates of the hydrologic state by weighting model estimates and observations by their respective degree of certainty [*Gelb*, 1974]. However, traditional Kalman filtering is often complicated because of the requirement of a linear model, or a linear approximation to a nonlinear model [e.g., *Galantowicz et al.*, 1999; *Hoeben and Troch*, 2000]. Ensemble-based data assimilation techniques such as the ensemble Kalman filter [e.g., *Evensen*, 1994; *Margulis et al.*, 2002; *Reichle et al.*, 2002; *Crow and Wood*, 2003; *Evensen*, 2004; *Moradkhani et al.*, 2005a], ensemble Kalman smoother [e.g., *Dunne and Entekhabi*, 2006], and other ensemble-based data assimilation techniques [*Dunne and Entekhabi*, 2005; *Moradkhani et al.*, 2005b], by contrast, have eliminated the need for model linearization in many circumstances. In ensemble-based data assimilation, Monte Carlo techniques are used to sample probability distributions that characterize the uncertainty in the model inputs (parameters and forcings), and the full nonlinear model dynamics are used to project these uncertainties onto the hydrologic state forward in time until an observation becomes available. Each Monte Carlo realization represents an equiprobable and physically plausible prediction of the hydrologic state of the system, given the uncertainty in the model inputs. The collection (ensemble) of plausible hydrologic states, taken together, characterizes the probabilistic structure of the hydrologic state, as represented by the dynamics of the model and uncertain model inputs: the "first guess" ensemble. At the time when an observation is available, some system of equations is used to simulate an observable quantity (e.g., brightness temperature, radar backscatter, river stage, etc.) based on the ensemble of model simulated hydrologic states, yielding a corresponding ensemble of predicted observations. Together, the first guess ensemble and predicted observations constitute the prior information about the hydrologic system under study. Ensemble-based data assimilation algorithms use this prior information to update or reweight the individual ensemble replicates to reflect new information contained in noisy observations. This updating or reweighting yields an ensemble of replicates characterizing the hydrologic state of the system in the model state space that is conditioned on the information contained in the observational data: the "analyzed" ensemble. The model is then reinitialized with these analyzed realizations of the model state, and the model is integrated forward in time under the influence of the uncertain inputs until another observation becomes available.

[1]Department of Geosciences, Boise State University, Boise, Idaho, USA.
[2]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[3]Henry Samueli School of Engineering, University of California, Irvine, California, USA.

Copyright 2010 by the American Geophysical Union.
0043-1397/10/2009WR008155

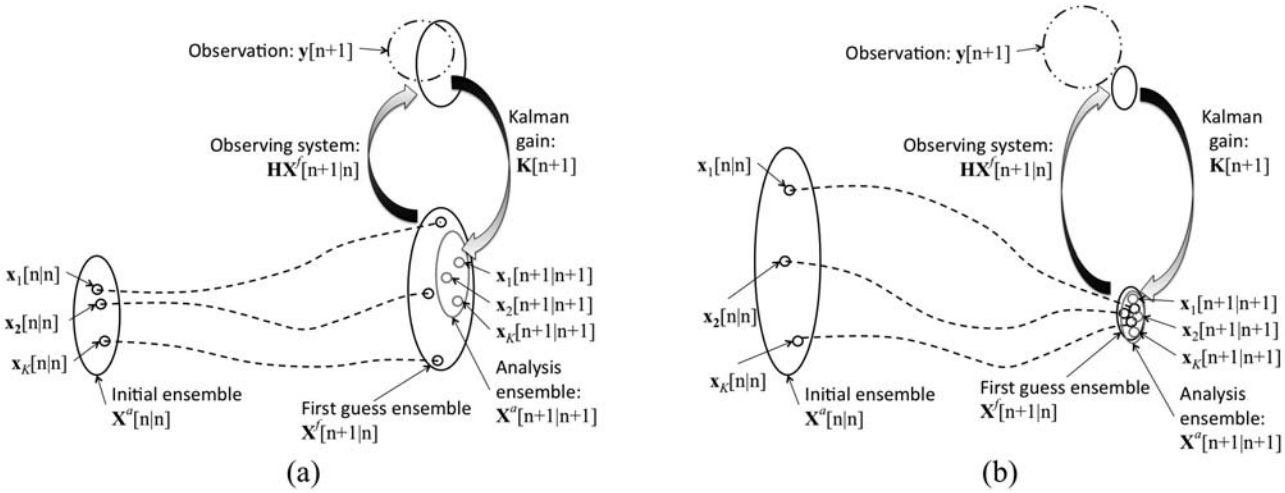

Figure 1. A conceptual diagram showing the ensemble Kalman filter process. (a) When uncertainty in the model forcings and parameters is well characterized, the uncertainty in the model estimate often increases as the time since the initialization increases and the assimilation of observations leads to a reduction of the uncertainty in the model predictions, relative to the first guess. However, (b) when uncertainty in forcings and parameters is not well characterized the uncertainty in the model estimate can decrease as time since initialization increases. Despite a large innovation (difference between the predicted observations and the observed data), the unreasonably low error in the first guess ensemble results in relatively little information propagating back to the model state estimate.

[3] Because it is composed of a finite number of realizations obtained by subjecting a model to uncertain hydrometeorological forcings (e.g., precipitation, temperature, radiation) and parameters (e.g., hydraulic conductivity, porosity, retention curve parameters), the first guess ensemble is an approximation of the probabilistic structure of the hydrologic state. As such, the manner in which uncertainty in the model forcings and parameters is represented, as well as the nature and form of structural errors embodied within the model, critically influence the ensemble-based approximation of the probabilistic structure of the hydrologic state and, therefore, the degree to which observational data will improve model estimates [e.g., see *Crow and van Loon*, 2006; *Reichle et al.*, 2008]. Neglecting or poorly representing potentially important sources of uncertainty in the model forcings or parameters can lead to a poor representation of the probabilistic structure of the hydrologic state. The importance of adequate representation of model input uncertainty in ensemble-based hydrologic data assimilation is particularly relevant given the high dimensionality and/or nonlinearity associated with hydrologic models, particularly those that simulate moisture and energy states in a spatially distributed fashion [e.g., *Carpenter et al.*, 2001; *Downer et al.*, 2002; *Ivanov et al.*, 2004a, 2004b; *Qu and Duffy*, 2007; *Ivanov et al.*, 2008a, 2008b]. Decreasing the ensemble size reduces computational costs associated with model simulation, but makes the Monte Carlo–based approximation to the probabilistic structure of the hydrologic state sensitive to errors in sampling the distributions meant to characterize uncertainty in model parameters and forcings (Figure 1a). For example, this is seen in Kalman-like ensemble updates as unrealistically low state error covariance, which leads to an unrealistically high confidence in the model estimate at the update and an insensitivity to observations (Figure 1b). Although the ensemble of predicted observations may be substantially different than the data, implying significant innovations, the unrealistically low state error covariance will limit the degree to which the innovations propagate back to the state. Conversely, unrealistically high state error covariance, would lead to an unrealistically high confidence in the observations.

[4] Imperfect characterization of soil hydraulic and thermal properties (SHTPs) is an important source of uncertainty in soil moisture estimates, and the associated latent and sensible heat fluxes, derived from land surface hydrology models [*Margulis et al.*, 2002; *Dunne and Entekhabi*, 2005, 2006]. The model parameters describing SHTPs in surface hydrology models affect the moisture-holding capacity of the soil and the rates of moisture redistribution and exchange between the land surface and the atmosphere. The characterization of soils in hydrology models leads to uncertainty in soil moisture for four important reasons: (1) simplifying assumptions made to enable mathematical description of complex and nonlinear hydrologic behavior of soils (e.g., hysteresis in the moisture retention curve), (2) imperfect laboratory and field techniques to estimate hydrologic parameters of in situ soils, (3) spatially sparse sampling and analysis of SHTPs, and (4) uncertainty in the definition and categorization of soil units in spatial databases even in data-rich regions of the world such as the Continental United States. In hydrologic modeling these uncertainties in SHTPs and the model parameters describing them can lead to significant uncertainty in model outputs of interest such as soil moisture, latent, and sensible heat fluxes.

