Advances in Water Resources 83 (2015) 421–427

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

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

Yuning Shi [a,*], Kenneth J. Davis [a,b], Fuqing Zhang [b], Christopher J. Duffy [c], Xuan Yu [c]

[a] Earth and Environmental Systems Institute, The Pennsylvania State Unviersity, University Park, PA, USA
[b] Department of Meteorology, The Pennsylvania State Unviersity, University Park, PA, USA
[c] Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA, USA

ARTICLE INFO

Article history:
Received 9 December 2014
Revised 10 May 2015
Accepted 19 June 2015
Available online 4 July 2015

Keywords:
Data assimilation
Parameter estimation
Land surface model
Hydrologic model
Ensemble Kalman filter

ABSTRACT

The capability of an ensemble Kalman filter (EnKF) to simultaneously estimate multiple parameters in a physically-based land surface hydrologic model using multivariate field observations is tested at a small watershed (0.08 km^2). Multivariate, high temporal resolution, *in situ* measurements of discharge, water table depth, soil moisture, and sensible and latent heat fluxes encompassing five months of 2009 are assimilated. It is found that, for five out of the six parameters, the EnKF estimated parameter values from different test cases converge strongly, and the estimates after convergence are close to the manually calibrated parameter values. The EnKF estimated parameters and manually calibrated parameters yield similar model performance, but the EnKF sequential method significantly decreases the time and labor required for calibration. The results demonstrate that, given a limited number of multi-state, site-specific observations, an automated sequential calibration method (EnKF) can be used to optimize physically-based land surface hydrologic models.

1. Introduction

Uncertainties in model parameters are a dominant source of uncertainty for hydrologic models [28]. The ensemble Kalman filter (EnKF) [13] provides a promising approach for the automated calibration of hydrologic models [26,29,39,46]. Most previous studies applied EnKF to conceptual or process-based hydrologic models. Shi et al. [39] performed a multiple-parameter estimation for a physically-based land surface hydrologic model, Flux-PIHM [37], *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 km^2) 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 the 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 [18].

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 performance of the data assimilation system when driven by atmospheric reanalysis and remotely-sensed forcing data, to evaluate the ability of the data assimilation method to adapt to commonly available continental-scale driver data.

* Corresponding author: Tel.: 8148657393.
E-mail address: yshi@psu.edu (Y. Shi).

http://dx.doi.org/10.1016/j.advwatres.2015.06.009
0309-1708/

Table 1
Flux-PIHM model parameters, their plausible ranges of calibration coefficients, estimates from different test cases, and manual calibration values [37]. 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 ($m^3\ m^{-3}$)	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 (dimensionless)	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 (dimensionless)	0.1–10	1.15	1.09	1.23	0.81	1.32	0.93	0.70

2. Flux-PIHM EnKF system

Flux-PIHM [37] is a coupled land surface hydrologic model. Flux-PIHM incorporates a land surface scheme into the Penn State Integrated Hydrologic Model (PIHM) [21,33,34], which is a fully-coupled, physically-based, spatially-distributed hydrologic model. The land surface scheme in Flux-PIHM is adapted from the Noah land surface model (LSM) [8,12]. 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.

A Flux-PIHM data assimilation system has been developed by incorporating EnKF for model parameter and state estimation [39] using the EnKF formulation from Snyder and Zhang [40]. In the Flux-PIHM EnKF system, the Flux-PIHM model variables and the global calibration coefficients of model parameters are concatenated into a joint state parameter vector **x**, and are updated simultaneously by EnKF using the state augmentation approach [1,3,19,25,46]. The global calibration coefficient [32,37,44] is a scalar multiplier applied to the corresponding soil or vegetation related parameter for all soil or vegetation types, and is used to decrease the dimension of the joint state parameter vector. The covariance relaxation method of Zhang et al. [48, Eq. (5)] is applied on model parameters and variables in order to avoid filter divergence [2]. In addition, the conditional covariance inflation method [1] is applied to model parameters. A quality control process [39] is performed after each EnKF analysis step to ensure the parameters and state variables remain within physically realistic or plausible ranges. Please see Shi et al. [37,39] for detailed descriptions.

