Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment

Yuning Shi,[1] Kenneth J. Davis,[1] Fuqing Zhang,[1] Christopher J. Duffy,[2] and Xuan Yu[2]

Received 3 May 2013; revised 1 December 2013; accepted 17 December 2013; published 29 January 2014.

[1] This paper presents multiple parameter estimation using multivariate observations via the ensemble Kalman filter (EnKF) for a physically based land surface hydrologic model. A data assimilation system is developed for a coupled physically based land surface hydrologic model (Flux-PIHM) by incorporating EnKF for model parameter and state estimation. Synthetic data experiments are performed at a first-order watershed, the Shale Hills watershed (0.08 km^2). Six model parameters are estimated. Observations of discharge, water table depth, soil moisture, land surface temperature, sensible and latent heat fluxes, and transpiration are assimilated into the system. The results show that, given a limited number of site-specific observations, the EnKF can be used to estimate Flux-PIHM model parameters. All the estimated parameter values are very close to their true values, with the true values inside the estimated uncertainty range (1 standard deviation spread). The estimated parameter values are not affected by the initial guesses. It is found that discharge, soil moisture, and land surface temperature (or sensible and latent heat fluxes) are the most critical observations for the estimation of those six model parameters. The assimilation of multivariate observations applies strong constraints to parameter estimation, and provides unique parameter solutions. Model results reveal strong interaction between the van Genuchten parameters α and β, and between land surface and subsurface parameters. The EnKF data assimilation system provides a new approach for physically based hydrologic model calibration using multivariate observations. It can be used to provide guidance for observational system designs, and is promising for real-time probabilistic flood and drought forecasting.

Citation: Shi, Y., K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu (2014), Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment, *Water Resour. Res.*, *50*, 706–724, doi:10.1002/2013WR014070.

1. Introduction

[2] Land surface models (LSMs) and hydrologic models are important tools for the forecasting and study of land surface and hydrologic processes. LSMs simulate the exchange of mass, momentum and energy between the land surface and the atmosphere. They play important roles in weather and climate forecasting, and provide necessary lower boundary conditions for atmospheric models. Hydrologic models simulate hydrologic system responses to incoming precipitation. They are essential for predicting flood and drought events and are routinely used for decisions that have considerable societal impacts. Both LSMs and hydrologic models are highly parameterized. Model structures are complex and the number of involved parameters is often large. The accuracy of LSMs and hydrologic models is limited by, among other factors, the uncertainties in model parameters. Parameter estimation of LSMs and hydrologic models has been the focus of many studies [e.g., *Gupta et al.*, 1999; *Xia et al.*, 2002; *Jackson et al.*, 2003].

[3] Uncertainties in model parameters are an especially dominant source of uncertainty for hydrologic models [*Moradkhani and Sorooshian*, 2008]. Hydrologic model parameters nearly always require calibration for specific watersheds before they can produce realistic responses to environmental inputs such as precipitation. For hydrologic models, the physical parameter values in actual field conditions might be substantially different from those measured in laboratory; the range of variation in parameter values spans orders of magnitude [*Bras*, 1990]. Some physical parameters may be heterogeneous in space, which complicates the task of obtaining accurate parameter estimates. Consequently, model calibration is one of the most demanding and time-consuming tasks in applying hydrologic models, and the resulting parameter values often have considerable uncertainty even after optimization.

[4] In the past few decades, many hydrologic model calibration methods have been proposed and studied. A basic calibration approach is the trial and error method, i.e.,

[1]Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.
[2]Department of Civil Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.

Corresponding author: Y. Shi, Department of Meteorology, Pennsylvania State University, 415 Walker Bldg., University Park, PA 16802, USA. (yshi@psu.edu)

0043-1397/14/10.1002/2013WR014070

manual calibration. In manual calibration, model performances are visually inspected, and then model parameter values are tuned to minimize the differences between model predictions and observations, based on human judgment [*Boyle et al.*, 2000; *Moradkhani and Sorooshian*, 2008]. This method is very labor intensive and time consuming. Manual calibration of physically based hydrologic models can be extremely difficult due to the high dimensionality of the parameter space and the strong interaction between model parameters. It is also important to note that manual calibration does not lead to a rigorous (if any) quantification of parameter uncertainty. Those difficulties motivated the development of automatic calibration methods.

[5] Generally, there are two strategies for automatic calibration: batch (iterative) calibration and sequential (recursive) calibration. Batch calibration aims to minimize the predefined objective functions by repeatedly searching in the parameter space and evaluating long period model performances [e.g., *Ibbitt*, 1970; *Johnston and Pilgrim*, 1976; *Pickup*, 1977; *Gupta and Sorooshian*, 1985; *Duan et al.*, 1992; *Sorooshian et al.*, 1993; *Franchini*, 1996; *Wagener et al.*, 2003; *Kollat and Reed*, 2006; *Yu et al.*, 2013]. Batch calibration requires previously collected historical data for model evaluation and is thus restricted to offline applications. Batch calibration has limited flexibility in dealing with the possible temporal evolution of model parameters [*Moradkhani et al.*, 2005a; *Moradkhani and Sorooshian*, 2008].

[6] Sequential calibration (data assimilation) methods can take advantage of measurements whenever they are available and are thus useful in both online and offline applications. They have more flexibility in dealing with time-variant parameters, compared with batch calibration methods. Some sequential calibration methods also explicitly address uncertainties in input data and model structures. Frequently used data assimilation methods for LSMs and hydrologic models include variational methods [e.g., *Reichle et al.*, 2002a; *Lee et al.*, 2011], the particle filter [e.g., *Moradkhani et al.*, 2005b; *Weerts and El Serafy*, 2006; *Salamon and Feyen*, 2009], and different forms of Kalman filter, especially the ensemble Kalman filter [EnKF; e.g., *Reichle et al.*, 2002a, 2002b; *Crow and Wood*, 2003; *Francois et al.*, 2003; *Moradkhani et al.*, 2005a; *Pan and Wood*, 2006; *Vrugt et al.*, 2006; *Clark et al.*, 2008; *Kumar et al.*, 2008; *Camporese et al.*, 2009a; *Xie and Zhang*, 2010; *Cammalleri and Ciraolo*, 2012; *Han et al.*, 2012; *Flores et al.*, 2012; *Hain et al.*, 2012]. Variational methods depend on the development of adjoint models. The application of variational methods to LSMs and hydrologic models is thus difficult, because adjoints of LSMs and hydrologic models are not always available and are difficult to derive [*Reichle et al.*, 2002a, 2002b; *Moran et al.*, 2004; *Vrugt et al.*, 2006; *Salamon and Feyen*, 2009]. The particle filter [*Arulampalam et al.*, 2002] has no assumptions on the form of the prior probability density function (PDF) of the model states and the model errors. It can maintain the predicted spatial pattern of distributed variables, because the particle filter updates the weights of different ensemble members, instead of directly updating the state variables. However, the particle filter requires many more ensemble members than EnKF to produce good estimates of model errors, and is thus more computationally expensive [*Weerts and El Serafy*, 2006; *Clark et al.*, 2008; *Salamon and Feyen*, 2009]. EnKF [*Evensen*, 1994] has been widely used for parameter estimation in recent years. EnKF has a simple conceptual formulation, relative ease of implementation (no adjoint needed), and affordable computational requirements [*Evensen*, 2003]. EnKF is not only useful in improving state and parameter estimations, but can also provide uncertainty estimations of variables and parameters. Compared with other forms of Kalman filters, EnKF is capable of handling strongly nonlinear dynamics, high dimensional state vectors, and to some degree non-Gaussian parameter and state probability distributions.

[7] Because of the high computational demands of process-based and physically-based hydrologic models, it is very difficult to use batch methods for calibration [*Tang et al.*, 2006]. Their high dimensional parameter space and high nonlinearity pose difficulties for sequential methods as well. Currently manual calibration, i.e., trial and error procedure, is still the prevalent choice for physically based hydrologic model calibration [e.g., *Pisinaras et al.*, 2010; *Leimer et al.*, 2011; *Shi et al.*, 2011; *Shih and Yeh*, 2011; *Dechmi et al.*, 2012; *Yao et al.*, 2012; *Shi et al.*, 2013a]. The EnKF provides a promising approach for distributed physically based hydrologic model auto calibration. *Moradkhani et al.* [2005a] applied EnKF to a lumped conceptual rainfall runoff (R-R) model to estimate the values of five model parameters using real observations, and found that the obtained parameter set from EnKF was similar to the results from batch calibration. The ensemble discharge prediction also agreed well with observations. *Xie and Zhang* [2010] applied EnKF to a spatially distributed conceptual hydrologic model to estimate a spatially distributed model parameter, which had different values in different hydrologic response units (HRUs). In the synthetic data experiments, at most HRUs, the estimated values of the parameter were very close to the true values when discharge observations were assimilated. *Lü et al.* [2013] applied EnKF to a lumped conceptual R-R model and found that using dual state-parameter estimation improves model streamflow estimation compared to the test case which only used state estimation. There are also studies implementing EnKF in groundwater models to estimate model parameters such as hydraulic conductivities [e.g., *Chen and Zhang*, 2006; *Liu et al.*, 2008; *Hendricks Franssen et al.*, 2011; *Kurtz et al.*, 2012]. Although EnKF has been proven effective for lumped and distributed conceptual watershed models and some physically based groundwater models, the effectiveness of EnKF in parameter estimation for spatially distributed physically based watershed models, or land surface hydrologic models is still untested.

[8] Data assimilation for fully coupled physically based hydrologic models using EnKF is difficult. Compared with conceptual models, physically based spatially distributed models generally have more model parameters, more model grids, and more state variables at each grid. A relatively large number of model grids with more state variables and model parameters results in a high dimensional joint vector of states and parameters, which makes the implementation of EnKF difficult and increases the computational cost. Physically based models require a long adjustment period or spin-up after observations are assimilated

and model states are updated. In physically based models, model formulations and parameters define the equilibrium among model state variables in the system. The equilibrium of the system does not only include the equilibrium between surface water, saturated water storage, and unsaturated water storage within a model grid, but also the equilibrium among different grids. The update of state variables and parameters via EnKF can disrupt the equilibrium in the system [*Pan and Wood*, 2006] which requires a time period for adjustment. The equilibrium needs to be reestablished through the exchange of water among different water components in a single water grid (e.g., infiltration, groundwater recharge, and root zone uptake), and through the exchange of water among different grids (e.g., horizontal groundwater flow and surface flow). During this adjustment or spin-up period, the covariance matrix between the model predictions and the joint vector of states and parameters is contaminated by this spin-up effect. If the assimilation interval is shorter than the adjustment period, the EnKF will update the state variables and model parameters using a contaminated covariance matrix, which will degrade the accuracy of the EnKF analysis. A long assimilation interval, however, means fewer observations can be assimilated, which can affect model performances due to the lack of observations. Therefore, finding the optimal assimilation interval is important.

[9] Identifying critical observations for model parameter estimation is important for model calibration, for enhancing the understanding of the inverse problem of parameter estimation, and for the observational system design at experimental sites. Classically, only discharge data are used for R-R model data assimilation, while soil moisture and land surface temperature data are used for LSMs [e.g., *Houser et al.*, 1998; *Crow and Wood*, 2003; *Pauwels and De Lannoy*, 2006; *Pan and Wood*, 2006; *Clark et al.*, 2008]. Some recent studies have assimilated multiple types of observations (multivariate observations) into hydrologic models (see a review by *Montzka et al.* [2012]). It has been shown that the assimilation of soil moisture in addition to discharge into R-R model improves the forecast of discharge [e.g., *Oudin et al.*, 2003; *Aubert et al.*, 2003; *Francois et al.*, 2003; *Camporese et al.*, 2009a, 2009b; *Bailey and Baù*, 2010; *Lee et al.*, 2011], especially during flood events [*Aubert et al.*, 2003]. *Xie and Zhang* [2010] also found that in synthetic experiments, the assimilation of soil moisture in addition to discharge improves the estimation of model parameters.

[10] This paper presents a demonstration of multiple parameter estimation for a coupled physically based land surface hydrologic model (Flux-PIHM) using multivariate observations via EnKF. The hydrologic land surface model used in this study is Flux-PIHM [*Shi et al.*, 2013a], which is based on the Penn State Integrated Hydrologic Model (PIHM) [*Qu*, 2004; *Qu and Duffy*, 2007; *Kumar*, 2009] and the land surface scheme from the Noah LSM [*Chen and Dudhia*, 2001; *Ek et al.*, 2003]. Six Flux-PIHM parameters, including three hydrologic parameters and three land surface parameters are estimated using EnKF. Synthetic experiments are performed to test the accuracy of EnKF in multiple parameter estimation and to find the optimal time interval for data assimilation. Synthetic observations of discharge, water table depth, soil moisture content, land surface temperature, sensible and latent heat fluxes, and canopy transpiration, and various subsets of those observations are assimilated to identify the observations critical for parameter estimation. The model is implemented at the Shale Hills watershed (0.08 km^2) in central Pennsylvania, where the broad array of observations provides the possibility for a future real-data test.

