1 Parameter estimation of a physically-based land surface
2 hydrologic model using an ensemble Kalman filter:
3 A multivariate real-data experiment

4 Yuning Shi^a, Kenneth J Davis^{a,b}, Fuqing Zhang^b, Christopher J Duffy^c,
5 Xuan Yu^c

6 ^a*Earth and Environmental Systems Institute, The Pennsylvania State University,*
7 *University Park, Pennsylvania, USA*

8 ^b*Department of Meteorology, The Pennsylvania State University, University Park,*
9 *Pennsylvania, USA*

10 ^c*Department of Civil and Environmental Engineering, The Pennsylvania State*
11 *University, University Park, Pennsylvania, USA*

12 **Abstract**

13 The capability of an ensemble Kalman filter (EnKF) to simultaneously esti-
14 mate multiple parameters in a physically-based land surface hydrologic model
15 using multivariate field observations is tested at a small watershed (0.08 km²).
16 Multivariate, high temporal resolution, *in situ* measurements of discharge,
17 water table depth, soil moisture, and sensible and latent heat fluxes encom-
18 passing five months of 2009 are assimilated. It is found that, for five out
19 of the six parameters, the EnKF estimated parameter values from different
20 test cases converge strongly, and the estimates after convergence are close to
21 the manually calibrated parameter values. The EnKF estimated parameters
22 and manually calibrated parameters yield similar model performance, but
23 the EnKF sequential method significantly decreases the time and labor re-
24 quired for calibration. The results demonstrate that, given a limited number
25 of multi-state, site-specific observations, an automated sequential calibra-
26 tion method (EnKF) can be used to optimize physically-based land surface
27 hydrologic models.

28 *Keywords:* Data assimilation, parameter estimation, land surface model,
29 hydrologic model, ensemble Kalman filter

1. Introduction

Uncertainties in model parameters are a dominant source of uncertainty for hydrologic models (Moradkhani and Sorooshian, 2008). The ensemble Kalman filter (EnKF) (Evensen, 1994) provides a promising approach for the automated calibration of hydrologic models (e.g., Moradkhani et al., 2005; Xie and Zhang, 2010; Lü et al., 2013; Shi et al., 2014b). Most previous studies applied EnKF to conceptual or process-based hydrologic models. Shi et al. (2014b) performed a multiple-parameter estimation for a physically-based land surface hydrologic model, Flux-PIHM (Shi et al., 2013), via EnKF and assimilating multivariate synthetic observations including discharge, water table depth, soil moisture, land surface temperature, sensible and latent heat fluxes, and transpiration. The modeling and data assimilation system was implemented at the Shale Hills watershed (0.08 km²) in central Pennsylvania, the site of the Susquehanna/Shale Hills Critical Zone Observatory (SSHCZO). Results from the synthetic data experiments indicated that EnKF is capable of providing accurate estimation of multiple Flux-PIHM model parameters, and the assimilation of multivariate observations including those currently available at the SSHCZO applied strong constraints to model parameters.

Real-data experiments, however, have notable difficulties that do not exist with synthetic data experiments, because the errors in model predictions expand to include errors from forcing data, domain configuration, observation bias, and model structure. When EnKF is used to estimate parameter values, over-adjustment may occur, which may cause large changes in parameter values and parameter uncertainties, and lead to system “shocks”, when the dynamic balance of model system is destroyed and the model attempts to restore the dynamic balance (Houser et al., 2010).

The goal of this research effort is to test the ability of the EnKF system to estimate multiple parameters in Flux-PIHM with the assimilation of real multivariate observations at a field site with co-located measurements. Extensive and detailed field site characterization along with a broad array of observations is available at the SSHCZO. This study site thus provides an unprecedented opportunity for real-data assimilation experiment. We test the EnKF system’s ability to estimate Flux-PIHM model parameters with SSHCZO observations. Model performances with the EnKF-estimated parameter values and manually calibrated values are compared to assess the quality of the EnKF-estimated parameter values. In addition, we test the

67 performance of the data assimilation system when driven by atmospheric re-
68 analysis and remotely-sensed forcing data, to evaluate the ability of the data
69 assimilation method to adapt to commonly available continental-scale driver
70 data.

71 **2. Flux-PIHM EnKF system**

72 Flux-PIHM (Shi et al., 2013) is a coupled land surface hydrologic model.
73 Flux-PIHM incorporates a land surface scheme into the Penn State Inte-
74 grated Hydrologic Model (PIHM) (Qu, 2004; Qu and Duffy, 2007; Kumar,
75 2009), which is a fully-coupled, physically-based, spatially-distributed hy-
76 drologic model. The land surface scheme in Flux-PIHM is adapted from the
77 Noah land surface model (LSM) (Chen and Dudhia, 2001; Ek et al., 2003).
78 The land surface and hydrologic components are coupled by exchanging wa-
79 ter table depth, infiltration rate, recharge rate, net precipitation rate, and
80 evapotranspiration rate between the two model components.

