



Evaluation of Model Parameter Convergence when Using Data Assimilation for Soil Moisture Estimation

GIFT DUMEDAH AND JEFFREY P. WALKER

Department of Civil Engineering, Monash University, Clayton, Victoria, Australia

(Manuscript received 30 November 2012, in final form 22 August 2013)

ABSTRACT

Data assimilation (DA) methods are commonly used for finding a compromise between imperfect observations and uncertain model predictions. The estimation of model states and parameters has been widely recognized, but the convergence of estimated parameters has not been thoroughly investigated. The distribution of model state and parameter values is closely linked to convergence, which in turn impacts the ultimate estimation accuracy of DA methods. This demonstration study examines the robustness and convergence of model parameters for the ensemble Kalman filter (EnKF) and the evolutionary data assimilation (EDA) in the context of the Soil Moisture and Ocean Salinity (SMOS) soil moisture assimilation into the Joint UK Land Environment Simulator in the Yanco area in southeast Australia. The results show high soil moisture estimation accuracy for the EnKF and EDA methods when compared with the open loop estimates during evaluation and validation stages. The level of convergence was quantified for each model parameter in the EDA approach to illustrate its potential in the retrieval of variables that were not directly observed. The EDA was found to have a higher estimation accuracy than the EnKF when its updated members were evaluated against the SMOS level 2 soil moisture. However, the EnKF and EDA estimations are comparable when their forward soil moisture estimates were validated against SMOS soil moisture outside the assimilation time period. This suggests that parameter convergence does not significantly influence soil moisture estimation accuracy for the EnKF. However, the EDA has the advantage of simultaneously determining the convergence of model parameters while providing comparably higher accuracy for soil moisture estimates.

1. Introduction

Data assimilation (DA) methods are used widely for finding a compromise between imperfect observations and uncertain model outputs. Generally, the DA procedure encompasses three major components: (i) accurate estimation of the model state, (ii) determination of the measurement/observation errors, and (iii) estimation of model parameter values. The model parameter estimation has a direct impact on the simulated outputs and thus significantly influences the forward model runs. Also, the model state and parameter estimation by ensemble members is closely linked to the distribution of model parameter values, which is indicative of their level of convergence. However, most DA methods are applied with the assumption that the model simulations

and the observation function are free of systematic errors (Dumedah and Coulibaly 2013a; Su et al. 2011).

Several studies, including He et al. (2012), Su et al. (2011), and Dumedah et al. (2011), have shown that model parameter errors can result in significant differences between the model prediction and observation. To account for several error sources in hydrologic models, Vrugt et al. (2005a,b) have proposed a simultaneous optimization and data assimilation (SODA) procedure to merge the search capabilities of the Shuffled Complex Evolution Metropolis algorithm with the ensemble Kalman filter (EnKF), in order to estimate both model parameters and states. Moreover, Andreadis et al. (2008) have examined the influence of model parametric uncertainty on multiscale snow simulation, and Dumedah et al. (2012) have assessed the time-variant properties of model parameters in streamflow estimation. Nonetheless, the contribution of model parameter convergence in relation to the estimation accuracy of DA methods has not been thoroughly examined in the DA literature. The convergence of model parameters influences the merging of

Corresponding author address: Jeffrey Walker, Building 60, Department of Civil Engineering, Monash University, Wellington Road, Clayton VIC 3800, Australia.
E-mail: jeff.walker@monash.edu