2. Development of the Flux-PIHM Data Assimilation System

2.1. Flux-PIHM

[11] Flux-PIHM [*Shi et al.*, 2013a] is a coupled land surface hydrologic model. Flux-PIHM incorporates a land-surface scheme into the Penn State Integrated Hydrologic Model (PIHM) [*Qu*, 2004; *Qu and Duffy*, 2007; *Kumar*, 2009]. PIHM is a fully coupled, physically based, and spatially distributed hydrologic model. It simulates channel flow, 2-D overland flow, 1-D unsaturated flow, and 2-D groundwater flow (with dynamic coupling to the unsaturated zone) in a physically based and fully coupled scheme. The land surface scheme in Flux-PIHM is adapted from the Noah LSM [*Chen and Dudhia*, 2001; *Ek et al.*, 2003]. The land surface and hydrologic components are coupled by exchanging water table depth, infiltration rate, recharge rate, net precipitation rate, and evapotranspiration rate between the two model components. Because Flux-PIHM is based on a spatially distributed, physically based, and fully coupled hydrologic model, the computational cost is relatively high. It has previously been manually calibrated and tested at the Shale Hills watershed (0.08 km^2) [*Shi et al.*, 2013a].

2.2. The Ensemble Kalman Filter

[12] After its introduction by *Evensen* [1994], EnKF has been widely used in atmospheric, geographic, and oceanic sciences. It was first developed for dynamic state estimation to improve initial conditions for numerical forecasts, and was later applied to model parameter estimation.

[13] The EnKF formulation used by *Snyder and Zhang* [2003] is adopted in this study. In EnKF, the posterior estimate $\mathbf{x}^a$, i.e., the analysis is given by

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^f\right), \tag{1a}$$

and the analysis error covariance $\mathbf{P}^a$ is given by

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f, \tag{1b}$$

where $\mathbf{x}^f$ is the prior estimate with n state variables, $\mathbf{y}$ is the observation vector of m observations, $\mathbf{H}$ is the observation operator matrix (dimension $m \times n$) which maps state variables to observations, $\mathbf{I}$ is an identity matrix (dimension $n \times n$), and $\mathbf{P}^f$ is the forecast background error covariance (dimension $n \times n$). The Kalman gain matrix $\mathbf{K}$ (dimension $n \times m$) is defined as

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{2}$$

where $\mathbf{R}$ is the observation error covariance (dimension $m \times m$). $\mathbf{P}^f\mathbf{H}^{\mathrm{T}}$ is the forecasted covariance between the

Table 1. Model Variables Included in the Joint Flux-PIHM State-Parameter Vector[a]

Variable	Description	Dimension	Physically Allowable Range
W_c	Water stored on canopy	N_g	$[0, \infty)$ m
h_{snow}	Snow stored on ground and canopy	N_g	$[0, \infty)$ m
h_{ovl}	Overland flow depth	N_g	$[0, \infty)$ m
h_{sat}	Groundwater level	$N_g + N_r$	$[0, D_{\text{BR}}]$
h_{us}	Unsaturated zone soil water storage	N_g	$[0, D_{\text{BR}}]$
h_{riv}	River water level	N_r	$[0, \infty)$
T_{s1-4}	Soil temperature at four layers	$4\ N_g$	$[-273.15, \infty)$°C
T_{sfc}	Surface skin temperature	N_g	$[-273.15, \infty)$°C
H	Sensible heat flux	N_g	$(-\infty, \infty)$ W m^{-2}
LE	Latent heat flux	N_g	$(-\infty, \infty)$ W m^{-2}
E_t	Canopy transpiration	N_g	$[0, \infty)$ m d^{-1}
Q	Outlet discharge	1	$[0, \infty)$ m d^{-1}

[a] N_g and N_r represent the numbers of triangular grids and river segments, respectively, and D_{BR} is the bedrock depth.

states and observed variables, and $\mathbf{HP}^f\mathbf{H}^{\text{T}}$ is the forecasted error covariance of the observed variables.

[14] The state augmentation approach, which has been tested in many studies [e.g., *Annan*, 2005; *Aksoy et al.*, 2006; *Hu et al.*, 2010; *Xie and Zhang*, 2010], is adopted for parameter estimation. In the state augmentation approach, parameters and state variables are concatenated into a joint state vector **x**, and are updated simultaneously by EnKF.

[15] In EnKF applications, filter divergence [*Anderson and Anderson*, 1999] occurs when the uncertainties of prior state variables and parameters become so small that the assimilated observations have little impacts on the posterior analysis. In order to avoid filter divergence, the covariance relaxation method of [*Zhang et al.*, 2004, equation (5)] is used. After the state variables and model parameters are updated by EnKF, the analysis error covariance is inflated using a weighted average between the prior perturbation and the posterior perturbation. The inflated analysis is given by:

$$(\mathbf{x}^a_{\text{new}})' = (1-a)(\mathbf{x}^a)' + a(\mathbf{x}^f)', \tag{3}$$

where a is a weighting coefficient. In this study, a is set to be 0.5 as in the study by *Zhang et al.* [2006]. Because model parameters are not dynamical variables, the values of parameters remain constant in each forecast step. Therefore, the adoption of covariance relaxation is not sufficient to avoid filter divergence caused by constantly decreasing covariance of model parameters. The conditional covariance inflation method [*Aksoy et al.*, 2006] is applied to model parameters in addition to equation (3): the posterior standard deviation of model parameter σ is inflated back to a predefined threshold when the standard deviation is smaller than the threshold. The threshold is chosen as 0.25 σ_0 as in *Aksoy et al.* [2006], where σ_0 is the initial standard deviation of model parameters.

2.3. Implementation of EnKF in Flux-PIHM

[16] The EnKF algorithm is implemented in the Flux-PIHM model system for state and parameter estimation. Flux-PIHM has a large number of model parameters and many of them are soil or vegetation dependent. To decrease the dimension of the joint state-parameter vector, EnKF in our case is applied to global calibration coefficients [*Pokhrel and Gupta*, 2010; *Wallner et al.*, 2012]. A global calibration coefficient is a scalar multiplier applied to the corresponding soil or vegetation related parameter for all soil or vegetation types. By applying global calibration coefficients, the dimension of parameter space for calibration is reduced, and the soil and vegetation parameters of all soil and vegetation types are adjusted in a coherent fashion. The calibration coefficients of those parameters for estimation are included in the joint state-parameter vector. For the sake of simplification, the calibration coefficients of those parameters are represented by the symbols for those original parameters in this paper. Note that this approach requires sound prior knowledge of the relative differences in soil and vegetation parameters across soil and vegetation types.

[17] The model variables included in the augmented state vector are listed in Table 1. Among them, outlet discharge (Q), sensible (H) and latent (LE) heat fluxes, and canopy transpiration (E_t) are diagnostic instead of prognostic variables, i.e., the values of those variables in the future time steps do not depend upon their values at present or previous time steps. They are included in the augmented vector because they are important observable variables, and the observations of those variables can be assimilated into the system to improve state and parameter estimations. Although the augmented vector includes some diagnostic variables, we will still use the term "joint state-parameter vector." The global calibration coefficients of those parameters that need to be estimated are also included in the joint state-parameter vector. If needed, meteorological forcing variables, e.g., precipitation and air temperature, can be regarded as model parameters and concatenated into the joint state-parameter vector as well.

[18] Physical constraints need to be added to ensure the analysis of parameters and model variables in physically realistic or plausible ranges. A quality control of EnKF analysis is performed after each analysis step. For a parameter ϕ constrained in the range between $\phi_{\min}$ and $\phi_{\max}$, the ensemble mean is constrained in the range of $(\phi_{\min}+\Delta, \phi_{\max}-\Delta)$ to make sure the ensemble has a reasonable spread. In this study, Δ is set to be 0.25 σ_0. If the analysis of ensemble mean given by EnKF is out of the range of $(\phi_{\min}+\Delta, \phi_{\max}-\Delta)$, the analysis will be rejected and the parameter values will not be updated. If the analysis of ensemble mean given by EnKF lies in the range of

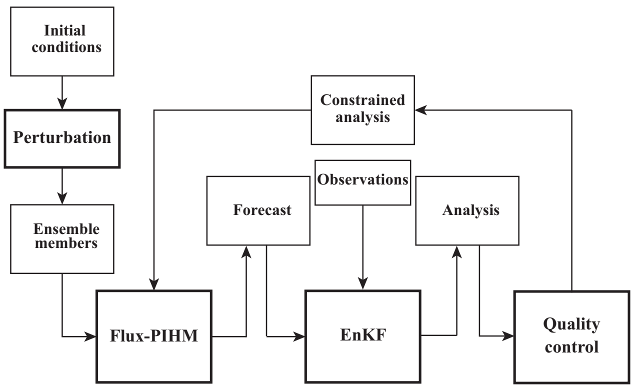

Figure 1. Flowchart of Flux-PIHM data assimilation framework for parameter and state estimation.

$(\phi_{\min}+\Delta, \phi_{\max}-\Delta)$, but some ensemble members are out of the range of $(\phi_{\min}, \phi_{\max})$, each ensemble member is adjusted using

$$\phi_i^{QC} = \frac{\max(\phi^a) - \overline{\phi^a}}{\phi_{\max} - \overline{\phi^a} - \epsilon}\left(\phi_i^a - \overline{\phi^a}\right) + \overline{\phi^a}, \tag{4a}$$

or

$$\phi_i^{QC} = \frac{\overline{\phi^a} - \min(\phi^a)}{\overline{\phi^a} - \phi_{\min} - \epsilon}\left(\phi_i^a - \overline{\phi^a}\right) + \overline{\phi^a}, \tag{4b}$$

where ϕ_i^{QC} is the parameter value of the ith ensemble member after quality control, $\overline{\phi^a}$ is the ensemble mean, and ϵ is a very small number. When equation (4a) or (4b) is applied, the standard deviation of parameters could be smaller than the predefined value in conditional covariance inflation. For model variables, the physically allowable ranges listed in Table 1 are applied. If the analysis of any ensemble member given by EnKF is out of range, the boundary value will be assigned to the analyzed ensemble member. For example, if the analysis of outlet discharge rate of any ensemble member is negative, it will be set to 0.

[19] The workflow of Flux-PIHM parameter estimation using EnKF is presented in Figure 1:

[20] 1. At the beginning, initial conditions of state variables (**x**), or model parameters (Φ), or both are perturbed to generate initial conditions and model parameters for the ith ensemble member, x_i and ϕ_i.

[21] 2. In the forecast step, each ensemble member is put into Flux-PIHM to perform hydrologic and land surface forecasting.

[22] 3. When observations are available, the forecasted variables for each ensemble member x_i^f and the parameters for each ensemble member ϕ_i^f are updated using EnKF by assimilating the observations.

[23] 4. The covariance relaxation method (equation (3)) is applied to both model variables and parameters while conditional covariance inflation [*Aksoy et al.*, 2006] is applied to model parameters if needed.

[24] 5. The quality control process is applied to the analysis of model variables x_i^a and model parameters ϕ_i^a to ensure both model variables and model parameters are constrained in their physically allowable or plausible ranges. The obtained state variables x_i^{QC} and parameters ϕ_i^{QC} are used as initial conditions and parameters for the next forecast step.

[25] 6. Steps 2–5 are repeated until the end of simulation.

[26] In the current methodology, EnKF analysis does not conserve mass and energy. Mass and energy conservation can be achieved by using constrained EnKF [*Pan and Wood*, 2006], which adds another constraint filter for mass and energy budgets after EnKF updates, or by simply rescaling model variables using the ratio between the prior total mass (energy) and the posterior total mass (energy). Those methods both depend on the linearization of mass and energy budget equations. The rescaling method has been tested with Flux-PIHM (results are not shown here), and the system needs a longer adjustment period when mass and energy conservation is applied. Because the objective of the current data assimilation system is to estimate the parameter values, mass and energy conservation does not need to be strictly satisfied at analysis steps. Therefore, mass and energy conservation is not applied to the current data assimilation and parameter estimation experiments, but the option is available if so desired.

3. Experimental Setup

[27] The Flux-PIHM EnKF data assimilation system is implemented at the Shale Hills watershed in central Pennsylvania (Figure 2). The Susquehanna Shale Hills Critical Zone Observatory (SSHCZO) now exists in this watershed. A real-time hydrologic monitoring network (RTHnet) is operating in the SSHCZO. The Shale Hills watershed (0.08 km^2) is a small-scale, forested, V-shaped catchment, characterized by relatively steep slopes and narrow ridges. The surface elevation varies from 256 m above sea level at the watershed outlet to 310 m above sea level at the ridge top. The Shale Hills watershed is in temperate continental climate, with a mean annual temperature of 10°C and a mean annual precipitation of 107 cm. Precipitation is relatively well-distributed year-round. A first-order headstream forms within the watershed, which is mostly dry during summer months. The small scale and the steep slopes make it challenging to perform model calibration, because streamflow and groundwater have larger variability in low-order watersheds than in larger basins [*Reed et al.*, 2004]. *Shi et al.* [2013a] have manually calibrated and evaluated Flux-PIHM at the Shale Hills watershed. The same domain setup and meteorological forcing as in *Shi et al.* [2013a] are adopted in this study. For the synthetic experiment, a truth model run is performed using the manually calibrated parameter values from *Shi et al.* [2013a] starting from the relaxation mode. The truth run starts from 0000 UTC 1 January 2009. The period from 0000 UTC 1 January to 1700 UTC 1 March 2009 is the spin-up period. After the spin-up, predictions from the truth run are used to generate synthetic observations from 1700 UTC 1 March to 0000 UTC 1 August 2009. The outputs from 0000 UTC 1 August to 0000 UTC 1 December 2009 are used to evaluate the estimated model parameters.