3. Experimental setup

The Flux-PIHM EnKF data assimilation system is implemented at the Shale Hills watershed (0.08 km^2) 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 m^2. 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. [37]. Given the small scale (0.08 km^2) 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-d leaf area index (LAI) data [20,30] are rescaled based on the comparison between the MODIS product and the CZO field measurements to drive the model [37]. The parameters to be estimated are: effective porosity Θ_e, van Genuchten [42] soil parameters α and β, Zilitinkevich [49] parameter C_{zil}, minimum stomatal resistance $R_{c\,min}$, and reference canopy water capacity S. The estimation of those parameters has been tested in synthetic experiments [39]. 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. [37,38].

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. [37]. 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 [37]. For each parameter (calibration coefficient) ϕ, the values are randomly drawn from a Gaussian distribution, with an initial standard deviation of $\sigma_0 = 0.2(\phi_{max} - \phi_{min})$, where ϕ_{max} and ϕ_{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. [39] 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 [37]. The model time step is 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. [39] found that the assimilation interval for synthetic experiments at the Shale Hills watershed should be larger than 72 h 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 h 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) [10,45] and rescaled

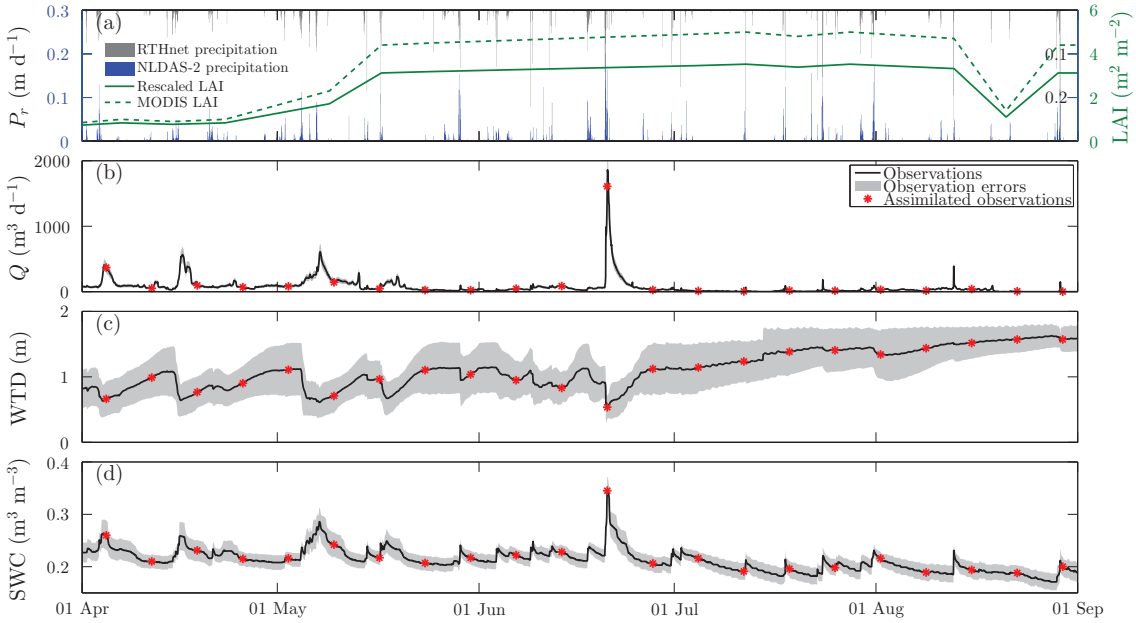

Fig. 1. (a) Comparison of precipitation between RTHnet observation and NLDAS-2 forcing, and comparison of LAI between rescaled and unchanged MODIS product. Observations of (b) discharge, (c) water table depth (WTD), and (d) soil water content (SWC). The shaded areas in (b–d) represent observation errors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

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

MODSI 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 interpolated from 1/8° 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. Fig. 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; the same as above).

Please see Shi et al. [37,39] 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 [7], and is converted to discharge rate using a rating curve developed by Nutter [31]. The observed discharge is converted into log space to improve EnKF performance [9]. Shi et al. [39, Eq. (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 Fig. 1b. Converting discharge to log space may improve the EnKF performance [9], but it can also exaggerate model errors for low flows. For example, when the forecast is 0.1 m^3 d^{-1} and the observation is 0.01 m^3 d^{-1}, the forecast error is the same in log space as when the forecast is 1000 m^3 d^{-1} and the observation is 100 m^3 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 m^3 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 m^3 m^{-3} for SWC [7,11]), 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 Fig. 1c and d.