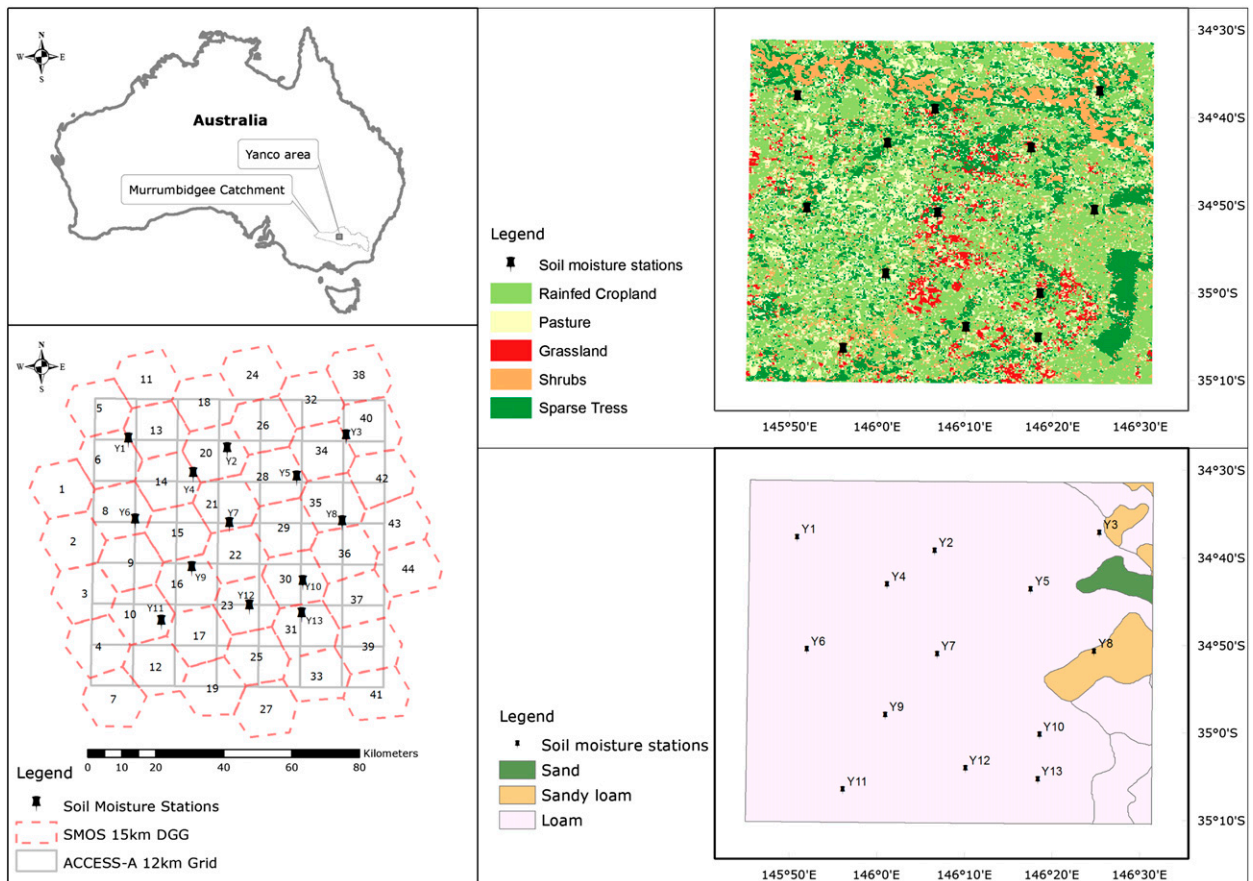


FIG. 1. The Yanco experimental area, showing the dominant land cover types, soil texture classes, the SMOS 15-km DGG, the 12-km ACCESS configuration, and the ground soil moisture stations (data sources: the Australian Bureau of Meteorology, Geoscience Australia, and Commonwealth Scientific and Industrial Research Organization).

observations with model predictions and thus the estimation accuracy of DA methods. In particular, the assessment of model parameter convergence across several assimilation time steps can provide the potential to retrieve variables that are not directly observed.

This study examines the contributions of model parameter convergence to overall performance of the EnKF and the evolutionary data assimilation (EDA) approaches, through estimation of soil moisture using the Joint UK Land Environment Simulator (JULES) in the Yanco area located in southeast Australia. The EnKF and EDA methods were employed to assimilate the Soil Moisture and Ocean Salinity (SMOS) level 2 soil moisture data into the JULES model. The distribution of JULES parameter values associated with the updated ensemble members for EnKF and EDA were evaluated, together with their contribution to soil moisture estimation. These findings are important to refine DA procedures through multi-objective evolutionary strategies to adequately account for contributions from convergence of model parameters.

The EnKF and the EDA have standard analytical procedures that are well documented in the literature. The two methods are briefly introduced in this section, but their implementations are outlined in section 2. The EnKF is a Monte Carlo integration approach that estimates the posterior density function (pdf) of the model states using randomly generated ensemble members (Evensen 2003, 1994; Burgers et al. 1998; Houtekamer and Mitchell 1998). The EnKF is popular and has been applied in numerous studies (Pipunic et al. 2011; Xie and Zhang 2010; Weerts et al. 2010; Clark et al. 2008; Weerts and El Serafy 2006; Moradkhani and Hsu 2005). The EDA is a relatively new multi-objective formulation of the evolutionary algorithm (EA) applied in a data assimilation framework (Dumedah and Coulibaly 2013a, 2014a; Dumedah 2012). The EDA combines stochastic and adaptive capabilities of the EA in a multi-objective fashion, together with the cost function from variational DA. Its population-based approach allows the selection of a subset of updated members from the entire ensemble membership, usually chosen from a parameter convergent population.

TABLE 1. Description of selected model parameters, forcing, and state variables for the JULES model. These model parameter intervals were defined in concert with land cover, soil, and meteorological forcing data in the Yanco area.

Parameter	Description	Interval (%)
Model parameters		
b	Exponent in soil hydraulic characteristics curve	± 5
Sathh	Absolute value of the soil matric suction at saturation (m)	± 5
Hsatcon	Hydraulic conductivity at saturation ($\text{kg m}^{-2} \text{s}^{-1}$)	± 5
sm-sat	Volumetric soil moisture content at saturation (m^3 water per m^3 soil)	± 5
sm-crit	Volumetric soil moisture content at critical point (m^3 water per m^3 soil)	± 5
sm-wilt	Volumetric soil moisture content at wilting point (m^3 water per m^3 soil)	± 5
Hcap	Dry heat capacity ($\text{J m}^{-3} \text{K}^{-1}$)	± 5
Hcon	Dry thermal conductivity ($\text{W m}^{-1} \text{K}^{-1}$)	± 5
Albsoil	Soil albedo	± 5
Meteorological forcing variables		
SWR	Downward component of shortwave radiation at the surface (W m^{-2})	± 5
LWR	Downward component of longwave radiation at the surface (W m^{-2})	± 5
Rain	Rainfall ($\text{kg m}^{-2} \text{s}^{-1}$)	± 5
Snow	Snowfall ($\text{kg m}^{-2} \text{s}^{-1}$)	± 5
Tempr	Atmospheric temperature (K)	± 5
Wind	Wind speed (m s^{-1})	± 5
Press	Surface pressure (Pa)	± 5
spHum	Atmospheric specific humidity (kg kg^{-1})	± 5
Model state variables		
Canopy	Amount of intercepted water that is held on each tile (kg m^{-2})	Updated
tstar-t	Surface or skin temperature of each tile (K)	Updated
t-soil	Temperature of each soil layer (K)	Updated
Sthuf	Soil wetness for each soil layer; mass of soil water expressed as a fraction of water content at saturation	Updated