81 A Flux-PIHM data assimilation system has been developed by incorporat-
82 ing EnKF for model parameter and state estimation (Shi et al., 2014b) using
83 the EnKF formulation from Snyder and Zhang (2003). In the Flux-PIHM
84 EnKF system, the Flux-PIHM model variables and the global calibration co-
85 efficients of model parameters are concatenated into a joint state parameter
86 vector $\mathbf{x}$, and are updated simultaneously by EnKF using the state augmen-
87 tation approach (Annan, 2005; Aksoy et al., 2006; Liu and Gupta, 2007; Hu
88 et al., 2010; Xie and Zhang, 2010). The global calibration coefficient (Pokhrel
89 and Gupta, 2010; Wallner et al., 2012; Shi et al., 2013) is a scalar multiplier
90 applied to the corresponding soil or vegetation related parameter for all soil
91 or vegetation types, and is used to decrease the dimension of the joint state
92 parameter vector. The covariance relaxation method of Zhang et al. (2004,
93 their Eq. (5)) is applied on model parameters and variables in order to avoid
94 filter divergence (Anderson and Anderson, 1999). In addition, the condi-
95 tional covariance inflation method (Aksoy et al., 2006) is applied to model
96 parameters. A quality control process (Shi et al., 2014b) is performed after
97 each EnKF analysis step to ensure the parameters and state variables remain
98 within physically realistic or plausible ranges. Please see Shi et al. (2013)
99 and Shi et al. (2014b) for detailed descriptions.

3. Experimental setup

The Flux-PIHM EnKF data assimilation system is implemented at the Shale Hills watershed (0.08 km²) in central Pennsylvania. The Shale Hills watershed is a small-scale, forested, V-shaped catchment characterized by relatively steep slopes and narrow ridges. The SSHCZO exists in this watershed. A real-time hydrologic monitoring network (RTHnet) is operating in the SSHCZO, which provides real-time and high-frequency observations from bedrock to the atmospheric boundary layer.

The Shale Hills watershed model domain is decomposed into 535 triangular grids and 20 river segments, with an average grid size of 157 m². There are five soil types and three vegetation types in the model domain. The grid configuration, vegetation map, soil map, meteorological forcing, and *a priori* input data are the same as in Shi et al. (2013). Given the small scale (0.08 km²) of the watershed, spatially uniform forcing is used. The meteorological forcing (precipitation, air temperature, relative humidity, downward longwave and solar radiation, wind speed, and surface air pressure) data are obtained from the RTHnet weather station and the surface radiation budget network (SURFRAD) Penn State University station. The moderate resolution imaging spectroradiometer (MODIS) 8-day leaf area index (LAI) data (Knyazikhin et al., 1999; Myneni et al., 2002) are rescaled based on the comparison between the MODIS product and the CZO field measurements to drive the model (Shi et al., 2013). The parameters to be estimated are: effective porosity Θ_e , van Genuchten (1980) soil parameters α and β , Zilitinkevich (1995) parameter C_{zil} , minimum stomatal resistance R_{cmin} , and reference canopy water capacity S . The estimation of those parameters has been tested in synthetic experiments (Shi et al., 2014b). The physically plausible ranges of the calibration coefficients are presented in Table 1. Detailed descriptions and *a priori* values of those parameters can be found in Shi et al. (2013, 2014a).

A total of 30 ensemble members are used for each test case. The ensemble members are generated by randomly perturbing the calibration coefficients of those six parameters within their plausible ranges (Table 1). The parameters that are not estimated are set to their manually calibrated values as in Shi et al. (2013). The manual calibration was performed using the “trial and error” strategy, using outlet discharge, water table depth, soil water content, soil temperature, and surface heat flux data from June to July 2009 to optimize model parameters (Shi et al., 2013). For each parameter (calibration

Table 1: Flux-PIHM model parameters, their plausible ranges of calibration coefficients, estimates from different test cases, and manual calibration values (Shi et al., 2013). The test cases are 1: Case0, 2: Case+, 3: Case−, 4: NLDAS, 5: MODIS, and 6: NLDAS+MODIS.

Parameter	Description	Range of calibration coefficient	Test cases						
			1	2	3	4	5	6	Manual
Θ_e	Effective porosity ($\text{m}^3 \text{ m}^{-3}$)	0.3–1.2	0.62	0.67	0.65	0.60	0.63	0.61	0.52
α	Van Genuchten soil parameter (m^{-1})	0–2.5	1.50	1.57	1.49	1.31	1.38	1.33	1.50
β	Van Genuchten soil parameter (-)	0.95–2.5	1.34	1.29	1.34	1.40	1.35	1.37	1.30
$R_{c\min}$	Minimum stomatal resistance (s m^{-1})	0.3–1.2	0.41	0.49	0.43	0.48	0.63	0.65	0.50
S	Reference canopy water storage (mm)	0–5	3.15	4.53	1.13	3.80	3.45	0.55	2.00
C_{zil}	Zilitinkevich parameter (-)	0.1–10	1.15	1.09	1.23	0.81	1.32	0.93	0.70

coefficient) ϕ , the values are randomly drawn from a Gaussian distribution, with an initial standard deviation of $\sigma_0 = 0.2 (\phi_{\max} - \phi_{\min})$, where $\phi_{\max}$ and $\phi_{\min}$ represent the upper and lower boundaries of the plausible range, respectively. Among those parameters, C_{zil} is perturbed in log space. Shi et al. (2014b) showed that EnKF is capable of identifying the interacting parameters and quantifying the correlations between parameters, without the need of *a priori* parameter correlation information. We thus perturb the parameters such that the initial correlation coefficient (the absolute value) between any two of those parameters is less than or equal to 0.25, to avoid artificially high correlations between parameters and observable variables.