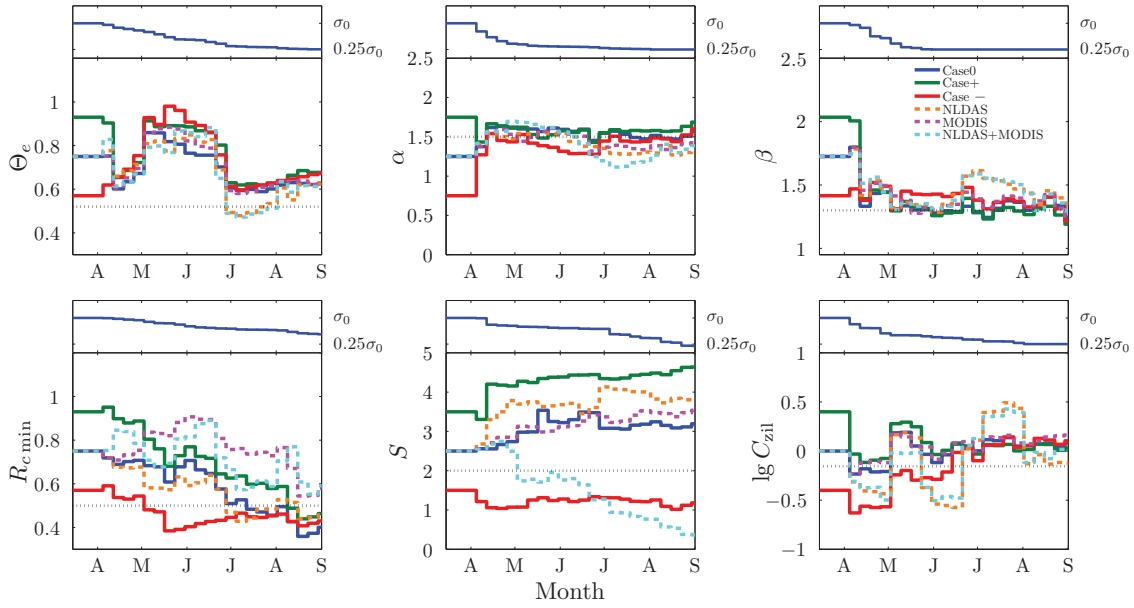

Fig. 2. Temporal evolution of estimated parameters in real-data experiment from different test cases. The dotted black lines represent the manually calibrated parameter values, and the insets on top represent the evolution of the standard deviation in the Case0 experiment.

In Flux-PIHM, the surface energy balance is closed, *i.e.*, $R_n - G = H + \mathrm{LE}$, 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 $H + \mathrm{LE}$ tends to be less than $R_n - G$ [15,16,27,41]. 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% [37]. We therefore treat G as negligible, and rescale $H + \mathrm{LE}$ using R_n. When the sum of hourly averaged surface heat fluxes $H_0 + \mathrm{LE}_0 < R_n$, the surface heat fluxes are rescaled as

$$H = \frac{H_0}{H_0 + \mathrm{LE}_0} R_n, \tag{1}$$

$$\mathrm{LE} = \frac{\mathrm{LE}_0}{H_0 + \mathrm{LE}_0} R_n. \tag{2}$$

The eddy covariance flux observations represent a flux footprint of approximately 1 km^2 [17]. 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 [43]. 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 [6,35]. 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 [4,5,14,22,23,35,36].

5. Results

Fig. 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 deviations (σ) of parameters for Case0 is also presented. The temporal evolutions of σ in the other test cases are similar. To avoid using time-variant parameters [25], the average of each parameter value from 1 August to 1 September, 2009 is taken to be the EnKF calibrated parameter value. The EnKF calibrated values are shown in Table 1, and are compared with the manually calibrated parameter values.

In Case0, Case+, and Case−, for all parameters except for S, the estimates from different test cases converge after about two months of simulation and data assimilation, while the uncertainty ranges (represented by the standard deviation) decrease and then stabilize (Fig. 2). The temporal evolution of estimated parameters becomes similar after convergence. For those five parameters, especially α and β, the estimates converge towards the manually calibrated values, and the calibrated parameter values are close to the manually calibrated values (Table 1). The performances of the reanalysis and remote-sensing driven test cases are similar to Case0 (Fig. 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.*, Fig. 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 (Fig. 1a).