[28] The hourly predictions of the following observable variables from the truth run are used:

[29] 1. Outlet discharge rate (Q);

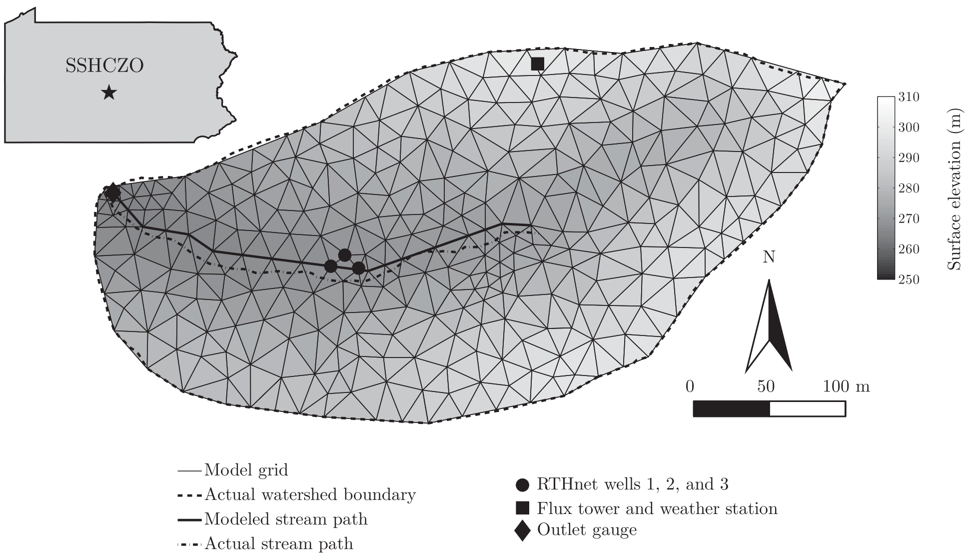

Figure 2. Grid setting for the Shale Hills watershed model domain. The watershed boundary, stream path, surface elevation, and locations of RTHnet measurements used in this study are shown.

[30] 2. Water table depth at the model grid that represents the RTHnet wells (WTD);

[31] 3. Integrated soil moisture content over the soil column at the model grid that represents the RTHnet wells (SWC);

[32] 4. Land surface temperature averaged over the model domain (T_{sfc});

[33] 5. Sensible heat flux averaged over the model domain (H);

[34] 6. Latent heat flux averaged over the model domain (LE); and

[35] 7. Canopy transpiration rate averaged over the model domain (E_t).

[36] To account for the observation uncertainties, synthetic observations are obtained by adding Gaussian white noise to the true time series. The imposed observation errors for different observation types are independent. The white noise added to the truth (Table 2) is designed to represent realistic errors in observational precision. Note that these errors do not represent potential systematic biases in observations. WTD, for example, is a highly precise measurement, but can have systematic offsets that are considerably larger than the precision of the measurement. We do not attempt to simulate the impact of systematic errors on the EnKF system in this manuscript. We determine observational precisions with a combination of instrument specifications and prior literature.

[37] At the Shale Hills watershed, discharge is measured with a V-notch weir at the outlet of the catchment. A Campbell CS420-L (0–10 psi) pressure transducer measures the water level, which is then converted to discharge rate using a rating curve developed by *Nutter* [1964] for the V-notch weir at the Shale Hills watershed. The calibrated rating curve is:

$$Q=\begin{cases} 2446.58\times10^{-5.56+181.67-2778.15x^2}, & 0<x\le0.034\,\mathrm{m}, \\ 3.08\times10^4x^{2.46}, & 0.034\,\mathrm{m}<x\le0.100\,\mathrm{m}, \\ 3.12\times10^6x^{4.47}, & x>0.100\,\mathrm{m}, \end{cases} \tag{5}$$

where x is the measured water level (m), and Q is the discharge rate ($\mathrm{m^3\ d^{-1}}$). The precision of Campbell CS420-L transducers is $\pm0.1\%$ full scale (0–10 psi), which is equivalent to about 7 mm of water level [*Campbell Scientific Inc.*, 2007]. Figure 3a shows the rating curves with 7 mm errors in measured water level (Q_+ and Q_-). *Clark et al.* [2008] found that converting discharge to log space improves EnKF performance. Their strategy is adopted in this study. Prior to each analysis step, the discharge observation Q^o is converted to $\ln(Q^o+\epsilon)$, and for each ensemble member i, model discharge forecast Q_i^f is converted to $\ln(Q_i^f+\epsilon)$, where ϵ is a very small discharge rate (set to $10^{-4}\ \mathrm{m^3\ d^{-1}}$ in this study) used to avoid taking the logarithm of a zero discharge rate. The precision of discharge measurement in log space at the Shale Hills watershed is approximated by

$$\sigma_{\ln Q}=0.5(\ln Q_+-\ln Q_-), \tag{6}$$

as shown in Figure 3b. To simplify the calculation of $\sigma_{\ln Q}$, two linear segments are used to fit the $\sigma_{\ln Q}$ curve:

Table 2. Standard Deviation of Gaussian White Noise Added to Each Observation Data Set

Data Set	Standard Deviation of Gaussian White Noise
Outlet discharge rate ($\mathrm{m^3\ d^{-1}}$)	Equation (7)[a]
Water table depth (m)	0.007 m[a]
Integrated soil moisture ($\mathrm{m^3\ m^{-3}}$)	0.01 $\mathrm{m^3\ m^{-3}}$[b]
Land surface temperature (°C)	1°C[c]
Sensible heat flux ($\mathrm{W\ m^{-2}}$)	10% of flux[d]
Latent heat flux ($\mathrm{W\ m^{-2}}$)	10% of flux[d]
Transpiration rate ($\mathrm{mm\ d^{-1}}$)	10% of flux[d]

[a]Precision of Campbell CS420-L pressure transducers.

[b]Precision of Decagon Echo2 EC-20 soil moisture sensors.

[c]*Wan and Li* [1997]; *Yu et al.* [2008]; *Coll et al.* [2009]; *Wang and Liang* [2009].

[d]*Lenschow and Stankov* [1986]; *Lenschow et al.* [1994]; *Baldocchi et al.* [1996]; *Finkelstein and Sims* [2001]; *Baldocchi* [2003]; *Richardson et al.* [2006]; *Salesky et al.* [2012].

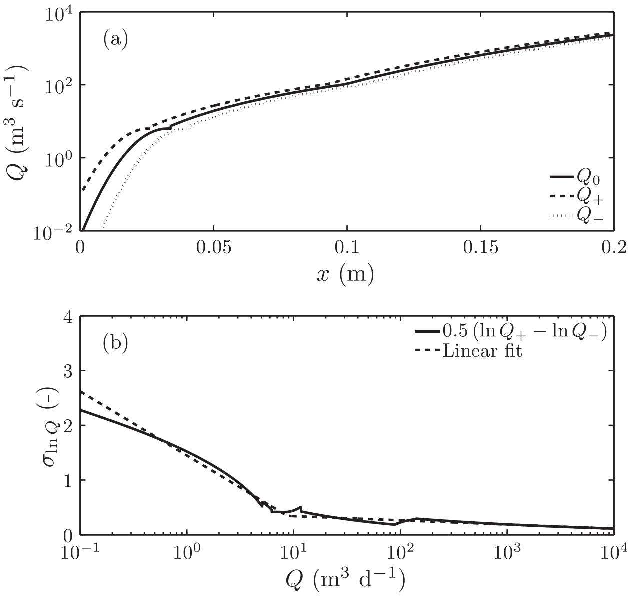

Figure 3. (a) Rating curves for the Shale Hills watershed outlet V-notch weir and (b) representative discharge measurement error. The dashed and dotted lines in Figure 3a represent rating curves with 7 mm error in measured water level. The dashed line in Figure 3b represents the manual fitting curve for the representative error.

$$\sigma_{\ln Q}=\begin{cases} -0.509\ln(Q+\epsilon)+1.448, & Q+\epsilon\le 8.731\,\mathrm{m^3d^{-1}},\\ -0.0332\ln(Q+\epsilon)+0.417, & Q+\epsilon> 8.731\,\mathrm{m^3d^{-1}}.\end{cases} \tag{7}$$

[38] Equation (7) is used in this study to estimate the observation error of discharge in log space. The groundwater level at the RTHnet wells are also measured with Campbell CS420-L (0–10 psi) pressure transducers, with a precision of 7 mm. The soil moisture contents are measured using Decagon Echo2 EC-20 soil moisture sensors. *Czarnomski et al*. [2005] examined the precision of the EC-20 sensors and found that the precision of which is about 4.8%. Because the annual average soil moisture content at the Shale Hills watershed is about 0.2 m^3 m^{-3}, we thus assume that the precision of the soil moisture sensors is 0.01 m^3 m^{-3}. The observation error of land surface temperature is assumed to be 1.0°C, based on prior validations of MODIS land surface temperature product [*Wan and Li*, 1997; *Yu et al*., 2008; *Coll et al*., 2009; *Wang and Liang*, 2009]. The observation errors of sensible and latent heat fluxes (H and LE) are assumed to be 10% based on extensive prior study of the nature of random errors in eddy covariance flux measurements [*Lenschow and Stankov*, 1986; *Lenschow et al*., 1994; *Baldocchi et al*., 1996; *Finkelstein and Sims*, 2001; *Baldocchi*, 2003; *Richardson et al*., 2006; *Salesky et al*., 2012]. Careful assessment of the precision of watershed-scale transpiration measurements (E_t) is lacking, so we have used the same precision estimate as for eddy covariance flux measurements.

[39] The parameters to be estimated in this study and their a priori values are presented in Table 3: the effective porosity Θ_e, the *van Genuchten* [1980] soil parameter α, the van Genuchten soil parameter β, the *Zilitinkevich* [1995] parameter C_{zil}, the minimum stomatal resistance $R_{c\,\min}$, and the reference canopy water capacity S. These six parameters show high distinguishability, observability, and simplicity [*Zupanski and Zupanski*, 2006; *Nielsen-Gammon et al*., 2010] in the parameter sensitivity analysis [*Shi et al*., 2013b]. High distinguishability, observability, and simplicity have been proven critical for EnKF parameter estimation [*Nielsen-Gammon et al*., 2010; *Hu et al*., 2010; *Aksoy et al*., 2006]. Therefore these six parameters are selected for EnKF parameter estimation. The physically plausible ranges of those parameters are obtained from previous studies [e.g., *Beven and Binley*, 1992; *Chen et al*., 1997; *Gupta et al*., 1999; *Eckhardt and Arnold*, 2001; *Anderton et al*., 2002; *Tang et al*., 2006] and experience from manual calibration [*Shi et al*., 2013a], and are presented in Table 3. Details about those parameters can be found in *Shi et al*. [2013a, 2013b]. The parameters that are not estimated are set to their manually calibrated values as in *Shi et al*. [2013a]. The Flux-PIHM Shale Hills watershed model domain has 535 triangular grids and 20 river segments. Including the variables in Table 1 and the six parameters (global calibration coefficients) in Table 3, the total dimension of the joint state-parameter vector is 7002.

[40] Several test cases are used for the synthetic data experiments (Table 4). For each test case, a total of 30 ensemble members are involved. Ensemble runs with 50 ensemble members have been tested, but the increase in ensemble members does not measurably improve the results in terms of the mean squared errors of the estimated parameters. Therefore, 30 ensemble members are used to reduce computational cost. To generate different ensemble members, calibration coefficients of those six parameters are randomly perturbed within their plausible ranges. For

Table 3. Flux-PIHM Model Parameters to be Estimated and the Plausible Ranges of Their Calibration Coefficients

Parameter	Description	A Priori Value					Range of Calibration Coefficient
		Soil Type Weikert	Berks	Rushtown	Blairton	Ernest	
Θ_e	Effective porosity (m^3 m^{-3})	0.48	0.32	0.33	0.29	0.34	0.3–1.2
α	Van Genuchten soil parameter (m^{-1})	2.46	2.51	2.84	2.79	3.27	0–2.5
β	Van Genuchten soil parameter	1.20	1.21	1.33	1.33	1.32	0.95–2.5
		Vegetation Type Decidous Forest		Evergreen Forest	Mixed Forest		
$R_{c\,\min}$	Minimum stomatal resistance (s m^{-1})	100		150	125		0.3–1.2
S	Reference canopy water storage (mm)	0.20		0.20	0.20		0–5
		Other					
C_{zil}	Zilitinkevich parameter		0.10				0.1–10

Table 4. Initial Ensemble Mean Of Parameters, Assimilation Intervals, and Assimilated Observations of Different Test Cases[a]

Case	Initial Ensemble Mean	Assimilation Interval	Q	WTD	SWC	T_{sfc}	H	LE	E_t
Control run (CR)	$0.5(\phi_{\min}+\phi_{\max})$	72 h	X	X	X	X	X	X	X
Case+	$0.5(\phi_{\min}+\phi_{\max})-\sigma_0$	72 h	X	X	X	X	X	X	X
Case−	$0.5(\phi_{\min}+\phi_{\max})+\sigma_0$	72 h	X	X	X	X	X	X	X
96 h	Same as CR	96 h	X	X	X	X	X	X	X
48 h	Same as CR	48 h	X	X	X	X	X	X	X
24 h	Same as CR	24 h	X	X	X	X	X	X	X
Q	Same as CR	72 h	X						
SSHCZO	Same as CR	72 h	X	X	X		X	X	
NoSWC	Same as CR	72 h	X	X		X	X	X	X
NoWTD	Same as CR	72 h	X		X	X	X	X	X
QST	Same as CR	72 h	X		X	X			

[a]X indicates the listed observation is assimilated, and blank indicates the observation is not assimilated.

each parameter (calibration coefficient) ϕ, the values are randomly drawn from a Gaussian distribution, with an initial standard deviation of $\sigma_0=0.2(\phi_{\max}-\phi_{\min})$. The initial ensemble means for different test cases are presented in Table 4. The parameter C_{zil} is perturbed in log space because the plausible range for C_{zil} spans orders of magnitude. The correlation coefficients between different parameters among all ensemble members are examined to guarantee that the initial correlation coefficient (absolute value) between any two of those parameters is ≤ 0.25. Same as the truth run, all of the ensemble members start from 0000 UTC 1 January 2009, and the calibration period is from 1700 UTC 1 March to 0000 UTC 1 August 2009. All model runs start from saturation in relaxation mode. Because accurate in situ meteorological observations (e.g., precipitation and temperature) are collected at this small-scale watershed (0.08 km^2), we assume that the uncertainties in meteorological forcing are negligible, compared with the uncertainties in model parameters. Therefore, the meteorological forcing used for each ensemble member is the same as for the truth run. The first set of observations is assimilated at 1700 UTC 1 March 2009. We choose to assimilate observations at the midday time steps (1700 UTC) because the land surface fluxes have the strongest correlation with model parameters at midday. Thus, assimilating observations at midday is the most effective in correcting the biases in model states and parameters. Different assimilation intervals (24, 48, 72, and 96 h) and different combinations of observations are tested (Table 4).