All ensemble members start from 0000 UTC 1 January 2009, from saturation in the relaxation mode (Shi et al., 2013). The model time step of 1 min and the output interval is 1 h. The first set of observations is assimilated at 1700 UTC 4 April 2009. The calibration period is from 4 April to 1 September 2009. Shi et al. (2014b) found that the assimilation interval for synthetic experiments at the Shale Hills watershed should be larger than 72 hours to avoid system “shocks” caused by EnKF updates. In real-data experiments, however, we found that the system shocks are often larger than with synthetic data, probably due to additional errors such as model structural errors. Thus we set the assimilation interval to 168 hours to avoid any potential shocks to the system. The time for assimilating the first set of observations is chosen to include the discharge peak on 20 June 2009 considering the assimilation interval.

Six test cases, Case0, Case+, Case−, NLDAS, MODIS, and NLDAS+MODIS are executed. The test cases Case0, Case+, and Case− have different initial guesses of parameter values. For Case0, the initial ensemble means of parameters are set to the center of the physically plausible range, i.e., $0.5 (\phi_{\max} - \phi_{\min})$. For Case+ and Case−, the initial ensemble means of parameters are set to $0.5 (\phi_{\max} - \phi_{\min}) + \sigma_0$ and $0.5 (\phi_{\max} - \phi_{\min}) - \sigma_0$, respectively. These three test cases are driven by locally-measured meteorological forcing and rescaled MODIS LAI data. The test cases NLDAS, MODIS, and NLDAS+MODIS have the same initial ensemble members as Case0. The test case NLDAS is driven by the forcing data for Phase 2 of the North American Land Data Assimilation System (NLDAS-2) (Cosgrove et al., 2003; Xia et al., 2012) and rescaled MODIS LAI data; the test case MODIS is driven by the locally-measured atmospheric forcing and unchanged MODIS LAI data; and the test case NLDAS+MODIS is driven by the NLDAS-2 forcing and unchanged MODIS LAI. The hourly NLDAS-2 forcing data are bilinearly

Table 2: Forcing data used for different test cases.

Test Cases	Meteorological forcing	LAI forcing
Case0	Locally measured	Rescaled MODIS LAI
Case+	Locally measured	Rescaled MODIS LAI
Case−	Locally measured	Rescaled MODIS LAI
NLDAS	NLDAS-2 forcing	Rescaled MODIS LAI
MODIS	Locally measured	Unchanged MODIS LAI
NLDAS+MODIS	NLDAS-2 forcing	Unchanged MODIS LAI

interpolated from $1/8^\circ$ resolution to the location of SSHCZO. These three test cases are used to test the robustness of the data assimilation system when driven by reanalysis and remotely-sensed forcing data. The forcing data used for each test case are listed in Table 2. Figure 1a shows the difference of precipitation between RTHnet observation and NLDAS-2 forcing, and the difference of LAI between MODIS product and rescaled LAI.

4. Assimilated observations and observation errors

The observations assimilated into the system are:

1. Hourly outlet discharge rate (Q);
2. Hourly average water table depth (distance from the land surface to the groundwater table) at three RTHnet wells (WTD);
3. Hourly average integrated soil moisture content over the soil column (0–50 cm) at three RTHnet wells (SWC);
4. Hourly average sensible heat flux (H) via above-canopy eddy covariance measurements (average of two 30-min flux measurements); and
5. Hourly average latent heat flux (LE; same as above).

Please see Shi et al. (2013) and Shi et al. (2014b) for the locations of the measurements.

Stream discharge integrates outflow across the entire watershed. These observations are compared to the simulations of whole-watershed discharge. The discharge is measured with a V-notch weir. The water level at the weir is measured using a Campbell CS420-L transducer with a precision of 7 mm (Campbell Scientific, Inc., 2007), and is converted to discharge rate using a rating curve developed by Nutter (1964). The observed discharge is converted into log space to improve EnKF performance (Clark et al., 2008).

Figure 1: (a) Comparison of precipitation between RTHnet observation and NLDAS-2 forcing, and comparison of LAI between rescaled and unchanged MODIS product. Observations of (b) discharge, (c) water table depth (WTD), and (d) soil water content (SWC). The shaded areas in (b), (c), and (d) represent observation errors and the stars represent assimilated observations in the real data experiments.

Shi et al. (2014b, their Equation (7)) calculated the random error of discharge in log space at the Shale Hills watershed. The observations and errors in observed discharge are shown in Figure 1b. Converting discharge to log space may improve the EnKF performance (Clark et al., 2008), but it can also exaggerate model errors for low flows. For example, when the forecast is $0.1 \text{ m}^3 \text{ d}^{-1}$ and the observation is $0.01 \text{ m}^3 \text{ d}^{-1}$, the forecast error is the same in log space as when the forecast is $1000 \text{ m}^3 \text{ d}^{-1}$ and the observation is $100 \text{ m}^3 \text{ d}^{-1}$. To avoid exaggerated errors and associated system shocks at low flows, and to avoid taking log of a zero discharge, a $1.0 \text{ m}^3 \text{ d}^{-1}$ discharge rate is added to both the observation and the forecasts before calculating the Kalman gain.