Fig. 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 (Fig. 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

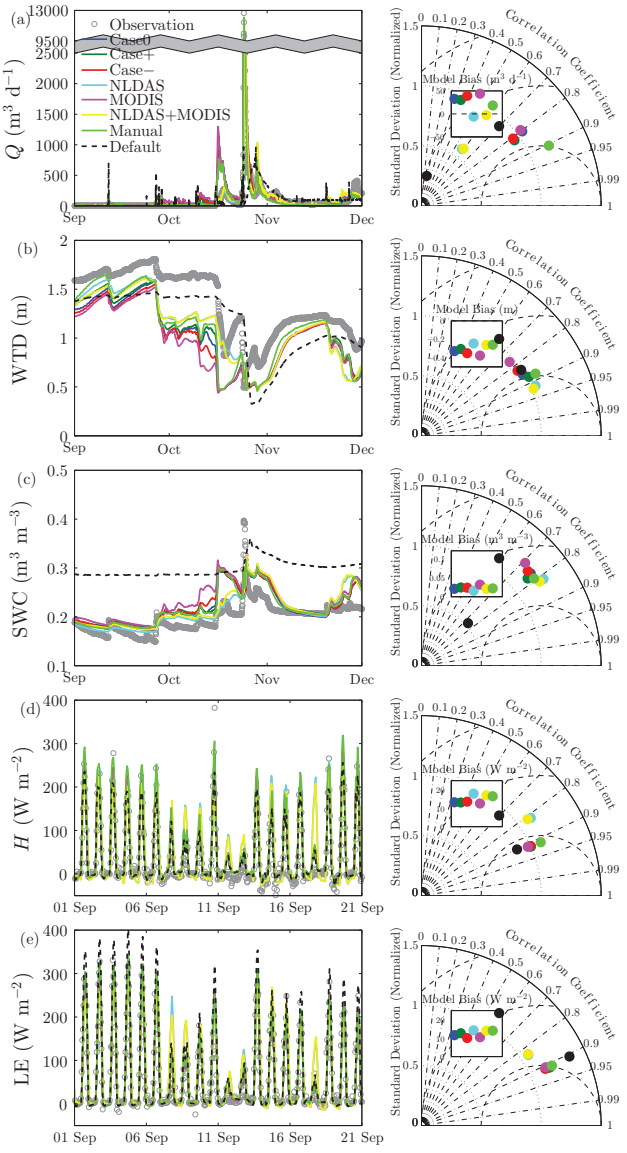

Fig. 3. Evaluation of the model predictions using the EnKF-estimated parameter sets and manual calibration for (a) discharge, (b) water table depth, (c) soil water content, and (d) sensible and (e) latent heat fluxes from 0000 UTC 1 September to 0000 UTC 1 December, 2009. Flux-PIHM predictions using 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.

six parameters. The other parameters that are not calibrated in this experiment are set to their manually calibrated values as in Shi et al. [37]. 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 Fig. 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 (Fig. 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 the two highest discharge peak events (16 and 24 October, 2009). The Taylor diagram in Fig. 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 (Fig. 3d and e).

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 (Fig. 3a), due to the errors in the 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 d^{-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 (Fig. 3d and e), primarily due to the errors of the NLDAS-2 downward solar radiation forcing. For example, as shown in Fig. 3d and e, 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 (Fig. 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 calibration 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 h 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 [9]. 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 $R_{c\,min}$ and S [38, Fig. 9]. In their synthetic experiments, Shi et al. [39] also showed that when Q, WTD, SWC, H, and LE were assimilated into the system, EnKF failed to provide accurate estimates of the parameters $R_{c\,min}$ and S [39, 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 (Fig. 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 [37]), which was not observed in the outlet discharge (Fig. 1b). This error might be caused not only by parameter values, but also by the 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 discharge observations make little impact on Θ_e. Another significant change in parameter value appears on 20 June, when the peak discharge event occurs (Fig. 1). After the observation of discharge peak is assimilated, EnKF once again adjusts Θ_e value towards the manually calibrated value, and the parameter value generally stabilizes afterwards. This suggests that discharge peak observations must be assimilated to effectively estimate the value of Θ_e.