[41] The assumption of homogeneous atmospheric forcing is well justified given the very small (microscale) grid sizes in our model [*Mahrt*, 2000]. We conducted a single test of our assumption that meteorological forcing uncertainties are negligible and found no impact of 10% random errors in precipitation, the meteorological input variable most likely to suffer from random observational errors.

[42] Each parameter is judged to have converged when the standard deviation of the parameter decreases to 0.25 σ_0, where 0.25 σ_0 is also the threshold specified for the conditional covariance inflation method. The temporal average of the ensemble mean after convergence is considered to be the calibrated value of each parameter. If the parameter does not converge during the calibration period, i.e., the standard deviation of parameter is always $>0.25\ \sigma_0$, the temporal average of the parameter values between 0000 UTC 1 July and 0000 UTC 1 August 2009 is calculated as the calibrated value for the parameter.

[43] Flux-PIHM runs using the calibrated parameter values from different test cases are performed and compared with the truth run to evaluate the estimated parameter values. Besides those test cases in Table 4, a NoPE (no parameter estimation) run is also performed. In the NoPE evaluation run, the calibration coefficients for those six parameters are set to 1.0, which means those parameters are uncalibrated and the a priori parameter values in Table 3 are used. All of the evaluation runs start from 0000 UTC 1 January 2009 from the relaxation mode. Model forecasts from 0000 UTC 1 August to 0000 UTC 1 December 2009, which is the period right after the calibration period, are used to evaluate the model performance. The predictions of all observable variables are compared with the truth run. For hydrologic variables (Q, WTD, and SWC), comparisons are made at every hour. For land surface variables (T_{sfc}, H, LE, and E_t), comparisons are made only at 1700 UTC on every day to filter out the diurnal cycles of surface fluxes.

4. Results

4.1. Accuracy of EnKF Parameter Estimation

[44] The control run is used to examine the accuracy of EnKF parameter estimation. In the control run, the initial guess of parameter values is at the center of their uncertainty ranges (Table 4). Figure 4 presents the true values and the temporal evolutions of the parameters from the control run. All of the six parameters approach their true values (Figure 4). After about 10 observation cycles (about 1 month simulations after the first set of observations is assimilated), all parameters are very close to their true values, with the true values inside or close to the 1 standard deviation (1-σ) spreads. The estimated parameter values oscillate around the true values after they approach them (Figure 4).

[45] The standard deviations of those parameters are decreasing over time before convergence (Figure 4), which indicates the decrease in parameter uncertainties. All of the parameters have converged (i.e., the standard deviations of the parameters have decreased to 0.25 σ_0). Among the six parameters, the standard deviations of α, β, and C_{zil} drop the fastest because of their strong correlations with state variables [*Shi et al.*, 2013b]. The standard deviations of Θ_e and S drop the slowest because the identifiability of Θ_e is only high at discharge peaks, and S is only effective when the canopy is wet, as found in *Shi et al.* [2013b].

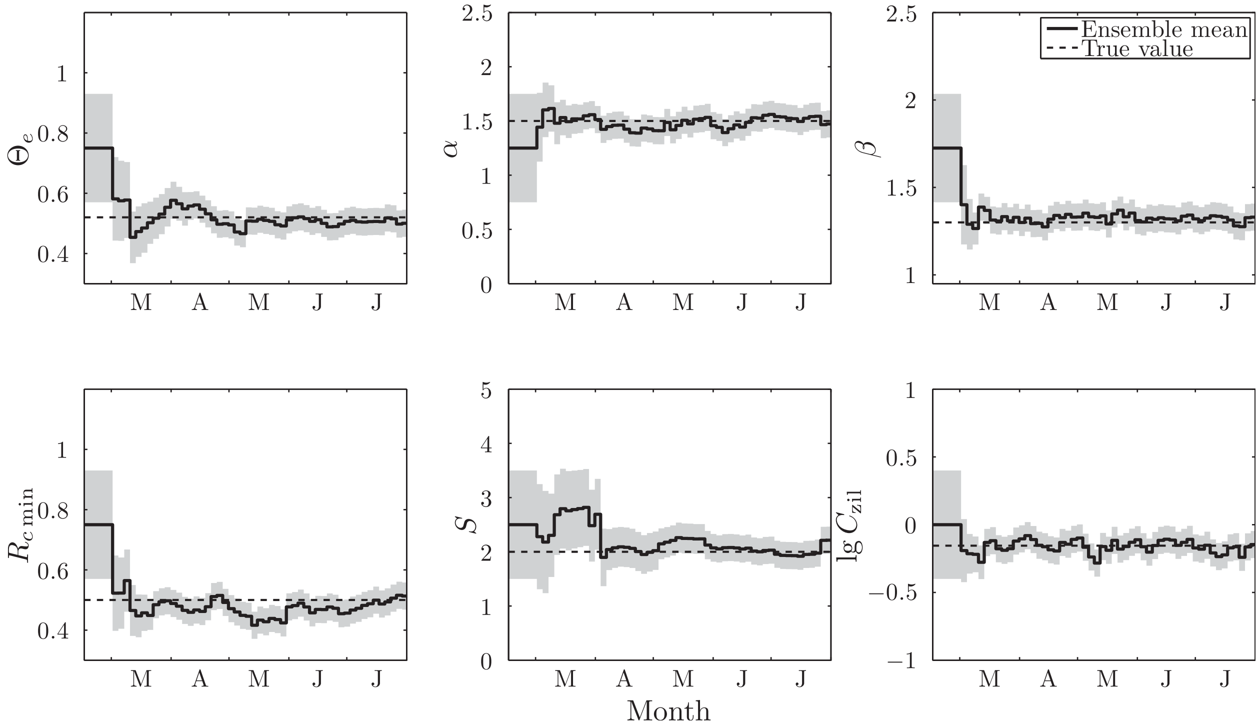

Figure 4. True values and temporal evolutions of parameters from the control run. The dashed lines represent the true parameter values, and the shaded areas represent the one standard deviation (1-σ) spread.

[46] Figure 5 demonstrates the capability of EnKF in state estimation, and two representative observable variables Q and T_{sfc} are presented. Figure 5 shows that EnKF is effective in optimizing model variables. It is especially apparent in the second assimilation cycle for Q (Figure 5a): although the prior (ensemble prediction) of Q has a relatively large error, EnKF successfully corrects the error in Q and significantly improves the estimate. Before the calibration period, the model errors and uncertainties (especially for T_{sfc}) are both relatively large. When more observations are assimilated, the errors and uncertainties of model parameters decrease (Figure 4), thus the model errors and uncertainties of variables decrease (Figure 5). Figure 5a shows some spin-up effects in Q. There is always

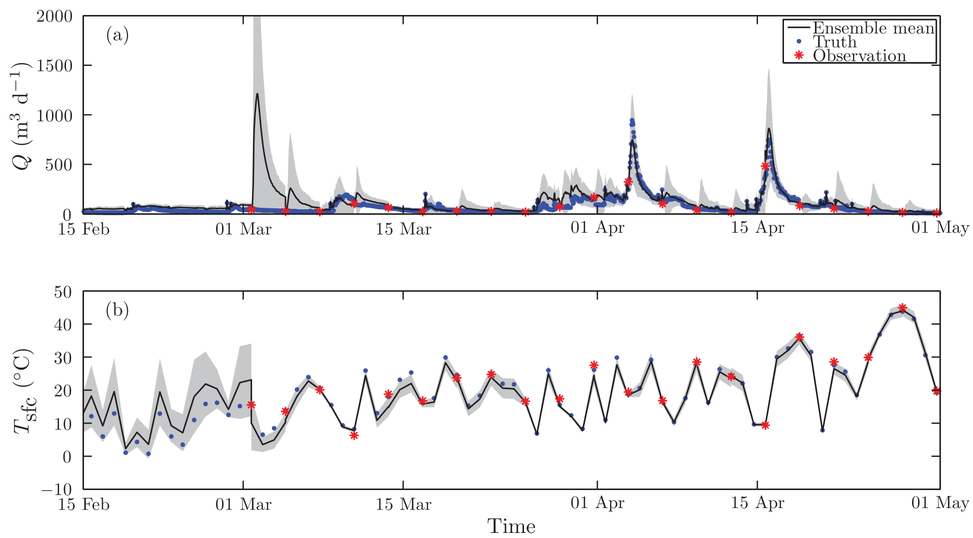

Figure 5. Temporal evolution of the ensemble prediction of (a) discharge and (b) mid-day land surface temperature in the control run from 0000 UTC 1 February to 0000 UTC 1 May 2009. The shaded area represents the 1-σ spread of the ensemble prediction.

Table 5. Calibrated Parameter Calibration Coefficients From Different Test Cases[a]

Case	Θ_e	α	β	C_{zil}	$R_{c\ min}$	S
True value	0.52	1.50	1.30	0.70	0.50	2.00
Control run	0.51	1.48	1.32	0.70	0.47	2.02
Case+	0.52	1.59	1.30	0.71	0.48	1.87
Case−	0.54	1.50	1.32	0.70	0.48	1.91
Q	(0.49)	**0.81**	**(1.51)**	**(0.46)**	**(0.60)**	(1.93)
SSHCZO	0.49	1.46	1.33	0.88	**0.45**	**(2.40)**
NoSM	0.52	**0.91**	**1.46**	0.70	0.48	2.01
NoWTD	0.48	1.46	1.33	0.70	0.47	1.89
QST	0.48	1.38	1.36	0.66	0.52	**(3.44)**
NoPE	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

[a]Calibrated values in bold font indicate that the estimated values have errors $>0.25\sigma_0$. Estimated values in parentheses indicate that the estimation of the parameter does not converge, i.e., the standard deviation of the parameter is always $>0.25\sigma_0$ during the calibration period.

an increase in model error and uncertainty at the beginning of each forecast cycle, because the update of state variables and parameters via EnKF disrupts the equilibrium in the system. This spin-up effect is more significant in hydrologic variables (Q, WTD, and SWC) than in land surface variables (T_{sfc}, H, LE, and E_t). Figure 5b shows an overshooting effect. When the first set of observations is assimilated, the analysis (posterior) of T_{sfc} overshoots the observation of T_{sfc}, due to the effects of other types of observations on T_{sfc}.

[47] The calibrated parameter values are listed in Table 5. Errors of calibrated parameter values in the control run are all $<0.25\ \sigma_0$. Because 0.25 σ_0 is the threshold specified for the conditional covariance inflation method, an error $<0.25\ \sigma_0$ indicates that the true parameter value is within the 1-σ spread of ensemble mean after convergence. The comparisons between the evaluation run using the calibrated parameter set and the "truth" are presented in Figure 6. Note that the evaluation run is the deterministic model run using the calibrated parameter set in Table 5, but not the ensemble calibration run shown in Figure 5. Compared with the NoPE run, calibrated parameter values from the control run significantly improve the model predictions, especially for the hydrologic variables (Figure 6). Predictions of all observable variables from the evaluation run agree well with the truth (Figure 6). Both the correlation coefficients and the normalized root-mean-square errors (RMSEs) of the predictions are very close to 1.0. The mean biases in different predictions are negligible.

4.2. Optimal Assimilation Interval

[48] The control run (with a 72 h assimilation interval), 96, 48, and 24 h cases (Table 4) are used to find the optimal assimilation interval for parameter estimation. The same 30 ensemble members are used to start each test case. Figure 7 presents the true values and the temporal evolutions of the parameters from those test cases. Generally, as shown in Figure 7, the performance of parameter estimation is the worst when the assimilation interval is 24 h. For the 24 h case, EnKF keeps increasing α, and decreasing β and Θ_e to compensate the spin-up effect. Differences among 48 h, control run (72 h), and 96 h are not significant in Figure 7.

[49] The RMSEs and absolute biases of the estimated parameter values for those test cases after convergence are calculated to quantify the effects of assimilation intervals. The results are presented in Figure 8. The RMSEs presented here are normalized by the RMSEs in the control run. For all the parameters except for C_{zil}, RMSEs and absolute biases decrease monotonically when the assimilation interval increases from 24 to 72 h (Figure 8). For the parameter C_{zil}, there is no obvious tendency with respect to the assimilation interval. This spin-up effect is the most prominent in the parameter α. When the assimilation interval increases from 72 to 96 h, no significant improvement in parameter estimation (in terms of RMSEs and absolute biases) is found (Figure 8). It suggests that the assimilation interval of 72 h is long enough to eliminate the impacts of spin-up effect in the synthetic experiments. Although longer assimilation intervals would also be sufficient to avoid the spin-up effect (e.g., the 96 h case), longer assimilation intervals mean that fewer observations would be assimilated into the system during the same simulation period. Therefore, 72 h is the optimal assimilation interval for the synthetic experiments at the Shale Hills watershed.