Observations of WTD and SWC are calculated by averaging multiple groundwater level measurements and volumetric soil moisture content measurements at the RTHnet wells near the stream. For the Shale Hills simulation, the model domain is discretized as such that the three RTHnet wells are located at three vertices of one model grid for the convenience of model data comparison. Thus, the measurements at the RTHnet wells represent the observed WTD and SWC at the model grid that is surrounded by RTHnet wells. These data are not compared to any other grid points in the simulation, thus the point nature of these measurements is preserved. Each well is equipped with one water level sensor (Druck pressure transducer CS420-L manufactured by Campbell Scientific) and three soil moisture sensors (Decagon Echo2 probes) at different levels below ground. The WTD and SWC observations from different wells show considerable spatial variability. Because the representation uncertainties (the standard errors among three wells) of WTD and SWC are always much larger than their instrumental errors (about 0.007 m for WTD and $0.01 \text{ m}^3 \text{ m}^{-3}$ for SWC (Campbell Scientific, Inc., 2007; Czarnomski et al., 2005)), we conclude that representation uncertainties dominate the uncertainty for these measurements. For every hourly WTD and SWC, we use the computed standard errors among the three RTHnet wells as the observation errors for WTD and SWC, which are shown in Figures 1c and 1d.

In Flux-PIHM, the surface energy balance is closed, i.e., $R_n - G = H + \text{LE}$, where R_n is the net radiation, and G is the ground heat flux. Eddy covariance measurements, however, always fail to close the energy budget, and $H + \text{LE}$ tends to be less than $R_n - G$ (McNeil and Shuttleworth, 1975; Fritschen et al., 1992; Twine et al., 2000; Foken, 2008). The surface heat fluxes measured at the SSHCZO using eddy-covariance (H and LE) likely have this consistent

low bias. Since it is a densely forested site, ground heat fluxes at the Shale Hills watershed are likely to be small. At the Shale Hills watershed in the growing season, simulated mid-day ground heat fluxes are always below 3% of R_n , and the average ratio between simulated ground heat flux and net radiation in 2009 is about 4% (Shi et al., 2013). We therefore treat G as negligible, and rescale $H + \text{LE}$ using R_n . When the sum of hourly averaged surface heat fluxes $H_0 + \text{LE}_0 < R_n$, the surface heat fluxes are rescaled as

$$H = \frac{H_0}{H_0 + \text{LE}_0} R_n, \quad (1)$$

$$\text{LE} = \frac{\text{LE}_0}{H_0 + \text{LE}_0} R_n. \quad (2)$$

The eddy covariance flux observations represent a flux footprint of approximately 1 km² (e.g., Horst and Weil, 1992). The flux measurements are thus compared to simulated watershed-average fluxes. The H and LE fluxes are calculated following the quality control methods documented by Vickers and Mahrt (1997). The processed 30-min H and LE fluxes are aggregated into hourly fluxes. Then the aggregated hourly sensible and latent heat fluxes are rescaled using hourly average net radiation observations according to Eq. (1). This rescaling closes the surface energy balance on average, but does not eliminate the random variability in H and LE that occurs with half-hourly flux measurements (e.g., Berger et al., 2001; Richardson et al., 2006). The rescaling, however, only works when both H and LE observations are available. If either H or LE is missing, the rescaling cannot be applied and neither H nor LE will be assimilated into the system. As a result, about 30% of 1700 UTC H and LE fluxes cannot be assimilated. The random observation errors in H and LE are estimated to be 10%, based on the site characteristics and 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).

5. Results

Figure 2 presents the temporal evolution of the calibration coefficients of the estimated parameters from April to August 2009; the manually calibrated values are shown for reference. The temporal evolution of the standard

Figure 2: Temporal evolution of estimated parameters in real-data experiment from different test cases. The dotted black lines represent the manually calibrated parameter values, and the insets on top represent the evolution of the standard deviation in the Case0 experiment.

268 deviations (σ) of parameters for Case0 is also presented. The temporal evo-
 269 lutions of σ in the other test cases are similar. To avoid using time-variant
 270 parameters (Liu and Gupta, 2007), the average of each parameter value from
 271 1 August to 1 September 2009 is taken to be the EnKF calibrated parameter
 272 value. The EnKF calibrated values are shown in Table 1, and are compared
 273 with the manually calibrated parameter values.

274 In Case0, Case+, and Case-, for all parameters except for S , the es-
 275 timates from different test cases converge after about two months of simu-
 276 lation and data assimilation, while the uncertainty ranges (represented by
 277 the standard deviation) decrease and then stabilize (Figure 2). The tempo-
 278 ral evolution of estimated parameters becomes similar after convergence. For
 279 those five parameters, especially α and β , the estimates converge towards the
 280 manually calibrated values, and the calibrated parameter values are close to
 281 the manually calibrated values (Table 1). The performances of the reanalysis
 282 and remote-sensing driven test cases are similar to Case0 (Figure 2). The

temporal fluctuations of parameter values, however, are stronger, especially in the NLDAS and NLDAS+MODIS test cases, probably due to the errors in NLDAS-2 meteorological forcing (e.g., Figure 1a). The estimates of the parameter $R_{c\min}$ in the MODIS and NLDAS+MODIS test cases are always larger than in Case0, because the unchanged MODIS LAI has larger values than the rescaled LAI (Figure 1a).

Figure 2 indicates that the parameters Θ_e , α , β , and C_{zil} are highly identifiable. The uncertainty in these parameters decreases fast and the estimates from different test cases converge quickly. In contrast, the uncertainty of S almost remains constant during the experiment and the estimates from different test cases do not converge at all. In the summer months (June, July, and August), the rate at which the uncertainty of $R_{c\min}$ decreases is larger than in April and May (Figure 2).