2. Materials and methods

a. Study area, datasets, and the land surface model

The Yanco area is located in the western plains of New South Wales, Australia (shown in Fig. 1). In this part of the Murrumbidgee catchment, the topography is flat, with scattered geological outcroppings. The soil texture class is predominantly loam, along with scattered clays, red brown earths, transitional red brown earth, sands over clay, and deep sands (McKenzie et al. 2000; McKenzie and Hook 1992). Information in the Digital Atlas of Australian Soils shows that the dominant soil landscape is characterized by plains with domes, lunettes, and swampy depressions, divided by discontinuous low river ridges associated with prior stream systems (McKenzie et al. 2000). The area is traversed with stream valleys, containing layered soil and sedimentary materials that are common at fairly shallow depths. The land cover is predominantly rain-fed cropping–pasture with scattered trees and grassland, with irrigated crops in the northwest.

The land surface model used in this study to simulate soil moisture is the JULES model—a widely used tiled model of subgrid heterogeneity that simulates water and energy fluxes between a vertical profile of soil layers, vegetation, and the atmosphere (Best et al. 2011). The JULES model uses meteorological forcing data, surface

land cover data, soil properties data, and values for prognostic variables. The soil properties data were derived from the Digital Atlas of Australian Soils (McKenzie et al. 2000), obtained through the Australian Soil Resource Information System. The soil data include information on soil texture class, along with proportion of clay content, bulk density, saturated hydraulic conductivity, and soil layer thickness for horizons A and B (McKenzie et al. 2000; McKenzie and Hook 1992). Information on land cover categories was obtained from the Australian National Dynamic Land Cover dataset (Lyburner et al. 2011), which was derived from the 250-m bands of the Moderate Resolution Imaging Spectroradiometer (MODIS).

The forcing data variables required in the JULES model include shortwave and longwave incoming radiation, air temperature, precipitation, wind speed, atmospheric pressure, and specific humidity. The forcing data were obtained from the Australian Community Climate and Earth-System Simulator–Australia (ACCESS-A) at hourly time steps with 12-km spatial resolution (Bureau of Meteorology 2010). The ACCESS-A precipitation dataset was bias corrected using the daily 5-km gridded rain gauge precipitation data from the Australian Water Availability Project obtained through the Bureau of Meteorology, herein denoted BAWAP (Jones et al. 2007, 2009). The land cover and soil data were mapped