[5] This study deals with the treatment of uncertainty in SHTPs and corresponding model parameters. The goal is not to perform parameter estimation, but rather to ensure that uncertainty in SHTPs is represented sufficiently well to ensure that (1) ensemble estimates of the variance in soil moisture are realistic and (2) ensemble estimates of the mean and variance in soil moisture are consistent (i.e., they

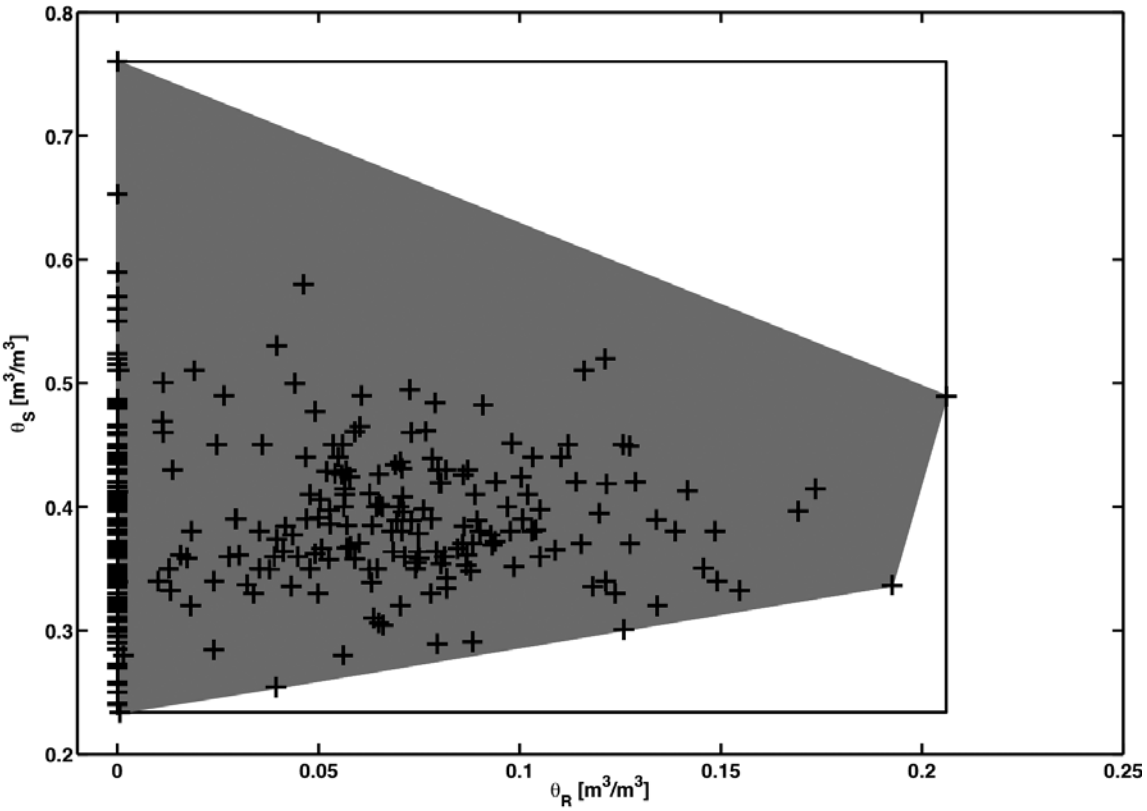

Figure 2. For the sandy loam soil texture, θ_S and θ_R are plotted based on the data of *Schaap and Leij* [1998]. The white area enclosed by the solid black lines represents the physically possible area of variation in θ_S and θ_R, given the extremes in the data set and assuming θ_S and θ_R are completely independent. The gray area is the convex hull enveloping the actual data from the database, which are plotted within this area as black plus symbols.

do not vary significantly across multiple ensembles, each using different stochastically generated or perturbed sets of the model parameters). In a modeling study we investigate two techniques to model uncertainty in parameters describing SHTPs that are input to a process-based ecohydrology model. Using an existing and widely used database of soil properties comprising 1209 soil samples from 9 different soil textural classes [*Schaap and Leij*, 1998] we fit marginal distributions to parameters required by the model, and examine the rank correlation structure among the parameters conditioned on soil textural class. The fit marginal distributions and rank correlation matrices are used to generate random soil parameter samples of varying size using the two sampling techniques. The first technique considered is based on a Latin Hypercube sampling (LHS) scheme and imposes correlation known or believed a priori to exist among the parameters. LHS provides an algorithmic technique to ensure that low-probability parameter values that are potentially of high consequence to model behavior are sampled at small sample sizes. The importance of correlation control in representing uncertainty in soil parameters is illustrated conceptually in Figure 2. For a sandy loam soil, the joint space of values of the saturation and residual soil moisture (θ_S and θ_R, respectively) demonstrates (1) region of physically possible combinations as illustrated by the gray area, (2) the region bounded by the data from the database of *Schaap and Leij* [1998] illustrated by the black area, and (3) the measured values of θ_S and θ_R from the database of *Schaap and Leij* [1998] (Figure 2). The correlation control algorithm provides a means of ensuring that stochastically generated soil parameters used in ensemble-based modeling of soil moisture exhibit joint behavior that is similar to previously collected soils data. The second random sampling technique considered neither controls correlation among the stochastic samples of input parameters nor guarantees that low-probability parameter values will be sampled from their respective marginal distribution. To assess the performance of each sampling scheme we perform a series of numerical experiments designed to investigate the sensitivity of the ensemble statistics of near-surface soil moisture (mean and variance) to ensemble size. Because our motivating interest lies in improving our ability to economically estimate soil moisture at hillslope scales (e.g., 10 to 100 m) using ensemble-based data assimilation techniques, we are particularly interested in the degree to which the ensemble mean and variance in near surface soil moisture (which is related to geophysically observable variables) is consistent at small ensemble sizes.

[6] In section 2 we describe the theory and methodology used in this study, including a brief description of the sampling techniques and a description of the physics of the ecohydrology model and related input parameters representing SHTPs. Section 3 presents an overview of the statistical analysis of a database of parameters describing SHTPs, an outline of the modeling experiments, and the results of the numerical experiments for a series of ensemble

simulations at a point scale. Conclusions and implications of the results for ensemble-based soil moisture uncertainty assessment, data assimilation, and parameter estimation are discussed in section 4.

2. Theory and Methodology

[7] This section presents a detailed overview of the two sampling techniques used in this study. The ecohydrology model used is then briefly introduced, followed by a description of the parameters required by its soil moisture simulation component.

2.1. Model Parameter Sampling Techniques

2.1.1. Latin Hypercube–Based Sampling With Correlation

[8] Latin Hypercube sampling (LHS) is a useful tool to generate replicates of uncertain model inputs under constraints of limited computational resources. In the context of ensemble soil moisture data assimilation computational burden arises from (1) the large state-space dimensionality that typifies spatially distributed processes, (2) the complexity and nonlinearity of models that continuously resolve energy and water balance under intermittent and spatially varying hydrometeorological forcings, and (3) requirements of multiple model runs to construct the ensemble that characterizes state variable uncertainty. Its use in the representation of hydrologic model parameter uncertainty has been previously suggested and applied in the literature [e.g., *Beven and Freer*, 2001; *Yu et al.*, 2001; *Christiaens and Feyen*, 2002; *van Griensven et al.*, 2006; *Abbaspour et al.*, 2007]. LHS has also been used to efficiently model errors in satellite rainfall estimates [*Hossain et al.*, 2006]. For a given number of samples (ensemble replicates), m, the LHS approach stratifies the marginal cumulative density function (CDF) of a parameter into m strata with equal probability mass. In probability space, the boundaries of these strata are located at 0, $1/m$, $2/m$ …1. Each of the m replicates is generated by sampling uniformly once within each of the m strata, and inverting the marginal CDF to arrive at the random sample of the uncertain input variable. By stratifying the CDF into equally probable regions, the LHS approach ensures that the m random model inputs will include values that have a relatively low probability of occurrence, but are potentially of high consequence to the model outputs conditioned on their occurrence. For example, soils with very high moisture-holding capacities and correspondingly low saturated hydraulic conductivity may be encountered infrequently but may be of high consequence in terms of the dynamics of local soil moisture and runoff generation (e.g., easily saturated). The shape of the marginal CDF will in some circumstances impact the number of strata (or replicates) m that must be used to adequately capture the tails of the distribution. In particular, care must be exercised with those marginal distributions having long tails, or defined on semi-infinite and infinite domains.

[9] The processes encompassed within modern watershed ecohydrology models frequently require many parameters as input. Moreover, there may be reason to believe that parameters that represent physical attributes of the soil column or are directly measurable are physically related to other parameters (e.g., residual moisture content is always less than saturation or total porosity for a particular soil sample) and thus demonstrate varying degrees of correlation with one another. *Iman and Conover* [1982] proposed the so-called restricted pairing (RP) algorithm that imposes a target rank (Spearman) correlation matrix, **T**, on a matrix containing m samples of n different parameters. The n parameters may individually possess any combination of marginal distributions, and because rank correlation measures monotonic correlation (rather than linear correlation) the n-dimensional joint distribution among the parameters need not be known. Although specification of the joint distribution of the parameters is not necessary in the RP algorithm, the n marginal distributions together with the imposed rank correlation matrix determine the joint first- and second-order behavior between the parameters. Thus, care must be taken in selection of marginal distributions of the parameters.