To test the EnKF calibrated parameter sets, evaluation runs with the calibrated parameter sets are executed. The parameter values calibrated by EnKF for different test cases in Table 1 are assigned to those six parameters. The other parameters that are not calibrated in this experiment are set to their manually calibrated values as in Shi et al. (2013). A simulation using the default parameters with no calibration is also performed, driven by locally-measured atmospheric forcing and rescaled MODIS LAI. The evaluation runs start from 0000 UTC 1 January 2009, from the relaxation mode, and are driven by the forcing listed in Table 2. The model predictions of Q , WTD, SWC, H , and LE from 0000 UTC 1 September to 0000 UTC 1 December 2009 are evaluated by comparing with the observations. Predictions of the evaluation runs are also compared with the Flux-PIHM evaluation run with the manually calibrated parameter set, which is driven by locally-measured atmospheric forcing and rescaled MODIS LAI (Table 1). Note that the surface heat flux observations used in the evaluation are not rescaled. The comparisons are presented in Figure 3.

Generally, performances of the Flux-PIHM evaluation runs with the EnKF calibrated parameter sets are comparable to the Flux-PIHM run with the manually calibrated parameter set (Figure 3), and show improvements in forecasting skill compared with the simulation using default parameters, especially for discharge and SWC. The Taylor diagrams demonstrate that manual calibration only exceeds the automated calibration results in the discharge prediction. For the other observable variables, the differences of manual calibration and EnKF calibration are almost indistinguishable. For the discharge prediction, the manually calibrated parameters provide better prediction for

Figure 3: Evaluation of the model predictions using the EnKF-estimated parameter sets and manual calibration for (a) discharge, (b) water table depth, (c) soil water content, and (d) sensible and (e) latent heat fluxes from 0000 UTC 1 September to 0000 UTC 1 December 2009. Flux-PIHM predictions using default parameters (not calibrated) are also shown. In each subfigure, the plot on the left show the model predictions and observations (for H and LE , only 1 September to 20 September is shown due to limitation of space), and the Taylor diagram on the right indicates correlation coefficient, normalized standard deviation, and root mean squared error as well as average model bias in the inset.

the two highest discharge peak events (16 and 24 October 2009). The Taylor diagram in Figure 3a shows that the manual calibration evaluation run has a higher correlation coefficient (about 0.9) with the observations than the EnKF calibrated parameter sets (about 0.8), smaller average bias, and smaller root mean square errors. For WTD and SWC, the Case- evaluation run performs slightly worse than the other runs, while performances of manual calibration, Case0 and Case+ are very similar. For H and LE, all evaluation runs yield higher surface heat fluxes than observed, but this is expected given the rescaling of H and LE observation data (Figures 3d and 3e).

The performances of the reanalysis and remote-sensing driven simulations are limited by the quality of the reanalysis and remote-sensing forcing, especially the NLDAS-2 forcing quality. The Taylor diagram shows that when driven by the NLDAS-2 forcing, Flux-PIHM discharge predictions are worse than the other evaluation runs for the evaluation period (Figure 3a), due to the errors in NLDAS-2 precipitation forcing. During the evaluation period, the NLDAS-2 forcing underestimates the total precipitation by 4.8 cm. The RMSE of the NLDAS-2 hourly precipitation forcing is about 0.33 mm day^{-1} , and the correlation coefficient between the observed hourly precipitation and the NLDAS-2 hourly precipitation is 0.58. Flux-PIHM surface heat flux predictions are also worse than the other evaluation runs when driven by the NLDAS-2 forcing (Figures 3d and 3e), primarily due to the errors of the NLDAS-2 downward solar radiation forcing. For example, as shown in Figures 3d and 3e, the NLDAS and NLDAS+MODIS evaluation runs significantly overestimate the mid-day surface heat fluxes on 17 September, when NLDAS-2 overestimates the mid-day downward solar radiation by about 360 W m^{-2} . The MODIS evaluation run performance is very similar to Case0. Although MODIS overestimates the LAI (Figure 1a), the parameter $R_{c \min}$ estimated in the MODIS and NLDAS+MODIS simulations are higher than the other simulations, which compensate the high bias of the MODIS LAI forcing.

6. Discussion and conclusions

The results of the multivariate real-data experiment demonstrate the capability of EnKF in parameter estimation for a physically-based land surface hydrologic model (Flux-PIHM) using multivariate field observations. The EnKF sequential calibration results are comparable to the manual calibra-

tion while significantly improving the overall efficiency in time and effort. Manual calibration took a large number of repeated runs and many days to perform, while the EnKF data assimilation at the Shale Hills watershed took less than 6 hours of wall time using 31 CPU processors (2.4 GHz) running in parallel.

The EnKF system is shown to be a powerful tool for multivariate data assimilation and parameter estimation. EnKF does not use explicit objective functions, or assign explicit weights to each assimilated observation data set. The weights of observations in EnKF are determined by their observation errors and the ensemble forecast uncertainties and covariances (Clark et al., 2008). Compared with other multivariate or multi-objective function calibration methods, which usually use empirical weights for different observations or objective functions, the weights in EnKF are more physically meaningful.

The fast decrease in parameter uncertainty and convergence of parameter estimates suggest that Θ_e , α , β , and C_{zil} are highly identifiable parameters. This agrees with the Flux-PIHM sensitivity analysis results, which showed that these four parameters have much higher identifiability, especially distinguishability, than the parameters R_{cmin} and S (Shi et al., 2014a, their Figure 9). In their synthetic experiments, Shi et al. (2014b) also showed that when Q , WTD, SWC, H , and LE were assimilated into the system, EnKF failed to provide accurate estimates of the parameters R_{cmin} and S (Shi et al., 2014b, their Table 5).

