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 $\mathbf{x}$, 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