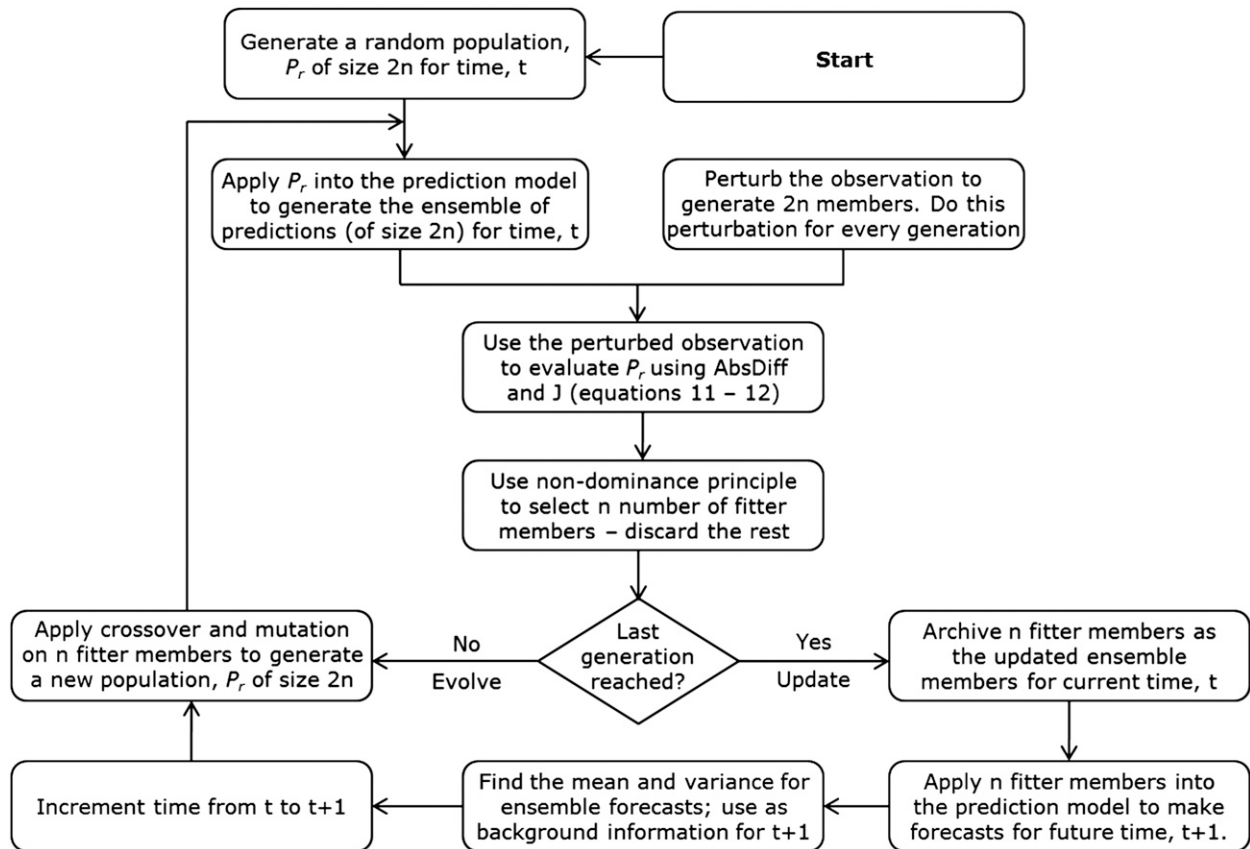


FIG. 2. Computational procedure for a sequential assimilation using the EDA method (adapted from Dumedah 2012).

to the 12-km ACCESS-A grids through spatial overlap and subsequent determination of the proportions of constituent land cover and soil classes within each grid. The forcing data together with the land cover and soil data were incorporated into JULES to simulate the temporal evolution of soil moisture.

The description of model parameters, forcing variables, and their associated uncertainty intervals within which they were modified in the JULES model are presented in Table 1. All of the forcing variables and model parameters in Table 1 were modified using a relative percentage measure, such that a $\pm\%$ uncertainty bound means that the specified variable or parameter was modified to within a maximum of $+\%$ and a minimum of $-\%$ of its original value. Thus, the modified values vary within a predetermined $\pm\%$ of the original values based on the soil and vegetation data. It is noteworthy that the perturbed values for parameters–states were constrained to within intervals acceptable to the JULES model in the context of data in the Yanco area. The original values for model parameters and forcing variables were determined based on the soil, land cover, and meteorological forcing data such that they are

physically meaningful for the JULES model in the context of the Yanco area. Note that the model parameters were at the same scale as the 12-km model resolution.

The observation dataset used to drive the assimilation is the SMOS level 2 soil moisture. The SMOS level 2 soil moisture is an estimate of SMOS-retrieved soil moisture at the reported 15-km discrete global grid (DGG). These SMOS data are the Soil Moisture level 2, version 4.0, User Data Product (SMUDP2), which was obtained for the period from January to December 2010. The SMOS level 2, version 4.0, soil moisture was retrieved using the Mironov model (Mironov et al. 2004). It is noted that the SMOS dataset was used at the 15-km DGG based on the findings from Dumedah et al. (2014), which showed that the error involved in representing 42-km SMOS observations at the 15-km DGG is no worse than the noise that currently exists in the original SMOS data.

b. The ensemble Kalman filter method

The EnKF procedure used in this study is based on the state–parameter estimation approach of Moradkhani et al. (2005), where state variables and model parameters