The temporal evolution of the parameter Θ_e implies that the performance of parameter estimation might be affected by errors other than observational errors. The parameter Θ_e has two significant changes in parameter value during the calibration period (Figure 2) in all test cases. During the first few analysis steps, the estimates of Θ_e from different test cases converge to about 0.65, which is close to the manually calibrated value. But a large change of Θ_e appears on 2 May, and the parameter value deviates from the manually calibrated value. The meteorology forcing shows there was a precipitation event in the early morning on that day. In the manual calibration process we noticed that however we tuned the parameter values, Flux-PIHM persistently predicted a discharge peak on 2 May (see the results in (Shi et al., 2013)), which was not observed in the outlet discharge (Figure 1b). This error might be caused not only by parameter values, but also errors in model structures, static input data, meteorological forcing. Because of the discrepancy in model prediction and observation, EnKF changes the Θ_e value dramatically, to compensate the other unidentified sources of error. The following low flow

395 discharge observations make little impact on Θ_e . Another significant change
 396 in parameter value appears on 20 June, when the peak discharge event occurs
 397 (Figure 1). After the observation of discharge peak is assimilated, EnKF
 398 once again adjusts Θ_e value towards the manually calibrated value, and the
 399 parameter value generally stabilizes afterwards. This suggests that discharge
 400 peak observations must be assimilated to effectively estimate the value of Θ_e .

401 Because of its role in transpiration prediction, $R_{c\min}$ is apparently more
 402 identifiable in summer than in spring (Figure 2), consistent with the Flux-
 403 PIHM sensitivity analysis (Shi et al., 2014a). Yu et al. (2013) divided PIHM
 404 parameters into event-scale parameters and seasonal time scale parameters.
 405 The parameter Θ_e was categorized as an event-scale parameter and $R_{c\min}$
 406 a seasonal time scale parameter. The temporal evolutions of Θ_e and $R_{c\min}$
 407 in our experiment (Figure 2) support their categorization: Θ_e can only be
 408 effectively estimated when observations from peak discharge events are as-
 409 similated, and $R_{c\min}$ can only be effectively estimated in summer. It implies
 410 that the calibration period and assimilation interval need to be chosen wisely,
 411 to cover important peak discharge events and also the time period when most
 412 parameters are identifiable. Although a fixed assimilation interval is used for
 413 this study, it is possible that a more sophisticated dynamic assimilation pe-
 414 riod that mixed wet and dry periods would prove effective. Experiments
 415 with the calibration period variations and the assimilation intervals will be
 416 addressed in future studies.

417 When estimating those six parameters using EnKF, other parameters are
 418 fixed at their manually calibrated values. The impacts of the other param-
 419 eters and their potential interaction with the six parameters estimated in this
 420 paper are left for future studies. It is possible that if more parameters are
 421 estimated using EnKF, the optimized parameters may provide better predic-
 422 tions. The efficiency of assimilating different observations is another future
 423 research topic.

424 The results of the NLDAS, MODIS, and NLDAS+MODIS test cases
 425 demonstrate the robustness of the EnKF in land surface hydrologic model
 426 parameter estimation. The EnKF is capable of providing reliable estimates
 427 of model parameters when using atmospheric reanalysis and remote sensing
 428 products to drive the model. In addition, most of the *a priori* soil, river
 429 bed, and vegetation parameters used for the study are empirical or from a
 430 research database (e.g., the Soil Survey Geographic (SSURGO) database and
 431 the modified International GeosphereBiosphere Programme (IGBP) MODIS
 432 20-category vegetation (land use) data, see Tables 1, 2, and 3 in Shi et al.,

2013). These input data are available with national and in some cases global spatial coverage. Therefore the EnKF data assimilation system is not limited to measurement-rich watersheds like Shale Hills, but could readily be extended to other watersheds. Data infrastructures like the HydroTerre system (Leonard and Duffy, 2013) enable convenient expansion to different watersheds. The rapid evolution of computing power and the rise of parallel computing technique will allow the Flux-PIHM EnKF data assimilation system to be implemented at larger river basins.

Acknowledgments

This research was supported by NOAA through grant NA10OAR4310166, and NSF Grant EAR0725019 (C. Duffy), EAR1239285 (S. Brantley), and EAR1331726 (S. Brantley) for the SSHCZO. Logistical support and/or data were provided by the NSF-supported SSHCZO. The NLDAS-2 forcing data used in this study were acquired as part of the mission of NASA’s Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

References

- Aksoy, A., Zhang, F., Nielsen-Gammon, J. W., 2006. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.* 134 (10), 2951–2970.
- Anderson, J. L., Anderson, S. L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* 127 (12), 2741–2758.
- Annan, J. D., 2005. Parameter estimation using chaotic time series. *Tellus A* 57 (5), 709–714.
- Baldocchi, D., Valentini, R., Running, S., Oechel, W., Dahlman, R., 1996. Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems. *Global Change Bio.* 2 (3), 159–168.
- Baldocchi, D. D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Bio.* 9 (4), 479–492.