4.3. Sensitivity to Initial Parameter Values

[50] The control run, Case+, and Case− (Table 4) are used to demonstrate the sensitivity of EnKF parameter estimation to different initial parameter values. Figure 9 presents the true values and the temporal evolutions of the estimated parameters from those three test cases. In all of the three test cases, all six parameters approach their true values (Figure 9). Starting from different initial guesses, the estimated parameter values from different test cases become close after about 2 month simulation and data assimilation. The temporal fluctuations of parameter values from different test cases are similar. Those fluctuations are mostly caused by the observation errors in the synthetic observations.

[51] All of the parameters from those three test cases have converged (i.e., the standard deviations of the parameters have decreased to $0.25\sigma_0$). Errors of the calibrated parameter values from those three test cases are all $<0.25\ \sigma_0$ (Table 5), indicating that the true parameter values are within the 1-σ spread of the ensemble prediction after convergence. The comparisons between the evaluation runs using the calibrated parameter sets and the truth are presented in Figure 6. The performances of the evaluation runs using Case+ and Case− parameters are very similar to the control run, and show significant improvements in model predictions over the NoPE run.

4.4. Efficiency of Assimilating Different Observations

[52] The control run, Q, SSHCZO, NoSWC, NoWTD, and QST cases are compared to illustrate the efficiency of

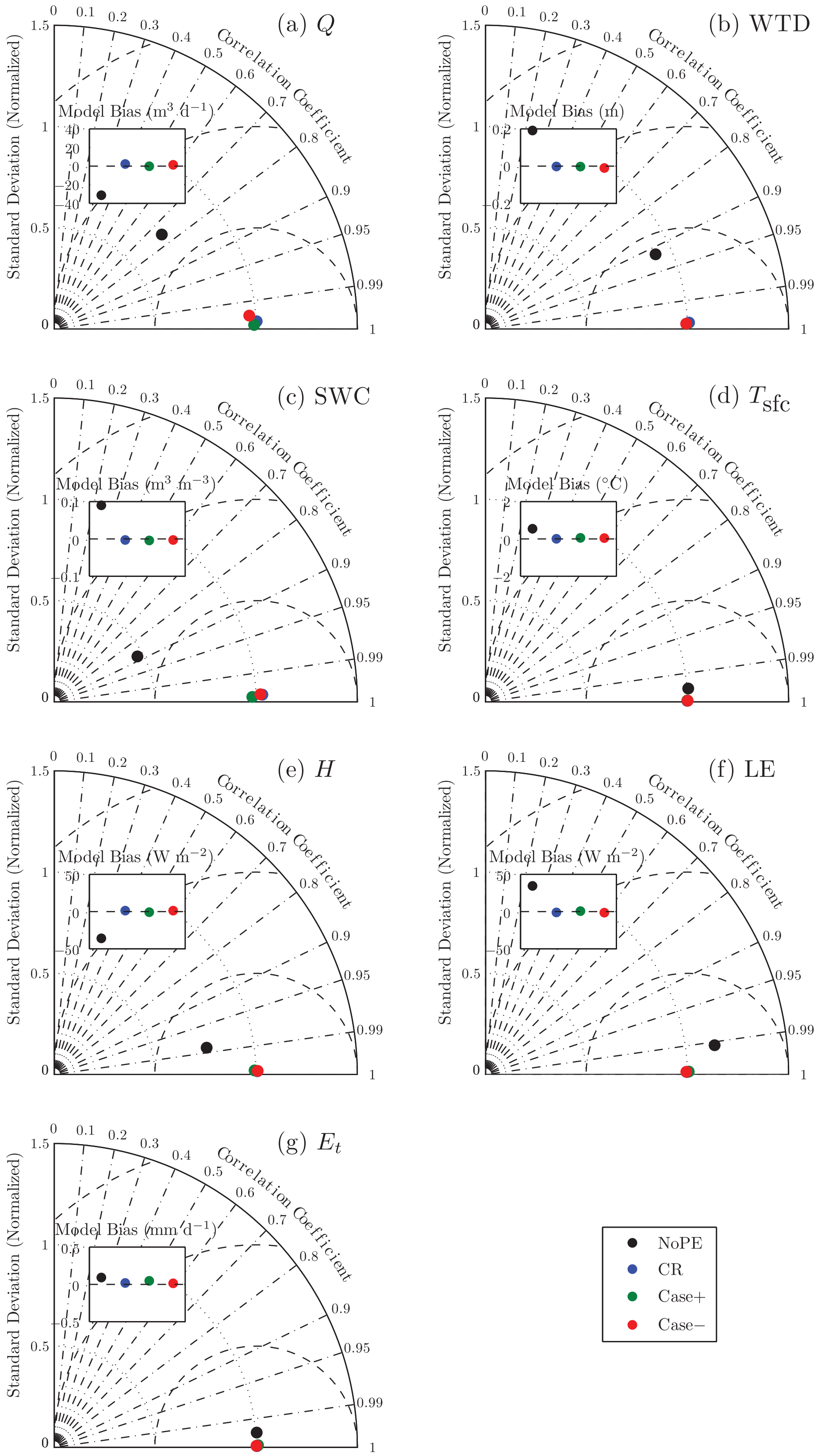

Figure 6. Evaluation of the model predictions using the estimated parameter sets from the control run (CR), Case+, Case−, and NoPE. Correlation coefficient, normalized standard deviation, and root mean squared error are presented in Taylor diagrams. Insets show the average hourly biases.

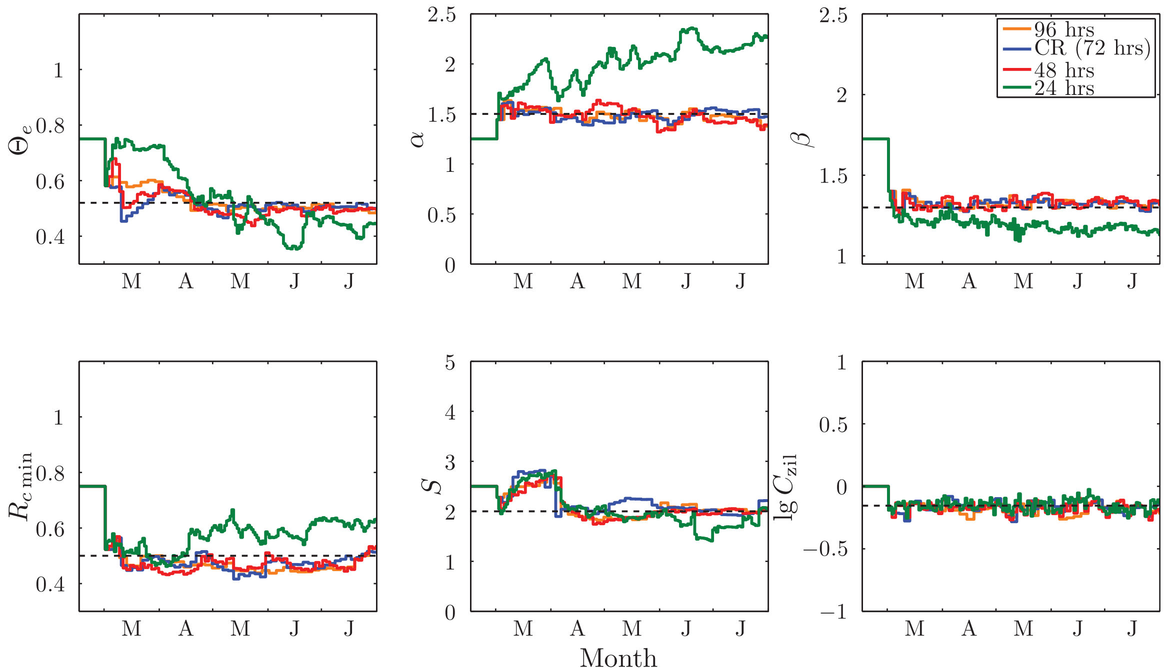

Figure 7. True values and temporal evolutions of parameters from the control run (CR; 72 h), 96, 48, and 24 h. The dashed lines represent the true parameter values.

assimilating different observations. Among them, the Q case only assimilates the discharge observations as in most previous studies of hydrologic model calibrations. The SSHCZO case uses those synthetic observations that represent the observations available at the Shale Hills Critical Zone Observatory (SSHCZO) within the Shale Hills watershed. The NoSWC and NoWTD test cases eliminate soil moisture and water table depth observations, respectively. The QST case assimilates the discharge, soil moisture, and land surface temperature observations, which are assumed to be the essential observations for Flux-PIHM at the Shale Hills watershed. Figure 10 presents the true values and the temporal evolutions of the parameters from those test cases. The same 30 ensemble members are used to start each test case. The calibrated values for each parameter in different test cases are listed in Table 5. The comparisons of observable variables between evaluation runs using the calibrated parameter sets and the truth are presented in Figure 11.

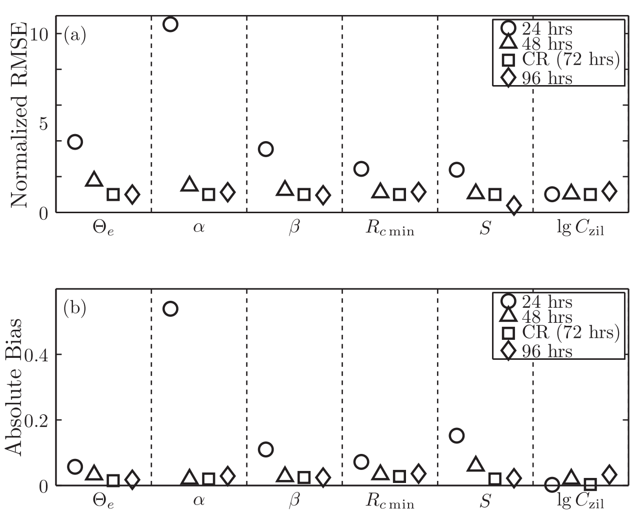

Figure 8. RMSEs and absolute biases of the estimated parameter values after convergence. RMSEs from all test cases are normalized by the RMSEs in control run (CR).

[53] Figure 10 and Table 5 show that when discharge is the only observation data set assimilated into the system, EnKF can only provide good estimates for model parameters Θ_e and S, the errors of which are $<0.25\sigma_0$. Except for the parameter α, the other calibrated parameters do not converge in this test case (Table 5), and the calibrated parameter values still have relatively large uncertainty. Compared with the NoPE evaluation run, although this test case (Q) provides good estimates for only two of the six model parameters, the calibrated parameters from this test case strongly improve the prediction of discharge (Figure 11a). Comparison of the discharge prediction with the truth shows a high correlation coefficient (about 0.99) and a normalized standard deviation comparable with other test cases, although this test case underestimates the total discharge by 10.31 $m^3\,d^{-1}$ (10.59%; Figure 11a). The assimilation of discharge observations helps the system obtain model parameters that can produce reasonable discharge predictions. For the other two hydrologic variables (WTD and SWC), the correlation coefficients are only better than the NoPE run, but lower than the other test cases, especially for SWC. It indicates that parameters obtained in the Q case have limited ability in resolving the temporal pattern of wetting and drying in WTD and SWC. The SWC simulation also significantly overestimates the amplitude of temporal variation in SWC, and has a relatively large model bias. Due to the lack of land surface variable observations, estimations of land surface parameters ($R_{c\,\min}$ and C_{zil}) are poor (Figure 10 and Table 5). The calibrated parameters in

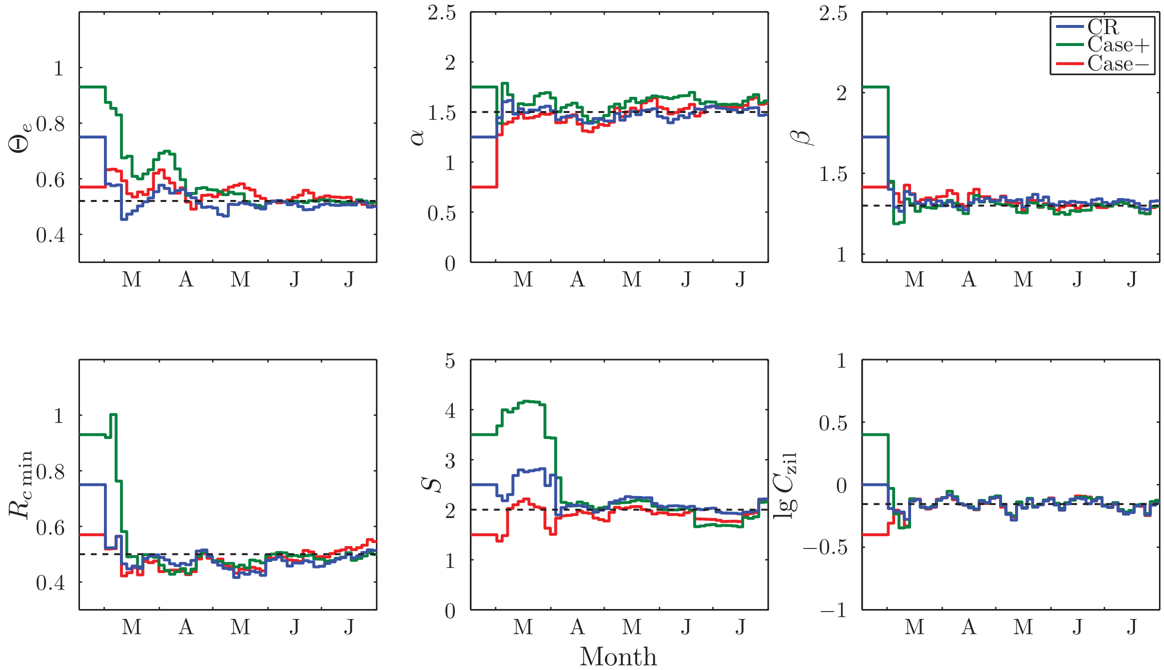

Figure 9. True values and temporal evolutions of parameters from the control run (CR), Case+, and Case−. The dashed lines represent the true parameter values.

this test case (Q) cannot reproduce the temporal variation of land surface variables well (Figures 11d–11g).