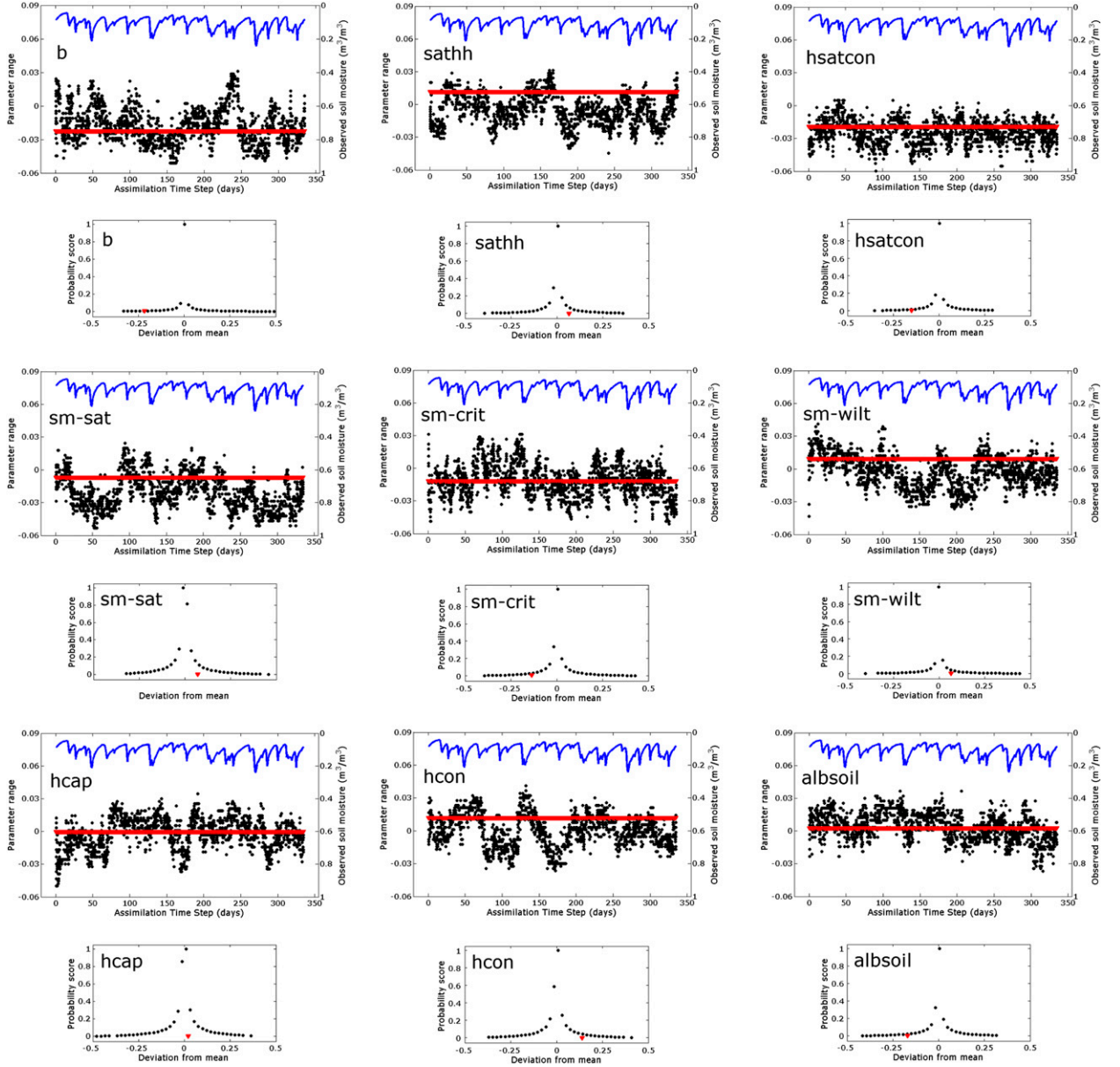


FIG. 3. Retrieval of the original truth model parameter values using the EDA method. Top two rows (left to right) parameters *b*, *Sathh*, and *Hsatcon* with the top row giving the parameter range as a function of time (days) and the bottom row showing the probability score vs deviation from the mean. The middle and bottom two rows are as in the top two but for parameters *sm-sat*, *sm-crit*, and *sm-wilt* in the middle and *Hcap*, *Hcon*, and *Albsoil* in the bottom. The red triangles in the smaller panels and the red horizontal lines indicate the original values, and the black circles represent the estimated EDA values. The synthetic soil moisture observation is shown in the larger panels on the right vertical axis.

were determined jointly. Accordingly, the soil moisture state was determined using Eq. (1):

$$\mathbf{x}_t = f_1[\mathbf{x}_{t-1}, \mathbf{u}_{t-1}, \mathbf{z}_{t-1}] + \boldsymbol{\omega}_{t-1}. \quad (1)$$

The ensemble members for model parameters evolve according to Eq. (2):

$$\begin{aligned} \mathbf{z}_t^- &= \mathbf{z}_{t-1}^+ + \mathbf{v}_{t-1}, \\ \mathbf{v}_{t-1} &\sim N(0, \boldsymbol{\beta}_{t-1}^z), \end{aligned} \quad (2)$$

where $\mathbf{x}_t$ is a vector of state variables at time t ; $f_1[\cdot]$ is the system transition function (i.e., the JULES model) that comprises the state vector at previous time $\mathbf{x}_{t-1}$, a vector