- 464 Berger, B. W., Davis, K. J., Yi, C., Bakwin, P. S., Zhao, C. L., 2001. Long-
465 term carbon dioxide fluxes from a very tall tower in a northern forest: Flux
466 measurement methodology. *J. Atmos. Oceanic Technol.* 18 (4), 529–542.
- 467 Campbell Scientific, Inc., 2007. CS420-L and CS425-L Druck’s Models PDCR
468 1830 and 1230 pressure transducers. Instruction manual 12/07, Campbell
469 Scientific, Inc., Palo Alto, Calif.
- 470 Chen, F., Dudhia, J., 2001. Coupling an advanced land surface-hydrology
471 model with the Penn State-NCAR MM5 modeling system. Part I: Model
472 implementation and sensitivity. *Mon. Wea. Rev.* 129 (4), 569–585.
- 473 Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater,
474 A. G., Schmidt, J., Uddstrom, M. J., 2008. Hydrological data assimilation
475 with the ensemble Kalman filter: Use of streamflow observations to update
476 states in a distributed hydrological model. *Adv. Water Resour.* 31 (10),
477 1309–1324.
- 478 Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F.,
479 Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., et al.,
480 2003. Real-time and retrospective forcing in the North American Land
481 Data Assimilation System (NLDAS) project. *J. Geophys. Res.* 108 (D22),
482 8842.
- 483 Czarnomski, N. M., Moore, G. W., Pypker, T. G., Licata, J., Bond, B. J.,
484 2005. Precision and accuracy of three alternative instruments for measuring
485 soil water content in two forest soils of the Pacific Northwest. *Can. J. For.*
486 *Res.* 35 (8), 1867–1876.
- 487 Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grummann, P., Koren,
488 V., Gayno, G., Tarpley, J. D., 2003. Implementation of Noah land sur-
489 face model advances in the National Centers for Environmental Prediction
490 operational Mesoscale Eta Model. *J. Geophys. Res.* 108 (D22), 8851.
- 491 Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-
492 geostrophic model using Monte Carlo methods to forecast error statistics.
493 *J. Geophys. Res.* 99 (C5), 10143–10162.
- 494 Finkelstein, P. L., Sims, P. F., 2001. Sampling error in eddy correlation flux
495 measurements. *J. Geophys. Res.* 106 (D4), 3503–3509.

- 496 Foken, T., 2008. The energy balance closure problem: An overview. *Ecol.*
497 *Appl.* 18 (6), 1351–1367.
- 498 Fritschen, L. J., Qian, P., Kanemasu, E. T., Nie, D., Smith, E. A., Stew-
499 art, J. B., Verma, S. B., Wesely, M. L., 1992. Comparisons of surface
500 flux measurement systems used in FIFE 1989. *J. Geophys. Res.* 97 (D17),
501 18697–18713.
- 502 Horst, T. W., Weil, J. C., 1992. Footprint estimation for scalar flux measure-
503 ments in the atmospheric surface layer. *Boundary-Layer Meteorol.* 59 (3),
504 279–296.
- 505 Houser, P. R., De Lannoy, G. J. M., Walker, J. P., 2010. Land surface data
506 assimilation. In: Lahoz, W., Khattatov, B., Menard, R. (Eds.), *Data As-*
507 *similation*. Springer Berlin Heidelberg, pp. 549–597.
- 508 Hu, X.-M., Zhang, F., Nielsen-Gammon, J. W., 2010. Ensemble-based simul-
509 taneous state and parameter estimation for treatment of mesoscale model
510 error: A real-data study. *Geophys. Res. Lett.* 37.
- 511 Knyazikhin, Y., Glassy, J., Privette, J. L., Tian, Y., Lotsch, A., Zhang, Y.,
512 Wang, Y., Morisette, J. T., Votava, P., Myneni, R. B., et al., 1999. MODIS
513 leaf area index (LAI) and fraction of photosynthetically active radiation
514 absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical
515 basis document. Theoretical basis document, NASA Goddard Space Flight
516 Center, Greenbelt, Maryland.
- 517 Kumar, M., 2009. Toward a hydrologic modeling system. Ph.D. thesis, The
518 Pennsylvania State University.
- 519 Lenschow, D. H., Mann, J., Kristensen, L., 1994. How long is long enough
520 when measuring fluxes and other turbulence statistics? *J. Atmos. Oceanic*
521 *Technol.* 11 (3), 661–673.
- 522 Lenschow, D. H., Stankov, B. B., 1986. Length scales in the convective bound-
523 ary layer. *J. Atmos. Sci.* 43 (12), 1198–1209.
- 524 Leonard, L., Duffy, C. J., 2013. Essential Terrestrial Variable data workflows
525 for distributed water resources modeling. *Environ. Modell. Software* 50,
526 85–96.