[10] The initial step in the RP algorithm consists of constructing an m by n matrix (**V**), whose columns are identical and contain the standard normal inverse of the increasing van der Waerden scores,

$$\mathbf{V}=\begin{bmatrix} \Phi^{-1}(1/(m+1)) & \Phi^{-1}(1/(m+1)) & \cdots & \Phi^{-1}(1/(m+1)) \\ \Phi^{-1}(2/(m+1)) & \Phi^{-1}(2/(m+1)) & \cdots & \Phi^{-1}(2/(m+1)) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi^{-1}(m/(m+1)) & \Phi^{-1}(m/(m+1)) & \cdots & \Phi^{-1}(m/(m+1)) \end{bmatrix}, \tag{1}$$

where $\Phi^{-1}(\cdot)$ is the standard normal inverse function. Row entries of each column are then subjected to a random permutation, to arrive at the matrix **V***. The m van der Waerden scores in each column of the matrix **V** are effectively nonrandom CDF ordinates from each of the m Latin Hypercube strata in probability space. Transforming the entries of **V** to standard normal deviates and randomly permuting the rows within each column creates the m-by-n matrix **V*** of standard normal variables that span the standard normal distribution as widely as possible with m samples, which possesses no serial correlation within each column, and which possesses no meaningful correlation between the n columns. The remainder of the RP algorithm consists of a series of matrix-vector manipulations to reorganize the entries of the matrix **V*** to approximate the target rank correlation matrix **T**. The n-by-n rank correlation matrix, $\mathbf{R_{VV}}$, of matrix **V*** is computed by computing the rank correlation coefficient, $r_{x_j\ x_k}$, for each pairing combination of the n parameters as

$$r_{x_j x_k}=\frac{\sum_{i=1}^{m}\left(Rank\left(x_{ij}\right)-\overline{R}\left(x_j\right)\right)\left(Rank\left(x_{ik}\right)-\overline{R}\left(x_k\right)\right)}{\left[\sum_{i=1}^{m}\left(Rank\left(x_{ij}\right)-\overline{R}\left(x_j\right)\right)^2\right]^{1/2}\left[\sum_{i=1}^{m}\left(Rank\left(x_{ik}\right)-\overline{R}\left(x_j\right)\right)^2\right]^{1/2}}, \tag{2}$$

where $Rank(\cdot)$ is the rank transformation of the argument, and $\overline{R}\ (x_j) = \overline{R}(x_k) = (m + 1)/2$. The random permutation of the columns of matrix **V**, denoted **V***, should leave little correlation structure. The n-by-n lower triangular matrix **Q** is then obtained through a Cholesky factorization of the matrix $\mathbf{R_{VV}}$,

$$\mathbf{R_{VV}} = \mathbf{QQ}^T. \tag{3}$$

[11] Similarly, the n-by-n lower triangular matrix $\mathbf{P}$ is found by performing a Cholesky factorization on the target rank correlation matrix, $\mathbf{T}$, that we wish to impose on the uncertain model inputs,

$$\mathbf{T} = \mathbf{P}\mathbf{P}^T. \tag{4}$$

[12] The matrices $\mathbf{Q}$ and $\mathbf{P}$ are effectively the square root matrices of the correlation matrices $\mathbf{R_{VV}}$ and $\mathbf{T}$, respectively. Post multiplying $\mathbf{P}$ by $\mathbf{Q}^{-1}$ yields an n-by-n lower triangular matrix $\mathbf{S}$,

$$\mathbf{S} = \mathbf{P}\mathbf{Q}^{-1}. \tag{5}$$

$\mathbf{S}$ is a linear operator that can be used to remove any correlation structure that exists in $\mathbf{V}^*$, imposing instead the correlation structure of $\mathbf{T}$, by post multiplying $\mathbf{V}^*$ by $\mathbf{S}^T$,

$$\mathbf{S}^* = \mathbf{V}^*\mathbf{S}^T. \tag{6}$$

[13] The m-by-n matrix $\mathbf{S}^*$ has a rank correlation matrix that approximates $\mathbf{T}$. The elements of $\mathbf{S}^*$ are then input to the standard normal operator, $\Phi(\cdot)$, to retrieve the matrix $\mathbf{V}^{**}$ whose columns contain nonrecurring values ranging from 0 to 1, and which approximately has the desired rank correlation matrix. The elements of $\mathbf{V}^{**}$ are then inverted based on the appropriate marginal distribution for each of the n parameter columns to arrive at the m-by-n matrix $\mathbf{X}^*$. The matrix $\mathbf{X}^*$ thus contains m replicates of the n parameters that are sampled from a Latin Hypercube and exhibit a rank correlation structure that approximates the assumed or known rank correlation matrix $\mathbf{T}$.

2.1.2. Random Sampling

[14] The second technique for obtaining m replicates of n uncertain model inputs is significantly less complex than the RP technique and is referred here as simple random sampling (SRS). This method assumes that each of the n uncertain parameters have the same marginal distributions as for the RP case above. However, in this case no effort is made to impose any correlation among the n parameters, nor are any efforts made to ensure that extremes of the marginal distributions for each parameter are sampled. Stated differently, for a given sample size there is no guarantee that low-probability but potentially high-consequence parameter values are sampled or that any of the m n-dimensional parameter combinations are statistically implausible. Rather, in the SRS case the m replicates of the each of the n parameters are generated by repeatedly and independently drawing randomly from each of the respective marginal distributions. Using the SRS approach with large m, the m samples of each parameter would well characterize the assumed marginal distribution underlying that parameter. However, because the SRS approach does not impose any correlation among the m combinations of n parameters and to the degree that the target correlation matrix ($\mathbf{T}$) is not approximately equal to the n-by-n identity matrix, statistically implausible combinations of parameters are more likely to occur when using the SRS approach versus the RP approach outlined above.

2.2. The tRIBS-VEGGIE Model

[15] The ecohydrology model used to propagate uncertainty in soil hydraulic and thermal properties to near-surface soil moisture is the Triangulated Irregular Network–based Real-time Integrated Basin Simulator (tRIBS) and Vegetation Generator for Interactive Evolution model (VEGGIE) [*Ivanov et al.*, 2004a, 2004b, 2007, 2008a, 2008b]. tRIBS-VEGGIE is a spatially distributed model that resolves mass, energy, and carbon balance over a watershed at the hillslope scale by representation of coupled (1) biophysical energy processes (e.g., partitioning of input solar radiation in the canopy and soils), (2) biophysical hydrologic processes (partitioning of rainfall into interception, throughfall, plant water uptake, etc.), and (3) biochemical processes and vegetation phenology. The model takes as input precipitation and meteorological forcings, as well as the topographic and soil boundary conditions. A full treatment of the tRIBS-VEGGIE model is beyond the present scope of work and the reader is directed to the work of *Ivanov et al.* [2004a, 2004b, 2007, 2008a, 2008b]. The present study focuses on point-scale moisture dynamics, however, and therefore the parameters dealing with lateral moisture redistribution in the subsurface are not required or introduced. Moreover, this study is limited to assessing the influence of uncertainty in soil parameters on the ensemble behavior of soil moisture independently of vegetation effects. Thus, unvegetated (i.e., bare soil) conditions are assumed. We briefly describe the process mechanisms represented in tRIBS-VEGGIE that require parameters representing SHTPs: infiltration of precipitation in variably saturated soils, ground heat flux, and bare soil latent and sensible heat fluxes.

[16] Infiltration of water into the soil is modeled using a one-dimensional Richards equation for a sloped surface that allows for lateral gravitational drainage. The lower boundary condition of the model is a flux boundary condition, consistent with the assumption of significant depth to the saturated zone in the semiarid environment for which the model is currently most applicable. Moisture in the finite element soil column can vary between the input residual volumetric moisture content ($m^3\ m^{-3}$), θ_R, and the volumetric moisture content at saturation ($m^3\ m^{-3}$), θ_S. tRIBS-VEGGIE uses the Brooks-Corey model [*Brooks and Corey*, 1964] to characterize the relationship between volumetric moisture content (θ) hydraulic conductivity, $K(\theta)$, and soil matric potential, $\psi(\theta)$. The Brooks-Corey parameterization requires specification of a hydraulic conductivity at saturation (cm h^{-1}), K_S, the pore distribution index parameter (dimensionless), λ, and the air entry pressure (mm), ψ_b.

[17] Ground heat flux in the tRIBS-VEGGIE model is calculated through the method outlined by *Wang and Bras* [1999], which is based on a numerical solution to the one-dimensional heat diffusion equation with constant heat diffusivity. The solution to the heat diffusion equation proposed by *Wang and Bras* [1999] is based on the recent history of soil surface temperatures, and requires specification of the volumetric thermal conductivity and heat capacity of the soil. Both the thermal conductivity and heat capacity depend on the moisture state (θ) at the time of calculation, and therefore require specification of soil-specific thermal parameters as input. Computation of the soil heat capacity is moisture-dependent linear combination of the input heat capacity of the soil solid materials, $C_{s,solids}$ ($Jm^{-3}K^{-1}$), the moisture content at saturation of the soil (θ_S), the heat capacity of liquid water, and the moisture state in the near surface (θ). Moisture-dependent calculations of thermal conductivity in tRIBS-VEGGIE are based on the method

Table 1. Soil Hydraulic and Thermal Properties Required by tRIBS-VEGGIE

Symbol	Description
K_S	saturated hydraulic conductivity (mm h^{-1})
θ_R	residual moisture content (m^3 m^{-3})
θ_S	saturated moisture content (m^3 m^{-3})
λ	Brooks-Corey pore distribution index parameter
h_b	Brooks-Corey air entry pressure parameter (mm)
$k_{s,dry}$	volumetric thermal conductivity of dry soil (J m^{-1} s^{-1} K^{-1})
$k_{s,sat}$	volumetric thermal conductivity of saturated soil (J m^{-1} s^{-1} K^{-1})
$C_{s,solids}$	volumetric heat capacity of soil solids (J m^{-3} K^{-1})

Table 2. Parameters in the *Schaap and Leij* [1998] Database

Parameter	Description
% clay	percent clay by mass
% sand	percent sand by mass
% silt	percent silt by mass
ρ_b	bulk density, not used (g cm^{-3})
K_S	saturated hydraulic conductivity
θ_R	residual moisture content (mm^3 mm^{-3})
θ_S	saturated moisture content (mm^3 mm^{-3})
α	van Genuchten fitting parameter
n	van Genuchten fitting parameter

suggested by *Farouki* [1981] and require specification of the thermal conductivity of the dry soil ($Jm^{-1}s^{-1}K^{-1}$), $k_{s,dry}$, and the corresponding thermal conductivity of the saturated soil ($Jm^{-1}s^{-1}K^{-1}$), $k_{s,sat}$.