Because of its role in transpiration prediction, $R_{c\,min}$ is apparently more identifiable in summer than in spring (Fig. 2), consistent with the Flux-PIHM sensitivity analysis [38]. Yu et al. [47] divided PIHM parameters into event-scale parameters and seasonal time scale parameters. The parameter Θ_e was categorized as an event-scale parameter and $R_{c\,min}$ a seasonal time scale parameter. The temporal evolutions of Θ_e and $R_{c\,min}$ in our experiment (Fig. 2) support their categorization: Θ_e can only be effectively estimated when observations from peak discharge events are assimilated, and $R_{c\,min}$ can only be effectively estimated in summer. It implies that the calibration period and assimilation interval need to be chosen wisely, to cover important peak discharge events and also the time period when most parameters are identifiable. Although a fixed assimilation interval is used for this study, it is possible that a more sophisticated dynamic assimilation period with mixed wet and dry periods would prove effective. Experiments with the calibration period variations and the assimilation intervals will be addressed in future studies.

When estimating those six parameters using EnKF, other parameters are fixed at their manually calibrated values. The impacts of the other parameters and their potential interaction with the six parameters estimated in this paper are left for future studies. It is possible that if more parameters are estimated using EnKF, the optimized parameters may provide better predictions. The efficiency of assimilating different observations is another future research topic.

The results of the NLDAS, MODIS, and NLDAS+MODIS test cases demonstrate the robustness of the EnKF in land surface hydrologic model parameter estimation. The EnKF is capable of providing reliable estimates of model parameters when using atmospheric reanalysis and remote sensing products to drive the model. In addition, most of the *a priori* soil, river bed, and vegetation parameters used for the study are empirical or from a research database [*e.g.*, the Soil Survey Geographic (SSURGO) database and the modified International GeosphereBiosphere Programme (IGBP) MODIS 20-category vegetation (land use) data, see Tables 1, 2, and 3 in 37]. 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 [24] 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 Grants 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

[1] Aksoy A, Zhang F, Nielsen-Gammon JW. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. Mon Weather Rev 2006;134(10):2951–70. http://dx.doi.org/10.1175/MWR3224.1.
[2] Anderson JL, Anderson SL. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon Weather Rev 1999;127(12):2741–58. http://dx.doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.
[3] Annan JD. Parameter estimation using chaotic time series. Tellus A 2005;57(5):709–14. http://dx.doi.org/10.1111/j.1600-0870.2005.00143.x.
[4] Baldocchi D, Valentini R, Running S, Oechel W, Dahlman R. Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems. Global Change Biol 1996;2(3):159–68. http://dx.doi.org/10.1111/j.1365-2486.1996.tb00069.x.
[5] Baldocchi DD. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Global Change Biol 2003;9(4):479–92. http://dx.doi.org/10.1046/j.1365-2486.2003.00629.x.
[6] Berger BW, Davis KJ, Yi C, Bakwin PS, Zhao CL. Long-term carbon dioxide fluxes from a very tall tower in a northern forest: Flux measurement methodology. J Atmos Ocean Technol 2001;18(4):529–42. http://dx.doi.org/10.1175/1520-0426(2001)018<0529:LTCDFF>2.0.CO;2.
[7] Campbell Scientific, Inc. CS420-L and CS425-L Druck's Models PDCR 1830 and 1230 pressure transducers. Instruction manual 12/07. Palo Alto, California:Campbell Scientific, Inc.; 2007.
[8] Chen F, Dudhia J. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon Weather Rev 2001;129(4):569–85. http://dx.doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.
[9] Clark MP, Rupp DE, Woods RA, Zheng X, Ibbitt RP, Slater AG, et al. Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. Adv Water Resour 2008;31(10):1309–24. http://dx.doi.org/10.1016/j.advwatres.2008.06.005.
[10] Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, et al. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. J Geophys Res 2003;108(D22):8842. http://dx.doi.org/10.1029/2002JD003118.
[11] Czarnomski NM, Moore GW, Pypker TG, Licata J, Bond BJ. Precision and accuracy of three alternative instruments for measuring soil water content in two forest soils of the Pacific Northwest. Can J For Res 2005;35(8):1867–76. http://dx.doi.org/10.1139/X05-121.
[12] Ek MB, Mitchell KE, Lin Y, Rogers E, Grummann P, Koren V, et al. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational Mesoscale Eta Model. J Geophys Res 2003;108(D22):8851. http://dx.doi.org/10.1029/2002JD003296.
[13] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 1994;99(C5):10143–62. http://dx.doi.org/10.1029/94JC00572.
[14] Finkelstein PL, Sims PF. Sampling error in eddy correlation flux measurements. J Geophys Res 2001;106(D4):3503–9. http://dx.doi.org/10.1029/2000JD900731.
[15] Foken T. The energy balance closure problem: an overview. Ecol Appl 2008;18(6):1351–67. http://dx.doi.org/10.1890/06-0922.1.
[16] Fritschen LJ, Qian P, Kanemasu ET, Nie D, Smith EA, Stewart JB, et al. Comparisons of surface flux measurement systems used in FIFE 1989. J Geophys Res 1992;97(D17):18697–713. http://dx.doi.org/10.1029/91JD03042.
[17] Horst TW, Weil JC. Footprint estimation for scalar flux measurements in the atmospheric surface layer. Boundary-Layer Meteorol 1992;59(3):279–96. http://dx.doi.org/10.1007/BF00119817.
[18] Houser PR, De Lannoy GJM, Walker JP. Land surface data assimilation. In: Lahoz W, Khattatov B, Menard R, editors. Data assimilation. Berlin, Heidelberg: Springer; 2010. p. 549–97. http://dx.doi.org/10.1007/978-3-540-74703-1_21.