- 527 Liu, Y., Gupta, H. V., 2007. Uncertainty in hydrologic modeling: Toward
528 an integrated data assimilation framework. *Water Resour. Res.* 43 (7),
529 W07401.
- 530 Lü, H., Hou, T., Horton, R., Zhu, Y., Chen, X., Jia, Y., Wang, W., Fu, X.,
531 2013. The streamflow estimation using the Xinanjiang rainfall runoff model
532 and dual state-parameter estimation method. *J. Hydrol.* 480, 102–114.
- 533 McNeil, D. D., Shuttleworth, W. J., 1975. Comparative measurements of the
534 energy fluxes over a pine forest. *Bound.-Layer Meteor.* 9 (3), 297–313.
- 535 Moradkhani, H., Sorooshian, S., 2008. General review of rainfall-runoff mod-
536 eling: Model calibration, data assimilation, and uncertainty analysis. In:
537 Sorooshian, S., Hsu, K.-L., Coppola, E., Tomassetti, B., Verdecchia, M.,
538 Visconti, G. (Eds.), *Hydrological Modelling and the Water Cycle*. Springer,
539 pp. 1–24.
- 540 Moradkhani, H., Sorooshian, S., Gupta, H. V., Houser, P. R., 2005. Dual
541 state-parameter estimation of hydrological models using ensemble Kalman
542 filter. *Adv. Water Resour.* 28 (2), 135–147.
- 543 Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian,
544 Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., et al., 2002. Global
545 products of vegetation leaf area and fraction absorbed PAR from year one
546 of MODIS data. *Remote Sens. Environ.* 83, 214–231.
- 547 Nutter, W. L., 1964. Determination of the head-discharge relationship for a
548 sharp-crested compound weir and a sharp-crested parabolic weir. Master’s
549 thesis, The Pennsylvania State University.
- 550 Pokhrel, P., Gupta, H. V., 2010. On the use of spatial regularization strategies
551 to improve calibration of distributed watershed models. *Water Resour. Res.*
552 46 (1).
- 553 Qu, Y., 2004. An integrated hydrologic model for multi-process simulation
554 using semi-discrete finite volume approach. Ph.D. thesis, The Pennsylvania
555 State University.
- 556 Qu, Y., Duffy, C. J., 2007. A semidiscrete finite volume formulation for mul-
557 tiprocess watershed simulation. *Water Resour. Res.* 43 (8).

- 558 Richardson, A. D., Hollinger, D. Y., Burba, G. G., Davis, K. J., Flana-
 559 gan, L. B., Katul, G. G., William Munger, J., Ricciuto, D. M., Stoy,
 560 P. C., Suyker, A. E., et al., 2006. A multi-site analysis of random error in
 561 tower-based measurements of carbon and energy fluxes. *Agric. For. Meteor.*
 562 136 (1), 1–18.
- 563 Salesky, S. T., Chamecki, M., Dias, N. L., 2012. Estimating the random
 564 error in eddy-covariance based fluxes and other turbulence statistics: the
 565 filtering method. *Bound.-Layer Meteor.* 144 (1), 113–135.
- 566 Shi, Y., Davis, K. J., Duffy, C. J., Yu, X., 2013. Development of a coupled
 567 land surface hydrologic model and evaluation at a critical zone observatory.
 568 *J. Hydrometeor.* 14 (5), 1401–1420.
- 569 Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., 2014a. Evaluation of the pa-
 570 rameter sensitivity of a coupled land surface hydrologic model at a critical
 571 zone observatory. *J. Hydrometeor.* 15 (1), 279–299.
- 572 Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., Yu, X., 2014b. Parameter
 573 estimation of a physically-based land surface hydrologic model using the
 574 ensemble Kalman filter: A synthetic experiment. *Water Resour. Res.* 50,
 575 706–724.
- 576 Snyder, C., Zhang, F., 2003. Assimilation of simulated Doppler radar obser-
 577 vations with an ensemble Kalman filter. *Mon. Wea. Rev.* 131 (8), 1663–
 578 1677.
- 579 Twine, T. E., Kustas, W. P., Norman, J. M., Cook, D. R., Houser, P. R.,
 580 Meyers, T. P., Prueger, J. H., Starks, P. J., Wesely, M. L., 2000. Correcting
 581 eddy-covariance flux underestimates over a grassland. *Agric. For. Meteor.*
 582 103 (3), 279–300.
- 583 van Genuchten, M. T., 1980. A closed-form equation for predicting the hy-
 584 draulic conductivity of unsaturated soils. *Soil Sci. Soc. Amer. J.* 44 (5),
 585 892–898.
- 586 Vickers, D., Mahrt, L., 1997. Quality control and flux sampling problems for
 587 tower and aircraft data. *J. Atmos. Oceanic Technol.* 14 (3), 512–526.

- 588 Wallner, M., Haberlandt, U., Dietrich, J., 2012. Evaluation of different cali-
589 bration strategies for large scale continuous hydrological modelling. *Adv.*
590 *Geosci.* 31, 67–74.
- 591 Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L.,
592 Alonge, C., Wei, H., Meng, J., et al., 2012. Continental-scale water and
593 energy flux analysis and validation for the North American Land Data
594 Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and
595 application of model products. *J. Geophys. Res.* 117.
- 596 Xie, X., Zhang, D., 2010. Data assimilation for distributed hydrological catch-
597 ment modeling via ensemble Kalman filter. *Adv. Water Resour.* 33 (6),
598 678–690.
- 599 Yu, X., Bhatt, G., Duffy, C., Shi, Y., 2013. Parameterization for distributed
600 watershed modeling using national data and evolutionary algorithm. *Com-
601 put. Geosci.* 58, 80–90.
- 602 Zhang, F., Snyder, C., Sun, J., 2004. Impacts of initial estimate and obser-
603 vation availability on convective-scale data assimilation with an ensemble
604 Kalman filter. *Mon. Wea. Rev.* 132 (5), 1238–1253.
- 605 Zilitinkevich, S. S., 1995. Non-local turbulent transport: pollution dispersion
606 aspects of coherent structure of convective flows. In: Power, H., Mous-
607 siopoulos, N., Brebbia, C. A. (Eds.), *Air Pollution III—Volume I Air
608 Pollution Theory and Simulation*. Computational Mechanics Publications,
609 Southampton, Boston, pp. 53–60.