[18] Latent and sensible heat fluxes from the bare soil in tRIBS-VEGGIE are computed through a resistance formulation, in which independent resistances to latent and sensible heat fluxes are calculated. The gradient between air temperature and soil skin temperature drives sensible heat flux, while the gradient between atmospheric humidity and air humidity in the near-surface pore space drives latent heat flux. Humidity in the pore spaces near the soil surface, in turn, depends on the soil skin temperature. In this formulation, the latent heat flux depends on the soil matric potential and moisture state in the near surface, and on the input parameters θ_S and θ_R. It should also be noted, that sensible and latent heat fluxes also indirectly depend on the soil thermal properties outlined above, because each flux component depends on the soil skin temperature.

[19] The SHTPs required as input to the tRIBS-VEGGIE model and considered as uncertain in the present study are summarized in Table 1.

3. Soil Database Analysis, Modeling Experiment Setup, and Results

[20] Efforts to fit marginal distributions and compute rank correlation matrices from the soil hydrologic parameter database used in this study are summarized here. This is followed by an overview of the modeling experiments and analytical techniques employed to evaluate the impacts of the different soil parameter sampling techniques on soil moisture ensemble statistics, along with the results of these experiments.

3.1. Statistical Analysis of Soil Properties

[21] The soils data used in this study constitute a metadatabase from 3 soil surveys [*Rawls and Brakensiek*, 1985; *Ahuja et al.*, 1989; *Leij et al.*, 1996]. These data have previously been analyzed by *Schaap and Leij* [1998], and underlie the ROSETTA software issued by the U.S. Department of Agriculture's Salinity Laboratory. This metadatabase contains 2134 records corresponding to individual soil samples, 1309 of which possess a measurement of saturated hydraulic conductivity (K_S). The parameters measured for each record are summarized in Table 2. Note that the metadatabase used by *Schaap and Leij* [1998] present parameter values for the van Genuchten–Mualem [*van Genuchten*, 1980] soil water retention model, whereas the tRIBS-VEGGIE model requires Brooks-Corey parameters [*Brooks and Corey*, 1964]. *Morel-Seytoux et al.* [1996], however, demonstrated equivalence between the parameters of the van Genuchten and Brooks-Corey soil water retention functions and provided transformations between the parameterizations.

[22] Each record was assigned to a soil textural class based on the recorded sand, silt, and clay fraction. The number of records within each textural class ranges from 10 (silt) to 514 (sandy loam) and is shown in Table 3. Of the 1309 records with K_S data, 9 textural classes are represented by at least 20 records and are set apart for further statistical analysis. In our analysis of these 9 textural classes we assume that the within-class ranges of parameter values and correlation structure characterize the ensemble behavior of each textural class. In light of this assumption, within each of the 9 selected textural classes we fit marginal distributions to each parameter and computed the Spearman correlation matrices, as required to generate uncertain replicates of the SHTPs necessary to simulate soil moisture with the tRIBS-VEGGIE model.

[23] We assumed that the log-transformed hydraulic conductivity data is normally distributed based on previous studies of hydraulic conductivity distributions [e.g., *Reynolds and Elrick*, 1985]. Further, the residual moisture content (θ_R) data exhibited a significant number of records possessed θ_R equals zero. We treated the marginal distribution of θ_R as a mixed discrete-continuous distribution, with an atom of probability at 0 with mass equal to the empirical frequency of occurrence of $\theta_S = 0$ for each textural class, and a two-parameter beta distribution for nonzero values of θ_S. Initial candidate distributions for the remainder of the parameters were the gamma, two-parameter beta, and exponential distributions. The chosen distribution for each parameter was

Table 3. Number of Data Records by Classified Textural Class

	Number of Records in Database	Number of Records in Database with K_S Data
Clay	94	63
Sandy clay	10	8
Silty clay	29	14
Sandy clay loam	181	135
Silty clay loam	92	42
Clay loam	142	56
Sandy loam	514	334
Loam	252	119
Silt loam	327	135
Sand	342	277
Loamy sand	141	123
Silt	10	3
Total	2134	1309

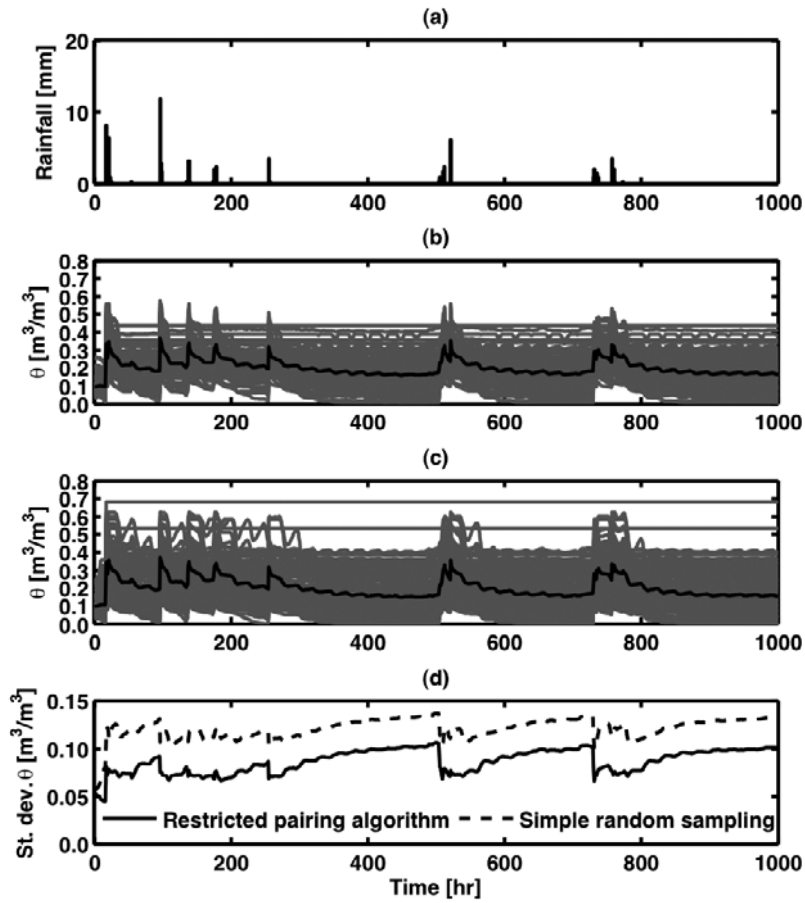

Figure 3. The time evolution of (a) rainfall used to drive the tRIBS-VEGGIE model, (b) soil moisture in the top 10 cm ($m^3\ m^{-3}$) for the simulations in which SHTPs were generated using the RP technique, (c) soil moisture in the top 10 cm ($m^3\ m^{-3}$) for the simulations in which SHTPs were generated using the random sampling technique, and (d) the standard deviation in soil moisture ($m^3\ m^{-3}$). In Figures 3b and 3c, gray lines depict individual ensemble replicates, while the black line depicts the ensemble mean. In Figure 3d the dashed line shows ensemble in which SHTPs were generated using random sampling, while the solid line indicates the ensemble in which they were generated using the RP technique.

based both on the significance of computed Kolmogorov-Smirnoff (KS) goodness-of-fit statistics and visible comparison between the fit marginal distributions and empirical histograms of each parameter. According to these criteria, the two-parameter beta distribution was chosen to represent the uncertainty in the remaining parameters. In the context of the soil properties considered here, the beta distribution is particularly advantageous because it is defined over a finite interval and can therefore constrain parameters to realistic values. The method of moments was used to estimate the beta distribution parameters a and b. The mean and variance for each soil parameter within each soil textural class, the estimates of the parameters describing the fit marginal distributions for each soil parameter and soil textural class together with plots of the fit marginal distributions along with the empirical histograms for each parameter and textural class, and the computed Spearman rank correlation matrices for each of the 9 considered textural classes are provided as auxiliary material.[1]

[1]Auxiliary materials are available in the HTML. doi:10.1029/2009WR008155.