[19] Hu X-M, Zhang F, Nielsen-Gammon JW. Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: a real-data study. Geophys Res Lett 2010;37. http://dx.doi.org/10.1029/2010GL043017.
[20] Knyazikhin Y, Glassy J, Privette JL, Tian Y, Lotsch A, Zhang Y, et al. MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical basis document Theoretical basis document. Greenbelt, Maryland: NASA Goddard Space Flight Center; 1999.
[21] Kumar M. Toward a hydrologic modeling system [Ph.D. thesis]. The Pennsylvania State University; 2009.
[22] Lenschow DH, Mann J, Kristensen L. How long is long enough when measuring fluxes and other turbulence statistics? J Atmos Oceanic Technol 1994;11(3):661–73. http://dx.doi.org/10.1175/1520-0426(1994)011<0661:HLILEW>2.0.CO;2.
[23] Lenschow DH, Stankov BB. Length scales in the convective boundary layer. J Atmos Sci 1986;43(12):1198–209. http://dx.doi.org/10.1175/1520-0469(1986)043<1198:LSITCB>2.0.CO;2.
[24] Leonard L, Duffy CJ. Essential terrestrial variable data workflows for distributed water resources modeling. Environ Modell Softw 2013;50:85–96. http://dx.doi.org/10.1016/j.envsoft.2013.09.003.
[25] Liu Y, Gupta HV. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour Res 2007;43(7):W07401. http://dx.doi.org/10.1029/2006WR005756.
[26] Lü H, Hou T, Horton R, Zhu Y, Chen X, Jia Y, et al. The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method. J Hydrol 2013;480:102–14. http://dx.doi.org/10.1016/j.jhydrol.2012.12.011.
[27] McNeil DD, Shuttleworth WJ. Comparative measurements of the energy fluxes over a pine forest. Bound-Layer Meteorol 1975;9(3):297–313. http://dx.doi.org/10.1007/BF00230772.
[28] Moradkhani H, Sorooshian S. General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. In: Sorooshian S, Hsu K-L, Coppola E, Tomassetti B, Verdecchia M, Visconti G, editors. Hydrological modelling and the water cycle. Springer; 2008. p. 1–24.
[29] Moradkhani H, Sorooshian S, Gupta HV, Houser PR. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 2005;28(2):135–47. http://dx.doi.org/10.1016/j.advwatres.2004.09.002.
[30] Myneni RB, Hoffman S, Knyazikhin Y, Privette JL, Glassy J, Tian Y, et al. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens Environ 2002;83:214–31. http://dx.doi.org/10.1016/S0034-4257(02)00074-3.
[31] Nutter WL. Determination of the head-discharge relationship for a sharp-crested compound weir and a sharp-crested parabolic weir [Master's thesis]. The Pennsylvania State University; 1964.
[32] Pokhrel P, Gupta HV. On the use of spatial regularization strategies to improve calibration of distributed watershed models. Water Resour Res 2010;46(1). http://dx.doi.org/10.1029/2009WR008066.
[33] Qu Y. An integrated hydrologic model for multi-process simulation using semi-discrete finite volume approach [Ph.D. thesis]. The Pennsylvania State University; 2004.
[34] Qu Y, Duffy CJ. A semidiscrete finite volume formulation for multiprocess watershed simulation. Water Resour Res 2007;43(8):W08419. http://dx.doi.org/10.1029/2006WR005752.