[54] When SWC is not assimilated into the system (the NoSWC test case), EnKF cannot provide good estimates of α and β, and the errors in α is much $>0.25\sigma_0$ (Figure 10 and Table 5). The sensitivity analysis of Flux-PIHM showed that the effect of α is the most significant in SWC [*Shi et al.*, 2013b]. In this test case, EnKF underestimates α, and thus produces a relatively large bias in SWC (Figure 11c). Although the parameter values of α and β estimated in this test case have relatively large errors, the discharge and WTD predictions using these estimated parameter values are comparable to the control run.

[55] The calibrated parameter values from the NoWTD case are very close to the control run (Figure 10 and Table 5). The predictions using those calibrated parameter

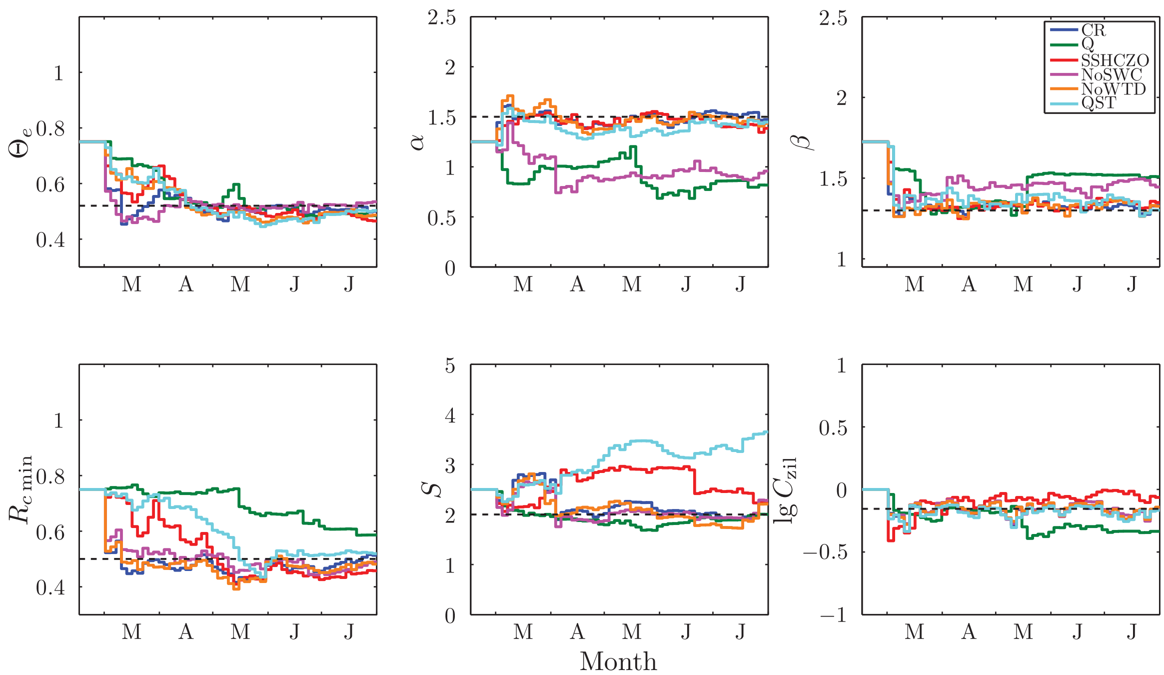

Figure 10. Same as Figure 9, but for the control run (CR), Q, SSHCZO, and NoSWC, NoWTD, and QST.

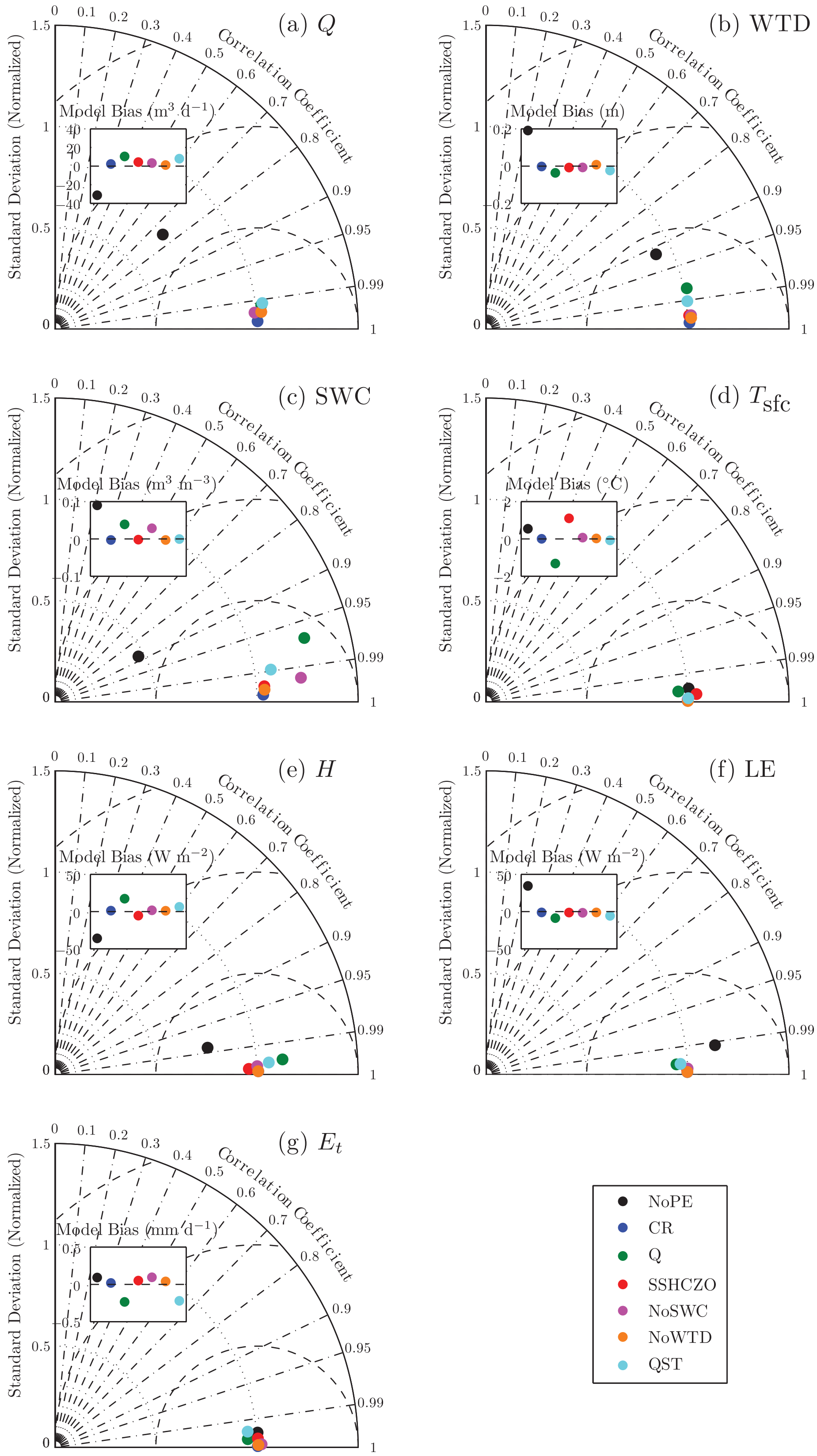

Figure 11. Same as Figure 6, but for the control run (CR), Q, SSHCZO, NoSWC, NoWTD, QST, and NoPE.

Table 6. Average Correlation Coefficients Between Estimated Parameter Values From 0000 UTC 1 June to 0000 UTC 1 August 2009 in Three Different Test Cases[a]

	Θ_e	α	β	C_{zil}	$R_{c\ min}$	S
Θ_e	1.00	−0.04	−0.06	−0.02	0.07	−0.06
	1.00	0.10	−0.11	0.02	0.03	0.00
	1.00	−0.03	−0.03	0.01	0.02	−0.02
α		1.00	−0.47	0.08	−0.29	0.06
		1.00	−0.36	0.00	−0.26	−0.06
		1.00	−0.31	0.02	−0.45	−0.02
β			1.00	−0.09	−0.05	0.00
			1.00	−0.06	−0.05	0.05
			1.00	−0.07	0.11	0.04
C_{zil}				1.00	−0.26	0.00
				1.00	−0.20	0.03
				1.00	−0.25	−0.02
$R_{c\ min}$					1.00	−0.20
					1.00	−0.16
					1.00	−0.03
S						1.00
						1.00
						1.00

[a]The correlation coefficients shown in each cell are in sequence for control run, Case+, and Case− (from top to bottom).

values are as good as the control run (Figure 11), which suggests that effect of assimilating WTD observations is limited.

[56] The QST case is used to test the most essential observations. Assimilating only three types of the seven available observations, the estimated parameter values from the QST case are close to the control run, except for the parameter *S*, which does not converge during the calibration period (Figure 10 and Table 5). The evaluation run predictions of the QST test case are comparable to other test cases (Figure 11), although the calibrated parameter set produces relatively large biases in discharge and transpiration rate compared with other test cases. On average, this test case overestimates the discharge by 8.11% (Figure 11a), and underestimates the midday transpiration by 6.26% (Figure 11g).

[57] The SSHCZO case does not assimilate T_{sfc} and E_t. The calibrated values of parameters, except for $R_{c\ min}$ and *S*, are very close to the true values (Figure 10 and Table 5). The predictions of the hydrologic variables are almost as good as in the control run (Figures 11a–11c). For the land surface variables, the prediction of the SSHCZO case overestimates T_{sfc} by 1.10°C, because observations of T_{sfc} are not assimilated, but the predictions of *H*, LE, and E_t are only slightly worse than the control run (Figures 11d–11g). In spite of the lack of T_{sfc} and E_t observations, the assimilation of *H* and LE are sufficient to represent land surface states.

4.5. Parameter Interaction

[58] Because EnKF is based on ensemble generation, the relationship among different ensemble members reveals the interactions between model parameters. Table 6 presents the temporal average of correlation coefficients between the estimated parameter values from 0000 UTC 1 June to 0000 UTC 1 August 2009 for three different test cases. This period is chosen because all parameters from those three test cases have converged during this period. Most of the parameter pairs have relatively low correlation coefficients between −0.2 and +0.2 after convergence (Table 6). Although the initial ensemble is generated such that the initial correlation coefficient (absolute value) between any two of the parameters is ≤0.25, there are three pairs of parameters, α-β, α-$R_{c\ min}$, and C_{zil}-$R_{c\ min}$ that show correlations (absolute values) >0.25 after convergence (Table 6). Among them, α-β and α-$R_{c\ min}$ show relatively high correlations in all three test cases. Because α and β are hydrologic parameters, and C_{zil} and $R_{c\ min}$ are land surface parameters, those three pairs of parameters, respectively, represent the interaction between hydrologic parameters, the interaction between land surface and subsurface, and the interaction between land surface parameters.

5. Discussion and Conclusions

[59] This paper presents the multiple parameter estimation of a coupled physically based land surface hydrologic model (Flux-PIHM) using multivariate observations via EnKF. Results demonstrate that, given a limited number of site-specific observations, the EnKF can be used to provide good estimates of Flux-PIHM model parameters, with associated uncertainties. The EnKF data assimilation system designed in this study provides a new approach for physically based hydrologic model calibration using multivariate observations. The sequential parameter estimation can save considerable manual labor required for the implementation of hydrologic models, especially physically based models, at different watersheds.

[60] The test cases with different assimilation intervals show that the spin-up effect degrades the accuracy of the estimated hydrologic parameters. The spin-up effect is more prominent for the hydrologic parameters than the land surface parameters. In this study at the Shale Hills watershed, the assimilation interval of 72 h is found to be optimal for the synthetic experiments.

[61] The performance of the test case Q indicates that assimilating discharge alone can improve the prediction of discharge, however, the improvement is limited compared with other test cases. The predictions of subsurface variables (especially SWC) and land surface variables in this test case are poor compared with other test cases (Figure 11). The prediction of discharge can be significantly improved when SWC observations are assimilated. Those findings agree with the findings of *Camporese et al.* [2009a, 2009b], *Bailey and Baù* [2010], and *Lee et al.* [2011]. This test case (Q) shows that assimilating discharge observation alone cannot provide reliable land surface parameter (C_{zil} and $R_{c\ min}$) estimation.