3.2. Soil Moisture Ensembles: Experimental and Analytical Framework

[24] To contrast the two SHTP sampling techniques in the context of ensemble soil moisture modeling, results of a single ensemble experiment are presented in Figure 3. Using each sampling algorithm, 100 combinations of soil parameter inputs to tRIBS-VEGGIE were generated. The near-surface soil moisture (top 10 cm) response during a 1000 h period was then simulated, assuming initial soil moisture conditions corresponding to 10% effective saturation (defined as $S_e = [\theta - \theta_R]/[\theta_S - \theta_R] = 0.10$). Both ensembles were subjected to the same hydrometeorological forcings and the rainfall is depicted in Figure 3a. Figure 3b shows individual replicates and ensemble mean soil moisture response in the top 10 cm of soil for the case in which the ensemble of soil parameters was produced using the RP algorithm. Figure 3c depicts the results when the ensemble of soil parameters was produced using SRS. Figure 3d shows the time evolution of the ensemble standard deviation in soil moisture for both sampling approaches. The ensemble mean soil moisture response appears virtually identical for the two SHTP sampling algorithms (Figures 3b and 3c). However, several soil

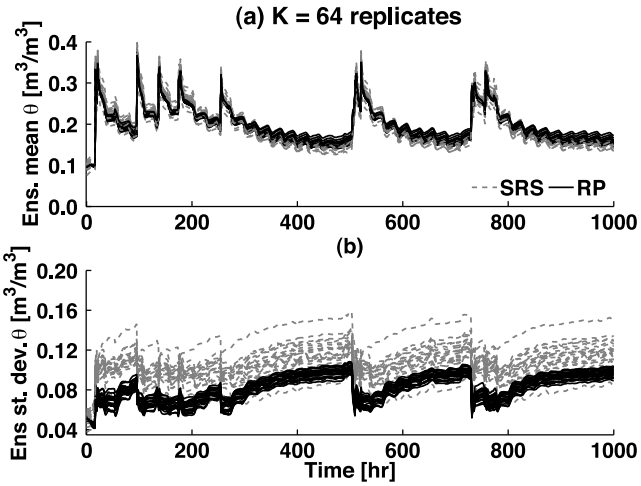

Figure 4. The time evolution of (a) soil moisture in the top 10 cm ($m^3\ m^{-3}$) and (b) the standard deviation in soil moisture ($m^3\ m^{-3}$). Gray dashed lines show ensembles in which SHTPs were generated using random sampling, while black solid lines indicate ensembles in which SHTPs were generated using the RP technique.

moisture replicates evolved with SRS-generated soil parameters appear to be physically unrealistic. Specifically, the soil moisture response of several replicates seems to saturate after the first rainfall event and remains saturated throughout the remainder of the simulation (Figure 3c). Furthermore, some replicates evolved with SRS-generated soils demonstrate large increases in near-surface soil moisture during interstorm periods, an implausibly large sensitivity to evaporative forcing (Figure 3c). The ensemble standard deviation in soil moisture, a measure of soil moisture uncertainty, generally responds similarly in time (Figure 3d). However, the ensemble standard deviation is higher for the experiment in which soil parameters were generated using the SRS approach (Figure 3d). Because a reasonably large number of soil parameter combinations were generated using each technique, the difference in the ensemble standard deviation in near-surface soil moisture is likely the result of parameter combinations that are statistically unlikely that, when subjected to meteorological forcings within the tRIBS-VEGGIE model, also lead to soil moisture dynamics that are unrealistic.

[25] The overarching objective of this work, however, is to investigate the degree to which the ensemble estimate of mean and variance in near-surface soil moisture vary depending on the technique by which soil property uncertainty is represented and the size of the ensemble. To this end, the present work requires producing sufficiently many ensemble first- and second-order statistics, across a range of ensemble sizes, to quantify estimator variances. We vary the ensemble size (K) in powers of 2, from 2^4 to 2^{10} (i.e., from 16 to 1024 replicates). To investigate potential difference in behavior associated with soil textural class variation we consider three distinct soil textures: loam, sandy loam, and clay. For each ensemble size (K) and soil textural class we generate 20 independent ensemble parameter combinations, each consisting of K combinations of the soil parameters required as input to tRIBS-VEGGIE, using both the RP and SRS technique. All replicates in these simulations are subjected to the same hydrometeorological forcings for a period of 1000 h, and the soil moisture state is not constrained to observations at any point during the simulation (i.e., soil moisture ensemble simulations are open loop). The rainfall time series used to drive the model is the same time series depicted in Figure 3a. Initial soil moisture conditions again correspond to 10% effective saturation.

[26] This set of simulations yields 20 time-evolving ensemble estimates of mean and variance in near-surface soil moisture for each ensemble size, K. Figure 4 depicts an example the time evolution of 20 estimates of ensemble mean (Figure 4a) and ensemble standard deviation (Figure 4b) soil moisture for one ensemble size (64 replicates). For this particular ensemble size, ensemble mean soil moisture is estimated fairly consistently using either RP or SRS to generate SHTPs input to the model (Figure 4a). However, for this ensemble size the estimate of ensemble standard deviation in soil moisture varies more when SRS is used to generate the soil parameters required as input to tRIBS-VEGGIE, than when the RP technique is used to simulate the soil parameters (Figure 4b). This reveals that, at this particular ensemble size, the estimate of ensemble soil moisture variance is sensitive to the particular combination of soil parameters sampled when SRS is used to generate the soil parameters required by the model.

[27] In the context of ensemble soil moisture data assimilation, achieving a consistent estimate of the ensemble statistics on which the state update step is based with the minimum number of ensemble replicates is a desirable goal from the perspective of computational efficiency and cost. In this case the number of ensemble replicates corresponds to the number of soil parameter combinations generated using each sampling technique and input to the tRIBS-VEGGIE model. A consistent estimate of the ensemble mean and variance, on which ensemble data assimilation algorithms such as the EnKF are based, is an estimate that is independent of the actual parameter values used. Specifically, the ensemble first- and second-order moments on which a hypothetical state update would be based should not vary significantly when different ensembles of soil parameters, generated in the same way, are used to characterize the soil. Significant variation in the soil moisture ensemble statistics across different ensemble simulations with different soil parameter ensembles drawn from the same marginal distributions could conceivably lead to significant differences in the innovations and analysis increments in a data assimilation state update step. Therefore, the appropriate measure of consistency in the first- and second-order ensemble statistics is the mean and variance (or standard deviation) in the ensemble estimates.

[28] For each of parameter sampling algorithms outlined above, and for each of the 20 independent ensembles and each ensemble size, $\hat{\mu}_{\theta i}(t)$ is the ensemble estimate of the mean soil moisture for ensemble i at time t, defined as

$$\hat{\mu}_{\theta i}(t) = \frac{1}{K}\sum_{k=1}^{K}\theta_k(t), \tag{7}$$

where $\theta_k(t)$ is the value of soil moisture of replicate k of K at time t. Similarly, $\hat{\sigma}^2_{\theta i}(t)$ is the ensemble estimate of the variance in soil moisture for ensemble i at time t, defined as

$$\hat{\sigma}^2_{\theta i}(t) = \frac{1}{K-1}\sum_{k=1}^{K}(\theta_k(t) - \hat{\mu}_\theta(t))^2. \tag{8}$$

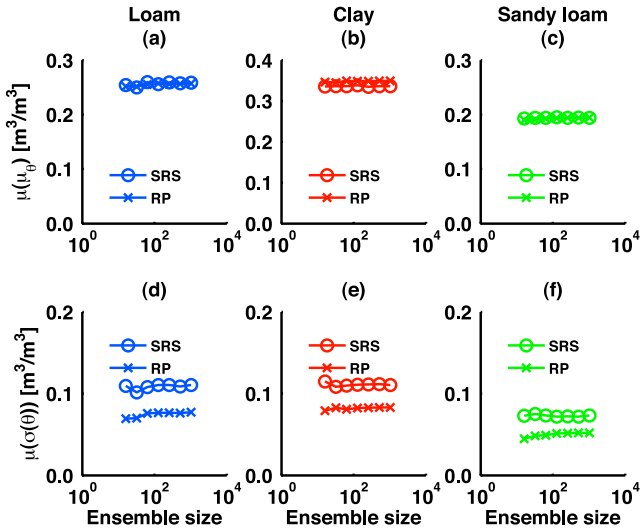

Figure 5. At 750 h (just after cessation of rainfall) the average ensemble mean soil moisture estimate across 20 ensembles as a function of ensemble size for (a) loam, (b) clay, and (c) sandy loam soils is shown. The average ensemble estimate of the standard deviation in soil moisture across 20 ensembles as a function of ensemble size for (d) loam, (e) clay, and (f) sandy loam soils.

The ensemble estimates of $\hat{\mu}_{\theta i}$ (t) and $\hat{\sigma}^2_{\theta i}$ (t) are computed for (1) each independent ensemble, (2) each ensemble size class, (3) each soil textural class considered, and (4) each parameter sampling technique. This yields 20 ensemble-based estimates of the mean and variance in near surface soil moisture for each soil class, ensemble size class, and sampling technique. For each soil class, ensemble size class, and sampling technique, we compute the average and variance (standard deviation) of the ensemble estimates of mean and variance in soil moisture across the 20 independent ensembles. The average ensemble mean and average ensemble variance in soil moisture, computed across the 20 independent ensembles and 7 ensemble size classes, provides a quantitative measure of the sensitivity of the ensemble estimates to ensemble size. Specifically of interest is (1) the degree to which the average estimate of the ensemble mean and average estimate of the ensemble variance changes as the ensemble size increases and (2) whether the average estimate of the ensemble mean and average estimate of the ensemble variance in soil moisture varies between the two parameter sampling schemes considered. For a particular ensemble size (K), the average ensemble mean and average ensemble variance in soil moisture across the N = 20 independent ensembles are computed as

$$\hat{\mu}(\hat{\mu}_\theta(t)) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}_{\theta i}(t)) \tag{9}$$

and

$$\hat{\mu}(\hat{\sigma}^2_\theta(t)) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}^2_{\theta i}(t)), \tag{10}$$

respectively.