[35] Richardson AD, Hollinger DY, Burba GG, Davis KJ, Flanagan LB, Katul GG, et al. A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. Agric For Meteorol 2006;136(1):1–18. http://dx.doi.org/10.1016/j.agrformet.2006.01.007.
[36] Salesky ST, Chamecki M, Dias NL. Estimating the random error in eddy-covariance based fluxes and other turbulence statistics: the filtering method. Bound-Layer Meteorol 2012;144(1):113–35. http://dx.doi.org/10.1007/s10546-012-9710-0.
[37] Shi Y, Davis KJ, Duffy CJ, Yu X. Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory. J Hydrometeorol 2013;14(5):1401–20. http://dx.doi.org/10.1175/JHM-D-12-0145.1.
[38] Shi Y, Davis KJ, Zhang F, Duffy CJ. Evaluation of the parameter sensitivity of a coupled land surface hydrologic model at a critical zone observatory. J Hydrometeorol 2014a;15(1):279–99. http://dx.doi.org/10.1175/JHM-D-12-0177.1.
[39] Shi Y, Davis KJ, Zhang F, Duffy CJ, Yu X. Parameter estimation of a physically-based land surface hydrologic model using the ensemble Kalman filter: a synthetic experiment. Water Resour Res 2014b;50:706–24. http://dx.doi.org/10.1002/2013WR014070.
[40] Snyder C, Zhang F. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. Mon Weather Rev 2003;131(8):1663–77. http://dx.doi.org/10.1175//2555.1.
[41] Twine TE, Kustas WP, Norman JM, Cook DR, Houser PR, Meyers TP, et al. Correcting eddy-covariance flux underestimates over a grassland. Agric For Meteorol 2000;103(3):279–300. http://dx.doi.org/10.1016/S0168-1923(00)00123-4.
[42] van Genuchten MT. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 1980;44(5):892–8. http://dx.doi.org/10.2136/sssaj1980.03615995004400050002x.
[43] Vickers D, Mahrt L. Quality control and flux sampling problems for tower and aircraft data. J Atmos Ocean Technol 1997;14(3):512–26. http://dx.doi.org/10.1175/1520-0426(1997)014<0512:QCAFSP>2.0.CO;2.
[44] Wallner M, Haberlandt U, Dietrich J. Evaluation of different calibration strategies for large scale continuous hydrological modelling. Adv Geosci 2012;31:67–74. http://dx.doi.org/10.5194/adgeo-31-67-2012.
[45] Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, et al. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J Geophys Res 2012;117:D03109. http://dx.doi.org/10.1029/2011JD016048.
[46] Xie X, Zhang D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Adv Water Resour 2010;33(6):678–90. http://dx.doi.org/10.1016/j.advwatres.2010.03.012.
[47] Yu X, Bhatt G, Duffy C, Shi Y. Parameterization for distributed watershed modeling using national data and evolutionary algorithm. Comput Geosci 2013;58:80–90. http://dx.doi.org/10.1016/j.cageo.2013.04.025.
[48] Zhang F, Snyder C, Sun J. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. Mon Weather Rev 2004;132(5):1238–53. http://dx.doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.
[49] Zilitinkevich SS. Non-local turbulent transport: pollution dispersion aspects of coherent structure of convective flows. In: Power H, Moussiopoulos N, Brebbia CA, editors. Air pollution III – volume I, Air pollution theory and simulation. Southampton, Boston: Computational Mechanics Publications; 1995. p. 53–60.