[62] The effect of WTD observations is not strong when discharge and SWC observations are assimilated. At the Shale Hills watershed over 80% of annual discharge comes from subsurface runoff [*Shi et al.*, 2013a], thus the temporal variations of discharge and WTD are well correlated. In addition, WTD and SWC observations are also highly correlated [*Shi et al.*, 2013a]. Therefore, WTD is not an independent data set, and the effect of WTD observations is very limited at this small watershed when discharge and SWC observations are both assimilated.

[63] The test cases QST and SSHCZO show that both T_{sfc} and surface heat fluxes are good indicators of land surface states. Assimilation of either T_{sfc} or surface heat fluxes

is sufficient for land surface parameter estimation. The test case QST also demonstrates that Q, SWC, and T_{sfc} are the essential observations for the estimation of those six model parameters at the Shale Hills watershed. The SSHCZO test case assimilates observations which are currently available at the Shale Hills watershed. The results are very encouraging. It indicates that using the currently available observation data sets for real data EnKF parameter estimation is promising.

[64] There are several test cases that do not assimilate E_t observations: the test cases Q, SSHCZO, and QST. Results from those test cases show that as long as T_{sfc} or surface heat fluxes are assimilated into the system, the system is able to obtain model parameters that could provide reasonably good E_t prediction (Figure 11g). Therefore, the measurement of E_t is not critical for model calibration purpose.

[65] Results from different test cases imply that the assimilation of multivariate observations improves the accuracy of parameter estimation, and provides unique parameter solutions. For example, when SWC observations are not assimilated into the system (NoSWC and Q test cases), EnKF cannot provide accurate estimates of the van Genuchten parameters (Table 5), although both test cases provide good discharge predictions (Figure 11a). More interestingly, the α and β values estimated for those two test cases (NoSWC and Q test cases) are very close (Table 5). From the equifinality [*Beven*, 1993] perspective, when SWC is not assimilated, EnKF finds another point in the α-β space, which produces almost equally good discharge predictions as the true parameter values. Only when SWC observations are assimilated, can EnKF find the unique and accurate solutions of α and β. The assimilation of multivariate observations can apply more constraints to model parameters, avoid the difficulty brought by model equifinality, and provide unique parameter solutions. By testing the influences of assimilated observations in synthetic experiments, the required observations to identify the unique parameter set can be found. The data assimilation system developed in this paper can thus be used to provide guidance of observational system designs.

[66] The EnKF provides the estimates of not only parameter values and model states, but also their uncertainties. When more types of observations are assimilated into the system (e.g., the control run), the uncertainties estimated by EnKF decrease faster. When only limited types of observations are assimilated into the system (e.g., the Q, SSHCZO, and QST test cases), the uncertainties estimated by EnKF decrease slowly, and some of the parameters do not even converge (Table 5). The uncertainties by EnKF result from various sources, e.g., the uncertainties of observations and model parameters. The quantification of uncertainties is very useful for practical application, because the accurate estimates of uncertainty is required in the operational flood and drought forecasting. This EnKF data assimilation system provides the possibility to perform real-time online probabilistic forecasting using a deterministic model, which explicitly accounts for uncertainties from different sources (e.g., parameter, model structure, meteorological forcing, and assimilated observations).

[67] It needs to be pointed out that those results are based on a perfect model, perfect forcing data, and a perfect model domain configuration. Model structural errors, forcing data errors, observation errors, and domain configuration errors (e.g., errors in input topography, soil map, and land cover map) would pose extra difficulties for parameter estimation using real data. The synthetic observations used in this study have Gaussian errors with no biases, and the synthetic errors only include the random instrumental errors. In reality, some observations (e.g., discharge and surface heat fluxes) may have non-Gaussian errors, the MODIS land surface temperature observations may have systematic biases [*Wan et al*., 2002; *Wang et al*., 2008], and the eddy covariance measurements fail to close the energy budget [*McNeil and Shuttleworth*, 1975; *Fritschen et al*., 1992; *Twine et al*., 2000]. Moreover, the soil moisture and water table depth observations may have representativity errors in addition to the instrumental errors. Those errors are not accounted for in this study, and need to be taken into account for real data applications. Different approaches have been used to assimilate observations which have consistent biases, e.g., rescaling the observations, subtracting the long-term means from observations and predictions, or assimilating the tendencies of observations instead of their absolute values [*Hain et al*., 2012; *Mackaro et al*., 2012]. The impact of non-Gaussian, non-zero-bias observations on Flux-PIHM parameter estimation, however, still needs to be tested.

[68] Although the initial parameters are perturbed independently, EnKF is able to identify the interacting parameters (Table 6). Results from the control run, Case+ and Case− reveal strong correlation between the van Genuchten parameters α and β (Table 6). The negative correlation found between α and β in these three test cases agree with the results in other test cases. As shown in Table 6, whenever α is underestimated, β is always overestimated, and vice versa. The strong correlation found between the van Genuchten parameters suggests strong interactions between those parameters. The interaction leads to model equifinality [*Beven*, 1993], and explains why the test cases Q and NoSWC provide poor estimates of α and β but acceptable discharge predictions. *Baldwin* [2011] derived the van Genuchten parameters at 61 sites in the Shale Hills watershed based on soil moisture observations at different depths, and analyzed the spatial relationship between the van Genuchten parameters. The results showed that the correlation coefficient between the van Genuchten parameters from all depths and locations is -0.28. The correlation coefficients between van Genuchten parameters at 10 cm and 20 cm below ground reach -0.44 and -0.48, respectively. The correlation between the van Genuchten parameters estimated by EnKF agrees well with *Baldwin* [2011] observational results.

[69] The relatively high correlation between the van Genuchten parameter α and the minimum stomatal resistance $R_{c\ \min}$ (Table 6) suggests interaction between land surface and subsurface. The model sensitivity analysis shows that the model soil moisture prediction is very sensitive to the change in the parameter α [*Shi et al*., 2013b], and soil moisture affects transpiration, which is also influenced by $R_{c\ \min}$. The parameters α and $R_{c\ \min}$ are therefore connected through the effect of soil moisture on transpiration. The correlations between the hydrologic and land surface parameters represent the interaction between the subsurface and land surface, which suggests that subsurface and land surface systems are closely coupled.

[70] The correlations between different parameters revealed by the EnKF system are useful for the study of the interaction between dynamic systems, and are useful for the simplification of model parameterization schemes. The ability of EnKF to identify the interacting parameters and quantify the correlations between parameters suggests that it may not be necessary to take into account the correlation between parameters when generating the initial ensemble. This is valuable because prior information describing parameter correlation is frequently not available.

[71] **Acknowledgments.** This research was supported by the National Oceanic and Atmospheric Administration through grant NA10OAR431 0166, and the National Science Foundation Shale Hills-Susquehanna Critical Zone Observatory project through grant EAR 0725019. Logistical support and/or data were provided by the NSF-supported Shale Hills Susquehanna Critical Zone Observatory.

References

Aksoy, A., F. Zhang, and J. W. Nielsen-Gammon (2006), Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model, *Mon. Weather Rev.*, *134*(10), 2951–2970, doi:10.1175/MWR3224.1.

Anderson, J. L., and S. L. Anderson (1999), A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Mon. Weather Rev.*, *127*(12), 2741–2758, doi:10.1175/1520-0493(1999)127¡2741:AMCIOT¿2.0.CO;2.

Anderton, S., J. Latron, and F. Gallart (2002), Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model, *Hydrol. Processes*, *16*(2), 333–353, doi:10.1002/hyp.336.

Annan, J. D. (2005), Parameter estimation using chaotic time series, *Tellus, Ser. A*, *57*(5), 709–714, doi:10.1111/j.1600-0870.2005.00143.x.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.*, *50*(2), 174–188, doi:10.1109/78.978374.

Aubert, D., C. Loumagne, and L. Oudin (2003), Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, *J. Hydrol.*, *280*(1), 145–161, doi:10.1016/S0022-1694(03)00229-4.

Bailey, R., and D. Baù (2010), Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution, *Water Resour. Res.*, *46*, W12543, doi:10.1029/2010WR009147.

Baldocchi, D., R. Valentini, S. Running, W. Oechel, and R. Dahlman (1996), Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems, *Global Change Biol.*, *2*(3), 159–168, doi:10.1111/j.1365-2486.1996.tb00069.x.

Baldocchi, D. D. (2003), Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future, *Global Change Biol.*, *9*(4), 479–492, doi:10.1046/j.1365-2486.2003.00629.x.

Baldwin, D. (2011), Catchment-scale soil water retention characteristics and delineation of hydropedological functional units in the Shale Hills Catchment, MS thesis, Pa. State Univ., University Park.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*(1), 41–51, doi:10.1016/0309-1708(93)90028-E.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, *36*(12), 3663–3674, doi:10.1029/2000WR900207.

Bras, R. L. (1990), *Hydrology: An Introduction to Hydrologic Science*, 643 pp., Addison-Wesley, Reading, Mass.

Cammalleri, C., and G. Ciraolo (2012), State and parameters update in a coupled energy/hydrologic balance model using ensemble Kalman filtering, *J. Hydrol.*, *416–417*, 171–181, doi:10.1016/j.jhydrol.2011.11.049.

Campbell Scientific Inc. (2007), *CS420-L and CS425-L Druck's Models PDCR 1830 and 1230 pressure transducers, Instr. Manual 12/07*, Palo Alto, Calif.

Camporese, M., C. Paniconi, M. Putti, and P. Salandin (2009a), Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow, *Water Resour. Res.*, *45*, W10421, doi:10.1029/2008WR007031.

Camporese, M., C. Paniconi, M. Putti, and P. Salandin (2009b), Comparison of data assimilation techniques for a coupled model of surface and subsurface flow, *Vadose Zone J.*, *8*(4), 837–845, doi:10.2136/vzj2009.0018.

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*(4), 569–585, doi:10.1175/1520-0493(2001)129¡0569:CAALSH¿2.0.CO;2.

Chen, F., Z. Janjić, and K. Mitchell (1997), Impact of atmospheric-surface layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta numerical model, *Boundary Layer Meteorol.*, *85*(3), 391–421, doi:10.1023/A:1000531001463.

Chen, Y., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv. Water Resour.*, *29*(8), 1107–1122, doi:10.1016/j.advwatres.2005.09.007.

Clark, M. P., D. E. Rupp, R. A. Woods, X. Zheng, R. P. Ibbitt, A. G. Slater, J. Schmidt, and M. J. Uddstrom (2008), Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Adv. Water Resour.*, *31*(10), 1309–1324, doi:10.1016/j.advwatres.2008.06.005.

Coll, C., Z. Wan, and J. M. Galve (2009), Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product, *J. Geophys. Res.*, *114*, D20102, doi:10.1029/2009JD012038.

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.

Czarnomski, N. M., G. W. Moore, T. G. Pypker, J. Licata, and B. J. Bond (2005), Precision and accuracy of three alternative instruments for measuring soil water content in two forest soils of the Pacific Northwest, *Can. J. For. Res.*, *35*(8), 1867–1876, doi:10.1139/X05-121.

Dechmi, F., J. Burguete, and A. Skhiri (2012), SWAT application in intensive irrigation systems: Model modification, calibration and validation, *J. Hydrol.*, *470–471*, 227–238, doi:10.1016/j.jhydrol.2012.08.055.

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, doi:10.1029/91WR02985.

Eckhardt, K., and J. G. Arnold (2001), Automatic calibration of a distributed catchment model, *J. Hydrol.*, *251*(1), 103–109, doi:10.1016/S0022-1694(01)00429-2.

Ek, M. B., K. Mitchell, Y. Lin, E. Rogers, P. Grummann, V. Koren, G. Gayno, and J. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational Mesoscale Eta Model, *J. Geophys. Res.*, *108*(D22), 8851, doi:10.1029/2002JD003296.

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. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*(4), 343–367, doi:10.1007/s10236-003-0036-9.

Finkelstein, P. L., and P. F. Sims (2001), Sampling error in eddy correlation flux measurements, *J. Geophys. Res.*, *106*(D4), 3503–3509, doi:10.1029/2000JD900731.

Flores, A. N., R. L. Bras, and D. Entekhabi (2012), Hydrologic data assimilation with a hillslope-scale-resolving model and L band radar observations: Synthetic experiments with the ensemble Kalman filter, *Water Resour. Res.*, *48*, W08509, doi:10.1029/2011WR011500.

Franchini, M. (1996), Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models, *Hydrol. Sci. J.*, *41*(1), 21–39, doi:10.1080/02626669609491476.

Francois, C., A. Quesney, and C. Ottlé (2003), Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter, *J. Hydrometeorol.*, *4*(2), 473–487, doi:10.1175/1525–7541(2003)4¡473:SAOESD¿2.0.CO;2.

Fritschen, L. J., P. Qian, E. T. Kanemasu, D. Nie, E. A. Smith, J. B. Stewart, S. B. Verma, and M. L. Wesely (1992), Comparisons of surface flux measurement systems used in FIFE 1989, *J. Geophys. Res.*, *97*(D17), 18,697–18,713, doi:10.1029/91JD03042.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z.-L. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,503, doi:10.1029/1999JD900154.

Gupta, V. K., and S. Sorooshian (1985), The relationship between data and the precision of parameter estimates of hydrologic models, *J. Hydrol.*, *81*, 57–77, doi:10.1016/0022-1694(85)90167-2.

Hain, C. R., W. T. Crow, M. C. Anderson, and J. R. Mecikalski (2012), An ensemble Kalman filter dual assimilation of thermal infrared and microwave satellite observations of soil moisture into the Noah land surface model, *Water Resour. Res.*, *48*, W11517, doi:10.1029/2011WR011268.