[29] By contrast, the variance (standard deviation) of the ensemble mean and the variance in the ensemble variance are measures of the consistency of the ensemble estimates. These statistical metrics provide insight into the degree to which the variance in the ensemble estimates of the mean and variance in near surface soil moisture are sensitive to ensemble size, sampling technique, and soil texture. The sample variance in the ensemble mean soil moisture estimate is computed as

$$\hat{s}^2(\hat{\mu}_\theta(t)) = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\mu}_{\theta i}(t) - \hat{\mu}(\hat{\mu}_\theta(t)))^2, \tag{11}$$

while the sample variance in the ensemble estimate of variance in soil moisture is computed as

$$\hat{s}^2(\hat{\sigma}^2_\theta(t)) = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\sigma}^2_{\theta i}(t) - \hat{\mu}(\hat{\sigma}^2_\theta(t)))^2. \tag{12}$$

[30] Both s^2 $(\hat{\mu}_\theta$ (t)) and s^2 $(\hat{\sigma}_\theta$ (t)) should be zero in the limit of infinitely large ensembles (K), regardless of how the uncertainty in the parameters is represented. At small ensemble sizes, sampling error can cause both to be appreciably different from zero. Furthermore, as mentioned above the values of $\hat{\mu}_\theta$ (t) and $\hat{\sigma}^2_\theta$ (t) to which the ensemble simulations converge as K increases may be different between the two sampling techniques because the SRS approach makes no attempt to impose correlation among the parameters.

[31] For given hydrometeorological forcings, the rate at which s^2 $(\hat{\mu}_\theta$(t)) and s^2 $(\hat{\sigma}_\theta$(t)) decrease as ensemble size increases can highlight tradeoffs between computational burden due to increased ensemble size and the associated decrease in the variance of the ensemble estimates of mean and variance in soil moisture. Comparing the values of s^2 $(\hat{\mu}_\theta$(t)) and s^2 $(\hat{\sigma}_\theta$(t)) for the RP and SRS approaches to representing soil parameter uncertainty at a given ensemble size and during different times in the wetting-drying cycle can demonstrate the potential benefits of more careful treatment of parameter uncertainty under ensemble size constraints.

[32] Within each soil textural class considered, there is almost no discernable difference between the two parameter sampling techniques in the value of $\hat{\mu}$ $(\hat{\mu}_\theta$(t)) at hour 750, which immediately follows substantial rainfall event (Figures 5a–5c). Furthermore, for both sampling techniques and all three soil textural classes, $\hat{\mu}(\hat{\mu}_\theta$(t)) does not vary significantly as a function of ensemble size (Figures 5a–5c). The fact that there is very little variation in $\hat{\mu}(\hat{\mu}_\theta$(t)) between sampling techniques for all soil types likely reflects the influence on the volume of rainfall during the preceding precipitation event on the ensemble mean value of soil moisture. At the same time during the simulation (750 h), the two parameter sampling techniques produce substantially different values of $\hat{\mu}(\hat{\sigma}^2_\theta$(t)), although within each soil textural class the values of $\hat{\mu}(\hat{\sigma}^2_\theta$(t)) do not change significantly as the ensemble size increases for either parameter sampling technique (Figures 5d–5f). Furthermore, the RP technique is associated with lower values of $\hat{\mu}(\hat{\sigma}^2_\theta$(t)) within each soil textural class (Figures 5d–5f).

[33] For all soil textures, there are small differences seen in $\hat{\mu}(\hat{\mu}_\theta$(t)) as a function of ensemble size and parameter

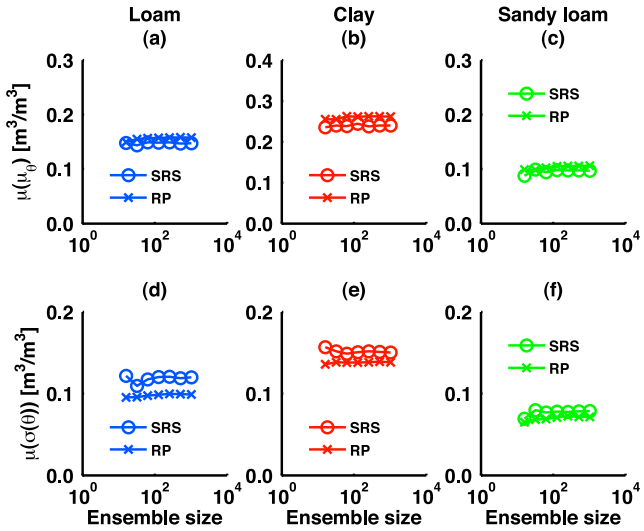

Figure 6. At 1000 h (during a significant dry down) the average ensemble mean soil moisture estimate across 20 ensembles as a function of ensemble size for (a) loam, (b) clay, and (c) sandy loam soils is shown. The average ensemble estimate of the standard deviation in soil moisture across 20 ensembles as a function of ensemble size for (d) loam, (e) clay, and (f) sandy loam soils.

sampling technique at hour 1000, which immediately follows an extended drying period (Figures 6a–6c). For all soil types, $\hat{\mu}(\hat{\mu}_\theta(\mathrm{t}))$ slightly increases with ensemble size (Figures 6a–6c). Furthermore, at any given ensemble size and for all ensemble sizes $\hat{\mu}(\hat{\mu}_\theta(\mathrm{t}))$ tends to be slightly higher for the soil moisture ensembles simulated with the RP-generated soil parameters (Figures 6a–6c). At the same time during the simulation (1000 h), for each soil type $\hat{\mu}(\hat{\sigma}^2_\theta(\mathrm{t}))$ does not substantially change as the ensemble size increases, except at the smallest ensemble sizes of 16 and 32 (Figures 6d–6f). Moreover, for loamy and clay soil types the two parameter sampling techniques seem to produce values of $\hat{\mu}(\hat{\sigma}^2_\theta(\mathrm{t}))$ that differ slightly, but noticeably (Figures 6d and 6e). However, for sandy loam soils there is little difference in $\hat{\mu}(\hat{\sigma}^2_\theta(\mathrm{t}))$ resulting from the different parameter sampling techniques.

[34] Figure 7 shows $s^2(\hat{\mu}_\theta(\mathrm{t}))$ and $s^2(\hat{\sigma}_\theta(\mathrm{t}))$ at $t = 750$ h into the simulation during a significant rain event. For all soil textures, using RP to generate soil parameters input to tRIBS-VEGGIE yields a lower value of $s^2(\hat{\mu}_\theta(\mathrm{t}))$ at all ensemble sizes (Figures 7a–7c). When comparing the techniques, the difference in $s^2(\hat{\mu}_\theta(\mathrm{t}))$ is greatest at the smallest ensemble size, and is relatively insignificant at $K = 1028$ (Figures 7d–7f). Although $s^2(\hat{\mu}_\theta(\mathrm{t}))$ is relatively low at the minimum ensemble size (16) when RP is used to sample SHTPs, it decreases only modestly as ensemble size increases. Similarly, using RP to generate soil parameters input to tRIBS-VEGGIE yields a lower value of $s^2(\hat{\sigma}_\theta(\mathrm{t}))$ at all ensemble sizes for all soil textures. When comparing the two techniques, the difference in the value of $s^2(\hat{\sigma}_\theta(\mathrm{t}))$ at a given ensemble size is largest at small ensemble sizes. However, this difference decreases as K increases and is negligibly small at ensemble sizes of $K = 512$ for clay and sandy loam soils, and $K = 256$ for loam soils. When using SRS to generate soil parameters, doubling or quadrupling the ensemble size from the minimum 16 yields much more consistency in the estimate of ensemble mean and variance in soil moisture. Figure 8 depicts similar results during a significant dry spell in the rainfall record ($t = 1000$ h). Conclusions are largely the same; however, note that at $K = 16$ for clay soils using SRS-generated soil parameters actually results in a lower value of $s^2(\hat{\sigma}_\theta(\mathrm{t}))$ when compared to using RP-generated parameters (Figures 8a–8c). It is possible this is due to sampling error associated with generating a relatively small number (20) of independent ensembles to compute $s^2(\hat{\mu}_\theta(\mathrm{t}))$ and $s^2(\hat{\sigma}_\theta(\mathrm{t}))$ (Figure 8). These results highlight the notion that careful representation of uncertainty in model parameters describing SHTPs required by a hydrologic model can lead to reduced estimator (mean and variance) variance at small ensemble sizes.