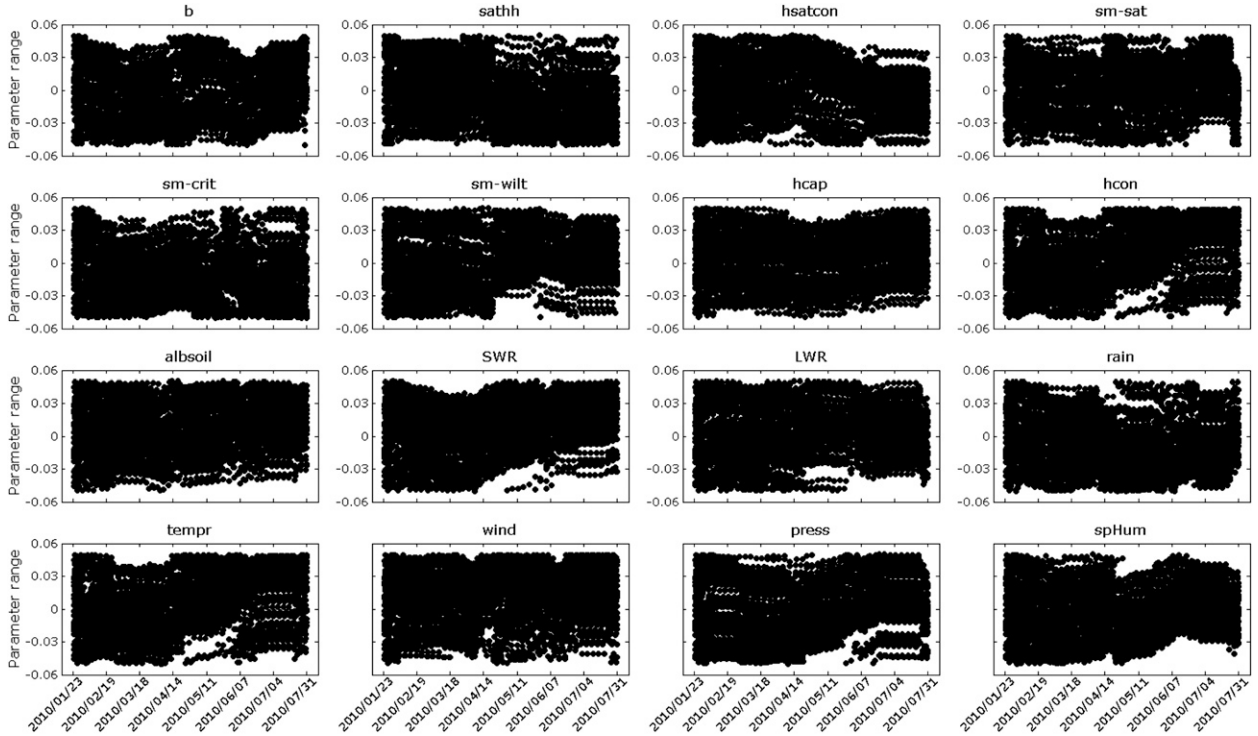


FIG. 4. Distribution of model parameter and forcing values for updated members obtained across assimilation time steps using the EnKF method. Note that the EnKF has 200 updated members at each assimilation time step so these are plotted for each time step.

of forcing data $\mathbf{u}_{t-1}$, and a vector of time-variant model parameters $\mathbf{z}_{t-1}$; $\boldsymbol{\omega}_{t-1}$ is the system noise representing all errors associated with model structure; $\mathbf{z}_t^-$ is a vector of model parameters for time t ; $\mathbf{z}_{t-1}^+$ is a vector of updated model parameters for the previous time; and $\mathbf{v}_{t-1}$ is the model parameter error with covariance $\boldsymbol{\beta}_{t-1}^z$. The model parameter ensemble members were combined with perturbed forcing data to determine the state ensemble in Eq. (3):

$$\mathbf{x}_t^- = f[\mathbf{x}_{t-1}^+, \mathbf{u}_{t-1}^*, \mathbf{z}_t^-], \quad (3)$$

where $\mathbf{x}_t^-$ is a vector of forecasted states at time t and $\mathbf{x}_{t-1}^+$ is a vector of updated states for the previous time. The forcing data were perturbed by adding the noise $\boldsymbol{\gamma}_t$ with covariance $\boldsymbol{\beta}_t^u$ at each time step to generate its ensemble according to Eq. (4):

$$\begin{aligned} \mathbf{u}_t^* &= \mathbf{u}_t + \boldsymbol{\gamma}_t, \\ \boldsymbol{\gamma}_t &\sim N(0, \boldsymbol{\beta}_t^u). \end{aligned} \quad (4)$$

The JULES model was run forward in time to determine the ensemble predictions ($\hat{\mathbf{y}}_t$) in Eq. (5):

$$\hat{\mathbf{y}}_t = f_2[\mathbf{x}_t^-, \mathbf{z}_t^-], \quad (5)$$

where f_2 represents the JULES prediction. The observation was perturbed to generate an ensemble of observations $\mathbf{y}_t$ according to Eq. (6):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{y}_t^- + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\varepsilon}_t &\sim N(0, \boldsymbol{\beta}_t^y), \end{aligned} \quad (6)$$

where $\boldsymbol{\varepsilon}_t$ is the observation noise with covariance $\boldsymbol{\beta}_t^y$. The ensemble predictions ($\hat{\mathbf{y}}_t$) were combined with perturbed observations $\mathbf{y}_t$ to determine the Kalman gain ($\mathbf{K}$) functions separately for model parameters ($\mathbf{K}_t^z$) in Eq. (7),

$$\mathbf{K}_t^z = \boldsymbol{\beta}_t^{zy} (\boldsymbol{\beta}_t^{yy} + \boldsymbol{\beta}_t^y)^{-1}, \quad (7)$$

and for states ($\mathbf{K}_t^x$) in Eq. (8),

$$\mathbf{K}_t^x = \boldsymbol{\beta}_t^{xy} (\boldsymbol{\beta}_t^{yy} + \boldsymbol{\beta}_t^y)^{-1}, \quad (8)$$

where $\boldsymbol{\beta}^{zy}$ is the covariance of model parameter ensemble $\mathbf{z}_t^-$ and the ensemble predictions $\hat{\mathbf{y}}_t$; $\boldsymbol{\beta}^{yy}$ is the forecast error covariance for ensemble predictions $\hat{\mathbf{y}}_t$; and $\boldsymbol{\beta}^{xy}$ is the covariance of the model state and prediction ensemble. The model parameters and states were directly updated using their respective Kalman gain functions and the innovation vector according to Eqs. (9) and (10), respectively:

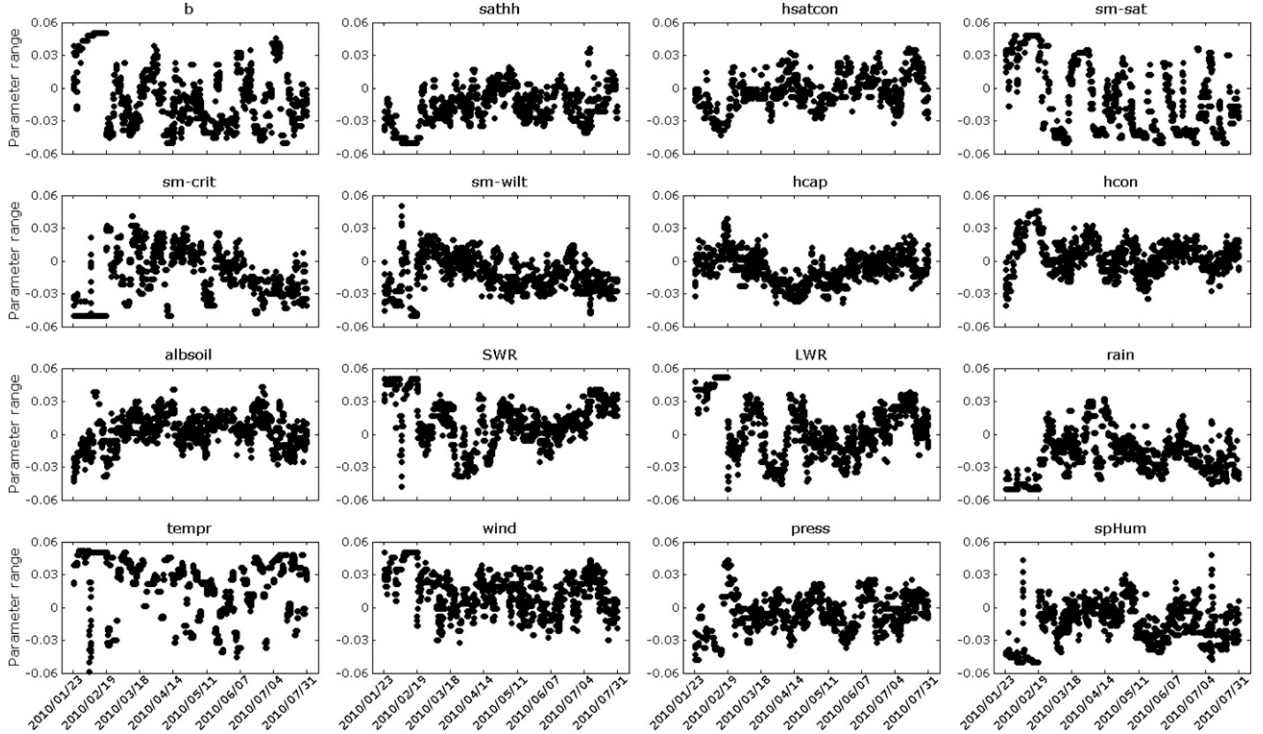


FIG. 5. As in Fig. 4, but using the EDA method. Note that the EDA has 20 updated members at each assimilation time step plotted for each time step.

$$\mathbf{z}_t^+ = \mathbf{z}_t^- + \mathbf{K}_t^z(\mathbf{y}_t - \hat{\mathbf{y}}_t) \quad (9)$$

and

$$\mathbf{x}_t^+ = \mathbf{x}_t^- + \mathbf{K}_t^x(\mathbf{y}_t - \hat{\mathbf{y}}_t). \quad (10)$$

c. The evolutionary data assimilation method

The EDA employs an evolutionary strategy based on the nondominated sorting genetic algorithm II (NSGA-II). The NSGA-II is an established multi-objective evolutionary algorithm that has been used in several hydrological applications (Confesor and Whittaker 2007; Efstratiadis and Koutsoyiannis 2010; Dumedah and Coulibaly 2014b; Dumedah et al. 2011; Tang et al. 2006). The EDA has been applied in soil moisture estimation (Dumedah and Coulibaly 2013a; Dumedah et al. 2011) and to facilitate streamflow estimations (Dumedah and Coulibaly 2012, 2013b; Dumedah 2012). A schematic outline of the EDA procedure is presented in Fig. 2.

In the EDA procedure, a random population P_r of size $2n$ was generated at the initial assimilation time step t , consisting of ensemble members that comprise model states, model parameters, and forcing data uncertainties. The variable n represents the number of updated members to be selected for each assimilation time step. As in

the EnKF procedure, the model states were obtained according to Eq. (3), and the forcing data perturbed using Eq. (4). Each member of P_r was applied into the JULES model to determine the predictions in Eq. (5). Similarly, the observation was perturbed using Eq. (6) to generate $2n$ members of the observation ensemble.

The members in P_r were evaluated using the absolute difference (AbsDiff) in Eq. (11),

$$\text{AbsDiff} = |\hat{y}_i - y_{o,i}|, \quad (11)$$

and the cost function (J) in Eq. (12),

$$J = \sum_{i=1}^t J(\hat{y}_i) = \sum_{i=1}^t \left[\frac{(\hat{y}_i - y_{b,i})^2}{\sigma_b^2} + \frac{(\hat{y}_i - y_{o,i})^2}{\sigma_o^2} \right], \quad (12)$$

where $y_{b,i}$ is the background value for the i th data point, $y_{o,i}$ is the perturbed observation value for the i th data point, σ_b^2 is the variance for the background value, σ_o^2 is the variance for the observation, $\hat{y}_i$ is the analysis (i.e., the searched) value for i th data point that minimizes $J(\hat{y}_i)$, and t is the number of data points or the time (note that $t = 1$ in this case for sequential assimilation).