Han, E., V. Merwade, and G. C. Heathman (2012), Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model, *J. Hydrol.*, *416*, 98–117, doi:10.1016/j.jhydrol.2011.11.039.

Hendricks Franssen, H. J., H. P. Kaiser, U. Kuhlmann, G. Bauser, F. Stauffer, R. Müller, and W. Kinzelbach (2011), Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating, *Water Resour. Res.*, *47*, W02532, doi:10.1029/2010WR009480.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich (1998), Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resour. Res.*, *34*(12), 3405–3420, doi:10.1029/1998WR900001.

Hu, X.-M., F. Zhang, and J. W. Nielsen-Gammon (2010), Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study, *Geophys. Res. Lett.*, *37*, L08802, doi:10.1029/2010GL043017.

Ibbitt, R. P. (1970), Systematic parameter fitting for conceptual models of catchment hydrology, Ph.D. thesis, Univ. of London, London.

Jackson, C., Y. Xia, M. K. Sen, and P. L. Stoffa (2003), Optimal parameter and uncertainty estimation of a land surface model: A case study using data from Cabauw, Netherlands, *J. Geophys. Res.*, *108*(D18), 4583, doi:10.1029/2002JD002991.

Johnston, P. R., and D. H. Pilgrim (1976), Parameter optimization for watershed models, *Water Resour. Res.*, *12*(3), 477–486, doi:10.1029/WR012i003p00477.

Kollat, J. B., and P. M. Reed (2006), Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design, *Adv. Water Resour.*, *29*(6), 792–807, doi:10.1016/j.advwatres.2005.07.010.

Kumar, M. (2009), Toward a hydrologic modeling system, PhD thesis, Pa. State Univ., University Park.

Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B. Eylander, and P. R. Houser (2008), A land surface data assimilation framework using the land information system: Description and applications, *Adv. Water Resour.*, *31*(11), 1419–1432, doi:10.1016/j.advwatres.2008.01.013.

Kurtz, W., H.-J. Hendricks Franssen, and H. Vereecken (2012), Identification of time-variant river bed properties with the ensemble Kalman filter, *Water Resour. Res.*, *48*, W10534, doi:10.1029/2011WR011743.

Lee, H., D. J. Seo, and V. Koren (2011), Assimilation of streamflow and in-situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states, *Adv. Water Resour.*, *34*(12), 1597–1615, doi:10.1016/j.advwatres.2011.08.012.

Leimer, S., T. Pohlert, S. Pfahl, and W. Wilcke (2011), Towards a new generation of high-resolution meteorological input data for small-scale hydrologic modeling, *J. Hydrol.*, *402*(3), 317–332, doi:10.1016/j.jhydrol.2011.03.026.

Lenschow, D. H., and B. B. Stankov (1986), Length scales in the convective boundary layer, *J. Atmos. Sci.*, *43*(12), 1198–1209, doi:10.1175/1520-0469(1986)043¡1198:LSITCB¿2.0.CO;2.

Lenschow, D. H., J. Mann, and L. Kristensen (1994), How long is long enough when measuring fluxes and other turbulence statistics?, *J. Atmos. Oceanic Technol.*, *11*(3), 661–673, doi:10.1175/1520-0426(1994)011¡0661:HLILEW¿2.0.CO;2.

Liu, G., Y. Chen, and D. Zhang (2008), Investigation of flow and transport processes at the MADE site using ensemble Kalman filter, *Adv. Water Resour.*, *31*(7), 975–986, doi:10.1016/j.advwatres.2008.03.006.

Lü, H., T. Hou, R. Horton, Y. Zhu, X. Chen, Y. Jia, W. Wang, and X. Fu (2013), The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method, *J. Hydrol.*, *480*, 102–114, doi:10.1016/j.jhydrol.2012.12.011.

Mackaro, S. M., R. T. McNider, and A. P. Biazar (2012), Some physical and computational issues in land surface data assimilation of satellite skin temperatures, *Pure Appl. Geophys.*, *169*, 401–414, doi:10.1007/s00024-011-0377-0.

Mahrt, L. (2000), Surface heterogeneity and vertical structure of the boundary layer, *Boundary Layer Meteorol.*, *96*(1–2), 33–62, doi:10.1023/A:1002482332477.

McNeil, D. D., and W. J. Shuttleworth (1975), Comparative measurements of the energy fluxes over a pine forest, *Boundary Layer Meteorol.*, *9*(3), 297–313, doi:10.1007/BF00230772.

Montzka, C., V. R. N. Pauwels, H.-J. Hendricks Franssen, X. Han, and H. Vereecken (2012), Multivariate and multiscale data assimilation in terrestrial systems: A review, *Sensors*, *12*(12), 16,291–16,333, doi:10.3390/s121216291.

Moradkhani, H., and S. Sorooshian (2008), General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis, in *Hydrological Modelling and the Water Cycle*, edited by S. Sorooshian et al., pp. 1–24, Springer, Heidelberg.

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.

Moran, M. S., C. D. Peters-Lidard, J. M. Watts, and S. McElroy (2004), Estimating soil moisture at the watershed scale with satellite-based radar and land surface models, *Can. J. Remote Sens.*, *30*(5), 805–826, doi:10.5589/m04-043.

Nielsen-Gammon, J. W., X. M. Hu, F. Zhang, and J. E. Pleim (2010), Evaluation of planetary boundary layer scheme sensitivities for the purpose of parameter estimation, *Mon. Weather Rev.*, *138*(9), 3400–3417, doi:10.1175/2010MWR3292.1.

Nutter, W. L. (1964), Determination of the head-discharge relationship for a sharp-crested compound weir and a sharp-crested parabolic weir, MS thesis, Pa. State Univ., University Park.

Oudin, L., A. Weisse, C. Loumagne, and S. Le Hégarat-Mascle (2003), Assimilation of soil moisture into hydrological models for flood forecasting: A variational approach, *Can. J. Remote Sens.*, *29*(6), 679–686, doi:10.5589/m03-038.

Pan, M., and E. F. Wood (2006), Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter, *J. Hydrometeorol.*, *7*(3), 534–547, doi:10.1175/JHM495.1.

Pauwels, V. R. N., and G. J. M. De Lannoy (2006), Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge, *J. Hydrometeorol.*, *7*(3), 458–477, doi:10.1175/JHM490.1.

Pickup, G. (1977), Testing the efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models, *Hydrol. Sci. Bull.*, *22*(2), 257–274, doi:10.1080/02626667709491716.

Pisinaras, V., C. Petalas, G. D. Gikas, A. Gemitzi, and V. A. Tsihrintzis (2010), Hydrological and water quality modeling in a medium-sized basin using the Soil and Water Assessment Tool (SWAT), *Desalination*, *250*(1), 274–286, doi:10.1016/j.desal.2009.09.044.

Pokhrel, P., and H. V. Gupta (2010), On the use of spatial regularization strategies to improve calibration of distributed watershed models, *Water Resour. Res.*, *46*, W01505, doi:10.1029/2009WR008066.

Qu, Y. (2004), An integrated hydrologic model for multi-process simulation using semi-discrete finite volume approach, PhD thesis, Pa. State Univ., University Park.

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.

Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and DMIP Participants (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, *298*(1), 27–60, doi:10.1016/j.jhydrol.2004.03.031.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114, doi:10.1175/1520-0493(2002)130¡0103:HDAWTE¿2.0.CO;2.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*(6), 728–740, doi:10.1175/1525–7541(2002)003¡0728:EVEKFF¿2.0.CO;2.

Richardson, A. D., et al. (2006), A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agric. For. Meteorol.*, *136*(1), 1–18, doi:10.1016/j.agrformet.2006.01.007.

Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, *376*(3), 428–442, doi:10.1016/j.jhydrol.2009.07.051.

Salesky, S. T., M. Chamecki, and N. L. Dias (2012), Estimating the random error in eddy-covariance based fluxes and other turbulence statistics: The filtering method, *Boundary Layer Meteorol.*, *144*(1), 113–135, doi:10.1007/s10546-012-9710-0.

Shi, P., C. Chen, R. Srinivasan, X. Zhang, T. Cai, X. Fang, S. Qu, X. Chen, and Q. Li (2011), Evaluating the SWAT model for hydrological modeling in the Xixian watershed and a comparison with the XAJ model, *Water Resour. Manage.*, *25*(10), 2595–2612, doi:10.1007/s11269-011-9828-8.

Shi, Y., K. J. Davis, C. J. Duffy, and X. Yu (2013a), Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory, *J. Hydrometeorol.*, *14*(5), 1401–1420, doi:10.1175/JHM-D-12-0145.1.

Shi, Y., K. J. Davis, F. Zhang, and C. J. Duffy (2013b), Evaluation of the parameter sensitivity of a coupled land surface hydrologic model at a critical zone observatory, *J. Hydrometeorol.*, doi:10.1175/JHM-D-12-0177.1, in press.

Shih, D.-S., and G.-T. Yeh (2011), Identified model parameterization, calibration, and validation of the physically distributed hydrological model WASH123D in Taiwan, *J. Hydrol. Eng.*, *16*(2), 126–136, doi:10.1061/(ASCE)HE.1943-5584.0000293.

Snyder, C., and F. Zhang (2003), Assimilation of simulated Doppler radar observations with an ensemble Kalman filter, *Mon. Weather Rev.*, *131*(8), 1663–1677, doi:10.1175//2555.1.

Sorooshian, S., Q. Duan, and V. K. Gupta (1993), Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, *Water Resour. Res.*, *29*(4), 1185–1194, doi:10.1029/92WR02617.

Tang, Y., P. Reed, and T. Wagener (2006), How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrol. Earth Syst. Sci.*, *10*, 289–307.

Twine, T. E., W. P. Kustas, J. M. Norman, D. R. Cook, P. R. Houser, T. P. Meyers, J. H. Prueger, P. J. Starks, and M. L. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agric. For. Meteorol.*, *103*(3), 279–300, doi:10.1016/S0168-1923(00)00123-4.

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.

Vrugt, J., H. V. Gupta, B. Ó. Nualláin, and W. Bouten (2006), Real-time data assimilation for operational ensemble streamflow forecasting, *J. Hydrometeorol.*, *7*(3), 548–565, doi:10.1175/JHM504.1.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Wallner, M., et al. (2012), Evaluation of different calibration strategies for large scale continuous hydrological modelling, *Adv. Geosci.*, *31*, 67–74, doi:10.5194/adgeo-31-67-2012.

Wan, Z., and Z.-L. Li (1997), A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data, *IEEE Trans. Geosci. Remote Sens.*, *35*(4), 980–996, doi:10.1109/36.602541.

Wan, Z., Y. Zhang, Q. Zhang, and Z. Li (2002), Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data, *Remote Sens. Environ.*, *83*(1), 163–180, doi:10.1016/S0034-4257(02)00093-7.

Wang, K., and S. Liang (2009), Evaluation of ASTER and MODIS land surface temperature and emissivity products using long-term surface long-wave radiation observations at SURFRAD sites, *Remote Sens. Environ.*, *113*(7), 1556–1565, doi:10.1016/j.rse.2009.03.009.

Wang, W., S. Liang, and T. Meyers (2008), Validating MODIS land surface temperature products using long-term nighttime ground measurements, *Remote Sens. Environ.*, *112*(3), 623–635, doi:10.1016/j.rse.2007.05.024.

Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resour. Res.*, *42*, W09403, doi:10.1029/2005WR004093.

Xia, Y., A. J. Pitman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas (2002), Calibrating a land surface model of varying complexity using multicriteria methods and the Cabauw dataset, *J. Hydrometeorol.*, *3*(2), 181–194, doi:10.1175/1525–7541(2002)003¡0181:CALSMO¿2.0.CO;2.

Xie, X., and D. Zhang (2010), Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter, *Adv. Water Resour.*, *33*(6), 678–690, doi:10.1016/j.advwatres.2010.03.012.

Yao, C., Z. Li, Z. Yu, and K. Zhang (2012), A priori parameter estimates for a distributed, grid-based Xinanjiang model using geographically based information, *J. Hydrol.*, *468–469*, 47–62, doi:10.1016/j.jhydrol.2012.08.025.

Yu, X., G. Bhatt, C. Duffy, and Y. Shi (2013), Parameterization for distributed watershed modeling using national data and evolutionary algorithm, *Comput. Geosci.*, *58*, 80–90, doi:10.1016/j.cageo.2013.04.025.

Yu, Y., J. L. Privette, and A. C. Pinheiro (2008), Evaluation of split-window land surface temperature algorithms for generating climate data records, *IEEE Trans. Geosci. Remote Sens.*, *46*(1), 179–192, doi:10.1109/TGRS.2007.909097.

Zhang, F., C. Snyder, and J. Sun (2004), Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter, *Mon. Weather Rev.*, *132*(5), 1238–1253, doi:10.1175/1520-0493(2004)132¡1238:IOIEAO¿2.0.CO;2.

Zhang, F., Z. Meng, and A. Aksoy (2006), Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation: Part I: Perfect model experiments, *Mon. Weather Rev.*, *134*(2), 722–736, doi:10.1175/MWR3101.1.

Zilitinkevich, S. S. (1995), Non-local turbulent transport: Pollution dispersion aspects of coherent structure of convective flows, in *Air Pollution III—Volume I Air Pollution Theory and Simulation*, edited by H. Power, N. Moussiopoulos, and C. A. Brebbia, pp. 53–60, Comput. Mech., Southampton, U. K.

Zupanski, D., and M. Zupanski (2006), Model error estimation employing an ensemble data assimilation approach, *Mon. Weather Rev.*, *134*(5), 1337–1354, doi:10.1175/MWR3125.1.