4. Discussion and Conclusions

[35] The results of this work indicate that when computational resources serve as a constraint on the size of the soil moisture ensemble, using a sampling technique that (1) samples low-probability but potentially high-consequence combinations of soil parameters and (2) imposes correlation known or believed to exist among those parameters can potentially result in more consistent estimation of ensemble mean and variance in soil moisture, as measured by the variance in the ensemble statistic estimates across 20 independent ensembles. This is a potentially important conclusion in the context of hydrologic data assimilation, because it demonstrates that potentially significant reductions in the computational cost associated with the Monte Carlo integration of hydrologic models during the forecast step can be realized through careful attention to the way in which model parameters are sampled. This work is particularly targeted toward future ensemble data assimilation studies using complex and spatially distributed ecohydrologic models, in which the Monte Carlo simulation of hydrologic moisture and energy states across the landscape will constitute the bulk of the computational expense, rather than the matrix computations associated with the state update. As such, this work represents a relatively novel, but potentially powerful, way of improving the computational economics of high-dimensional soil moisture data assimilation systems.

[36] The finding that the parameter sampling techniques considered here were associated with substantial differences in the average value of the ensemble standard deviation in soil moisture during a rainfall event is a potentially important finding from the perspective of soil moisture data assimilation. For example, because assimilation algorithms such as the EnKF rely on a variance-covariance matrix to resolve the innovations in the model space, the way in which parameter uncertainty is represented may play an important, if often unrecognized, role in the model state update. Additionally, a critical component of current and planned microwave soil moisture remote sensing satellites is to allow for independent quantification of errors in precipitation data used to force land surface models [e.g., see *Crow*, 2007]. Employing different techniques to sample the soil parameters required as input to land surface models could lead to systemic differences in the ensemble statistics used to update those models with available observations, and systemic differences in the state update itself. When the magnitude of near-surface soil

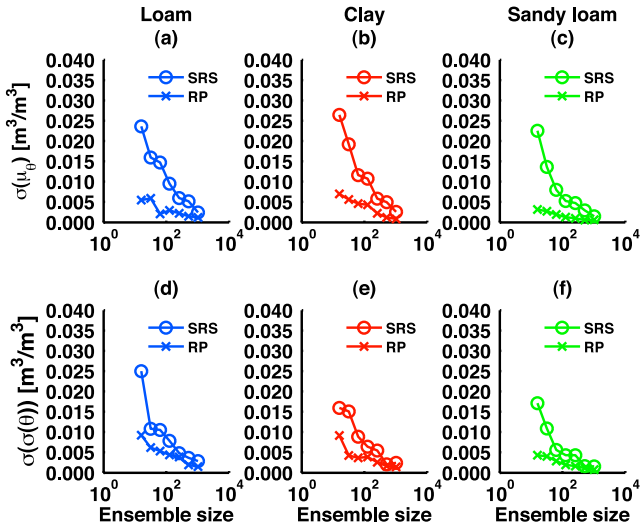

Figure 7. At 750 h (just after cessation of rainfall) the standard deviation in the ensemble mean soil moisture estimate across 20 ensembles as a function of ensemble size for (a) loam, (b) clay, and (c) sandy loam soils is shown. The standard deviation in the ensemble estimate of standard deviation in soil moisture across 20 ensembles as a function of ensemble size for (d) loam, (e) clay, and (f) sandy loam soils.

moisture updates form a basis upon which to quantify errors in the precipitation data used to force the model, the technique employed to represent uncertainty in the soil parameters required by the model may influence the interpretation of those precipitation errors. This work has shown that while the size of the ensemble plays a critical role in the estimator variance of the ensemble mean and variance in near-surface soil moisture (Figures 7 and 8), correlation known or believed to exist among the soil parameters dictates the values of the ensemble mean and variance in soil moisture, particularly during rainfall events (Figure 5). It is important to note that the structure of the rank correlation matrix **T** in the RP approach, which was calculated empirically from a soils database in this study, influences the consistency of the ensemble statistics. However, the magnitude of the empirically calculated rank correlation coefficient is sensitive to the number of soil measurement records within each soil textural class. It is possible, therefore, that ensembles of soil parameters generated through the RP approach and the associated ensemble soil moisture statistics derived from the model are influenced by spurious soil property correlations when the number of soil records used to compute **T** for each textural class is small. This may lead to consistent estimates of ensemble statistics that are nevertheless unrealistic because the correlation among the model parameters has either been overestimated or underestimated. This consequence indicates that some caution is warranted in the use of the RP approach, particularly in cases where the rank correlation matrix **T** must be calculated with a small number records of the soil or other model parameters.

[37] Depending on the way in which the physical mechanisms responsible for lateral redistribution of moisture in the subsurface are represented within a particular model, however, explicit representation of uncertainty soil parameters may have implications for the spatial patterns of ensemble soil moisture statistics. For instance, the tRIBS-VEGGIE model formulates the slope-parallel hydraulic conductivity is assumed to be proportional to the (uncertain) slope-normal hydraulic conductivity. The constant of proportionality is termed the anisotropy ratio, a_r, and can either amplify ($a_r > 1$) or diminish ($a_r < 1$) uncertainty in the slope-normal hydraulic conductivity. At any given time, uncertainty in the slope-normal hydraulic conductivity arises from imperfect knowledge of the Brooks-Corey parameters as well as uncertainty in moisture status.

[38] It is important to note that an argument can be made that the low ensemble soil moisture variance depicted in Figure 1b is physically consistent with the expectation of very little near-surface moisture toward the end of a drying cycle, irrespective of the soil. However, while it is true that the total amount of moisture in the near surface would be small at the end of a drying cycle, the residual moisture content parameter is meant to capture the fact that most natural soils do retain some small amount of moisture in pore spaces not connected to the continuous pore network of the soil, such as intra-aggregate spaces, even under intense natural drying conditions [*Hillel*, 1998]. Moreover, because the hydraulic conductivity of the soil depends on the soil moisture state, the potentially most important role of the residual soil moisture parameter in hydrologic simulation is in constraining the infiltration rate at the beginning of a storm.

[39] This work is also of potential importance in the calibration of spatially distributed hydrologic models, a topic that has received a significant amount of attention in hydrologic science literature. Early data assimilation studies in the hydrologic sciences employed Kalman filtering procedures to simultaneously estimate the state and parameters

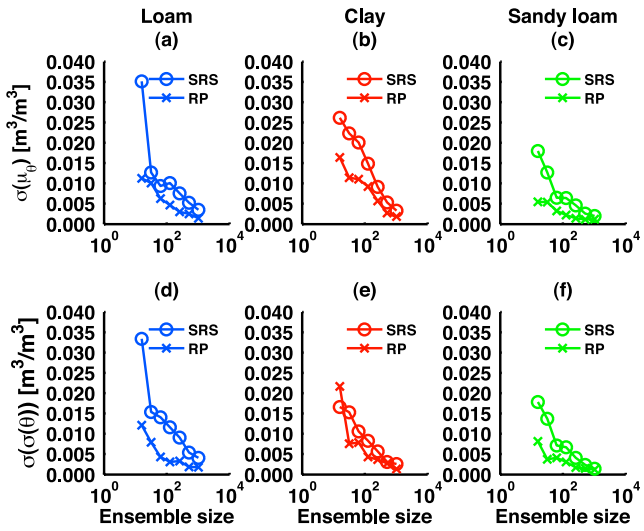

Figure 8. At 1000 h (during a significant dry down) the standard deviation in the ensemble mean soil moisture estimate across 20 ensembles as a function of ensemble size for (a) loam, (b) clay, and (c) sandy loam soils is shown. The standard deviation in the ensemble estimate of standard deviation in soil moisture across 20 ensembles as a function of ensemble size for (d) loam, (e) clay, and (f) sandy loam soils.

of the Sacramento model by assimilating discharge observations. This automatic calibration approach to parameter estimation has been used to identify a parameter and state estimate that is best in a linear least squares sense as well as a measure of the uncertainty in that estimate. A substantial body of work has been devoted to global optimization approaches for calibration of complex hydrological models [see, e.g., *Duan et al.*, 1992; *Gupta et al.*, 1998; *Yapo et al.*, 1998; *Boyle et al.*, 2000]. These global optimization approaches, which minimize an often multiobjective cost function that penalizes deviations between predicted observations and observational data to arrive at some Pareto-optimal parameter estimate, may be advantageous when using complex and high-dimensional hydrologic models because augmented state vector approaches may be substantially more computationally expensive. *Vrugt et al.* [2005] exploited the strengths of both data assimilation and global optimization strategies to estimate hydrologic model states and parameters in an ensemble-based framework that they term simultaneous optimization and data assimilation (SODA). The SODA framework, however, is computationally expensive because it requires serial ensemble data assimilation experiments interspersed with the use of a shuffled complex evolution-based global optimization scheme. The RP soil parameter generation technique outlined in this work may stand to substantially reduce computational costs associated with parameter estimation through the SODA framework by reducing the size of the ensemble simulation between successive iterations of the optimization scheme. More recently, Markov Chain Monte Carlo (MCMC) techniques have received a great deal of attention in the calibration of hydrologic models and the estimation of their parameters [*Vrugt et al.*, 2008]. MCMC is an optimization technique designed to minimize errors between model predictions and observations in order to estimate the parameters of a model and their posterior distributions. The RP approach outlined here could be used to generate ensembles of soil parameters for sequential state estimation with algorithms such as the EnKF based on the estimated parameter posterior distributions obtained through the MCMC technique.

[40] As pointed out by a reviewer, the approach taken in this study assumes that the spatial distribution of soil textures is known. This is generally not the case, particularly in locations outside the Continental United States. It is possible to extend the techniques outlined here to include for the possibility of uncertainty in the spatial distribution of soil types. For example, through geostatistical analysis of topography and surface lithology one could construct a map delineating distinct lithotopo units and a corresponding probability mass function that expresses the likelihood that a given lithotopo unit is characterized by a particular soil texture. This probability mass function could be used to generate ensembles of maps delineating the possible spatial distribution of categorical soil texture classes. Then, for each potential soil texture map, the RP approach could be used to generate an ensemble of soil parameters required as input to the land surface model. This ensemble of soil moisture ensembles could then be constrained to satellite microwave observations in a data assimilation system, which would conceivably lead to improved understanding about the spatial distribution of soil textures. Such an approach may lead to a process-based mechanism for mapping soil texture globally.

[41] **Acknowledgments.** The authors wish to thank Wade Crow and two anonymous reviewers for their instructive comments. Additionally, we would like to express our gratitude to Marcel Schaap at the University of Arizona for providing the soil data used in this study. This work was made possible through the support of Army Research Office grant W911NF-04-1-0119, NASA grant NNG05GA17G, and an MIT Martin Family Society of Fellows for Sustainability fellowship.

References

Abbaspour, K. C., J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan (2007), Modeling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT, *J. Hydrol.*, *333*(2–4), 413–430, doi:10.1016/j.jhydrol.2006.09.014.

Ahuja, L. R., D. K. Cassel, R. R. Bruce, and B. B. Barnes (1989), Evaluation of spatial distribution of hydraulic conductivity using effective porosity data, *Soil Sci.*, *148*, 404–411, doi:10.1097/00010694-198912000-00002.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*(1–4), 11–29, doi:10.1016/S0022-1694(01)00421-8.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*, 3663–3674, doi:10.1029/2000WR900207.

Brooks, R. H., and A. T. Corey (1964), *Hydraulic properties of porous media*, *Hydrol. Pap. 3*, Colo. State Univ., Fort Collins.

Carpenter, T. M., K. P. Georgakakos, and J. A. Spersflage (2001), On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use, *J. Hydrol.*, *253*(1–4), 169–193, doi:10.1016/S0022-1694(01)00476-0.

Christiaens, K., and J. Feyen (2002), Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model, *Water Resour. Res.*, *38*(9), 1169, doi:10.1029/2001WR000478.

Crow, W. T. (2007), A novel method for quantifying value in spaceborne soil moisture retrievals, *J. Hydrometeorol.*, *8*(1), 56–66, doi:10.1175/JHM553.1.

Crow, W. T., and E. van Loon (2006), The impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, *J. Hydrometeorol.*, *7*(3), 421–432, doi:10.1175/JHM499.1.

Crow, W. T., and E. F. Wood (2003), The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, *Adv. Water Resour.*, *26*(2), 137–149, doi:10.1016/S0309-1708(02)00088-X.

Downer, C. W., F. L. Ogden, W. D. Martin, and R. S. Harmon (2002), Theory, development, and applicability of the surface water hydrologic model CASC2D, *Hydrol. Processes*, *16*, 255–275, doi:10.1002/hyp.338.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*, 1015–1031, doi:10.1029/91WR02985.

Dunne, S., and D. Entekhabi (2005), An ensemble-based reanalysis approach to land data assimilation, *Water Resour. Res.*, *41*, W02013, doi:10.1029/2004WR003449.

Dunne, S., and D. Entekhabi (2006), Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, *42*, W01407, doi:10.1029/2005WR004334.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162, doi:10.1029/94JC00572.

Evensen, G. (2004), Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dyn.*, *54*, 539–560, doi:10.1007/s10236-004-0099-2.

Farouki, O. T. (1981), Thermal properties of soils, *Rep. 81-011*, Cold Reg. Res. and Eng. Lab., U.S. Army Corps of Eng., Hanover, N. H.

Galantowicz, J. F., D. Entekhabi, and E. G. Njoku (1999), Tests of sequential data assimilation retrieving profile soil moisture and temperature from observed L-band radiobrightness, *IEEE Trans. Geosci. Remote Sens.*, *37*, 1860–1870, doi:10.1109/36.774699.

Gelb, A. (1974), *Applied Optimal Estimation*, 374 pp., MIT Press, Cambridge, Mass.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763, doi:10.1029/97WR03495.

Hillel, D. (1998), *Environmental Soil Physics*, Academic, Amsterdam.

Hoeben, R., and P. A. Troch (2000), Assimilation of active microwave observation data for soil moisture profile estimation, *Water Resour. Res.*, *36*, 2805–2819, doi:10.1029/2000WR900100.

Hossain, F., E. N. Anagnostou, and A. C. Bagtzoglou (2006), On Latin Hypercube sampling for efficient uncertainty estimation of satellite rainfall observations in flood prediction, *Comput. Geosci.*, *32*(6), 776–792, doi:10.1016/j.cageo.2005.10.006.

Iman, R. L., and W. J. Conover (1982), A distribution-free approach to inducing rank correlation among input variables, *Commun. Stat. Simul. Comput.*, *11*(3), 311–334, doi:10.1080/03610918208812265.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, and D. Entekhabi (2004a), Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: A fully distributed physically based approach, *J. Hydrol.*, *298*(1–4), 80–111, doi:10.1016/j.jhydrol.2004.03.041.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, and D. Entekhabi (2004b), Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resour. Res.*, *40*, W11102, doi:10.1029/2004WR003218.

Ivanov, V. Y., R. L. Bras, and D. C. Curtis (2007), A weather generator for hydrological, ecological, and agricultural applications, *Water Resour. Res.*, *43*, W10406, doi:10.1029/2006WR005364.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008a), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks, *Water Resour. Res.*, *44*, W03429, doi:10.1029/2006WR005588.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008b), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 2. Energy-water controls of vegetation spatiotemporal dynamics and topographic niches of favorability, *Water Resour. Res.*, *44*, W03430, doi:10.1029/2006WR005595.

Leij, F. J., W. J. Alves, M. T. van Genuchten, and J. R. Williams (1996), Unsaturated soil hydraulic database: UNSODA 1.0 user's manual, *Rep. EPA/600/R-96/095*, 103 pp., U.S. Environ. Prot. Agency, Ada, Okla.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, *38*(12), 1299, doi:10.1029/2001WR001114.

Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005a), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*(2), 135–147, doi:10.1016/j.advwatres.2004.09.002.

Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005b), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, *41*, W05012, doi:10.1029/2004WR003604.

Morel-Seytoux, H. J., P. D. Meyer, M. Nachabe, J. Tourna, M. T. van Genuchten, and R. J. Lenhard (1996), Parameter equivalence for the Brooks-Corey and van Genuchten soil characteristics: Preserving the effective capillary drive, *Water Resour. Res.*, *32*(5), 1251–1258, doi:10.1029/96WR00069.

Qu, Y., and C. J. Duffy (2007), A semidiscrete finite volume formulation for multiprocess watershed simulation, *Water Resour. Res.*, *43*, W08419, doi:10.1029/2006WR005752.

Rawls, W. J., and D. L. Brakensiek (1985), Prediction of soil water properties for hydrologic modeling, in *Watershed Management in the Eighties*, edited by E. B. Jones, and T. J. Ward, pp. 293–299, Am. Soc. of Civ. Eng., Denver, Colo.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*, 103–114, doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, *44*, W03423, doi:10.1029/2007WR006357.

Reynolds, W. D., and D. E. Elrick (1985), In situ measurement of field-saturated hydraulic conductivity, sorptivity, and the α-parameter using the Guelph permeameter, *Soil Sci.*, *140*, 292–302, doi:10.1097/00010694-198510000-00008.

Schaap, M. G., and F. J. Leij (1998), Database-related accuracy and uncertainty of pedotransfer functions, *Soil Sci.*, *163*, 765–779, doi:10.1097/00010694-199810000-00001.

van Genuchten, M. T. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, *44*(5), 892–898.

van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*(1–4), 10–23, doi:10.1016/j.jhydrol.2005.09.008.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, W00B09, doi:10.1029/2007WR006720.

Wang, J., and R. L. Bras (1999), Ground heat flux estimated from surface soil temperature, *J. Hydrol.*, *216*(3–4), 214–226, doi:10.1016/S0022-1694(99)00008-6.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol.*, *204*(1–4), 83–97, doi:10.1016/S0022-1694(97)00107-8.

Yu, P.-S., T.-C. Yang, and S.-J. Chen (2001), Comparison of uncertainty analysis methods for a distributed rainfall-runoff model, *J. Hydrol.*, *244*(1–2), 43–59, doi:10.1016/S0022-1694(01)00328-6.

R. L. Bras, Henry Samueli School of Engineering, University of California, 305 Rockwell Engineering Center, Irvine, CA 92697, USA.

D. Entekhabi, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA.

A. N. Flores, Department of Geosciences, Boise State University, 1910 University Dr., Boise, ID 83725-1535, USA. (lejoflores@boisestate.edu)