The AbsDiff estimates the absolute residual/difference between the predicted and observed value, whereas J determines a compromised value between the background

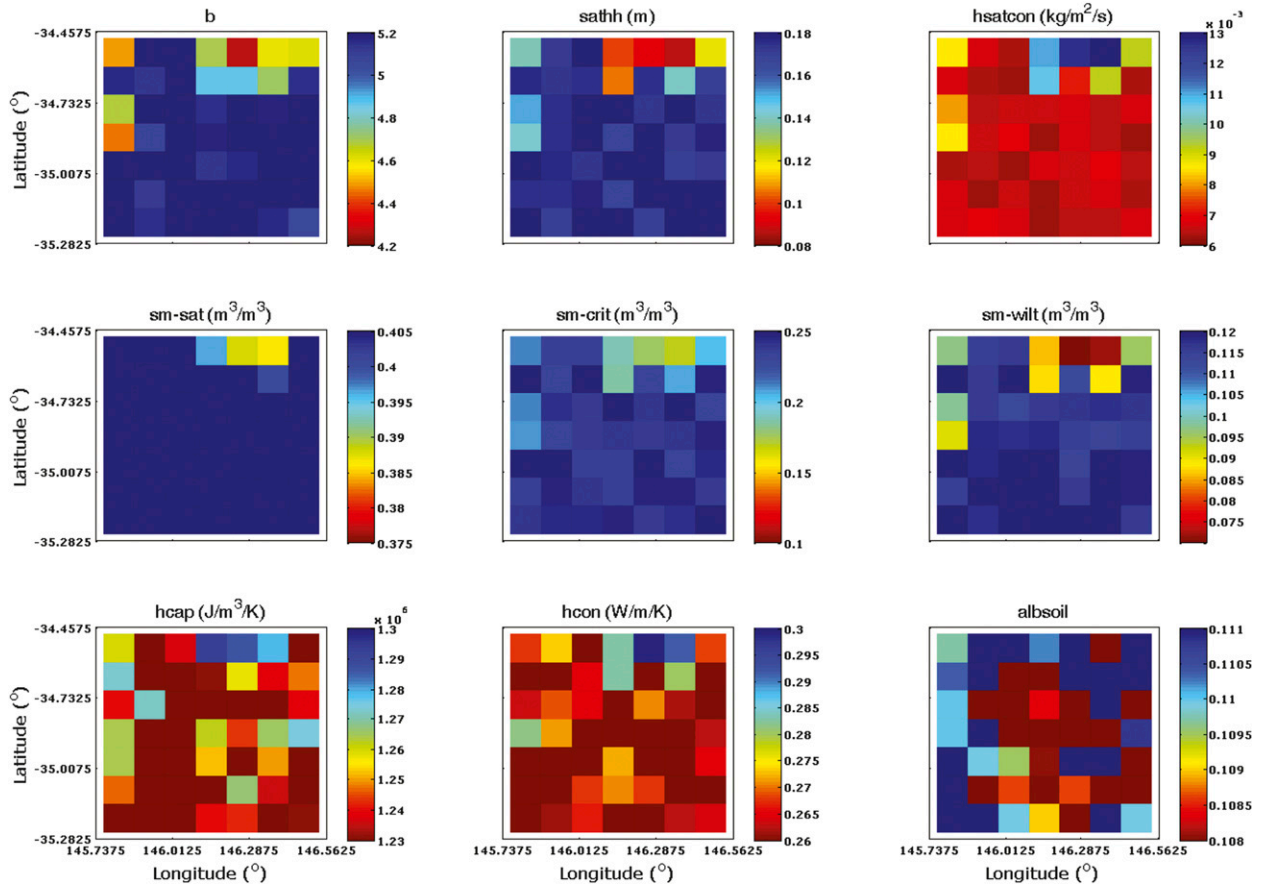


FIG. 6. Comparison between model parameter values from the EnKF and EDA methods for the median ensemble member at the end of the assimilation period, showing the spatial latitude–longitude distribution of model parameter values across model grids.

and the observed value. The background value is the ensemble average of soil moisture estimated by applying updated members of the population from the previous assimilation time step into the JULES model to make a prediction for the current assimilation time step. It is noteworthy that the background value for the initial time step was determined from a randomly generated population of members.

The minimization of AbsDiff and J allows the ranking of the members into different nondominance levels through Pareto dominance. The top n fittest members were selected, varied, and recombined through crossover and mutation to determine new members for the population P_r of size $2n$. The above procedures were repeated to evolve the population P_r through several generations where each generation improved/preserved the overall quality of members in P_r . At the referenced generation, the final n fittest members, which are a subset of all evaluated members, were chosen as the updated members where they were archived into the population P_e .

To increment the assimilation time step from t to $t + 1$, the n members in P_e at t were used to generate n number of forecasts of soil moisture for future time $t + 1$. The estimated ensemble average and its associated variance from the ensemble members were used as background information. The P_e from t was also used as a seed population for $t + 1$, where it was varied and recombined to generate a new population P_r of size $2n$. At $t + 1$, the new P_r was again evolved through several generations to determine the final n updated members for $t + 1$. These procedures were repeated for all assimilation time steps to evolve members through several generations and to determine the updated members.

d. Setup of model and data assimilation runs

The procedures for the EnKF and the EDA methods were used to assimilate SMOS level 2 soil moisture into the JULES model. For the initial assimilation time step, the uncertainties for model parameters and forcing variables presented in Table 1 were used. For subsequent assimilation time steps, the uncertainties were determined

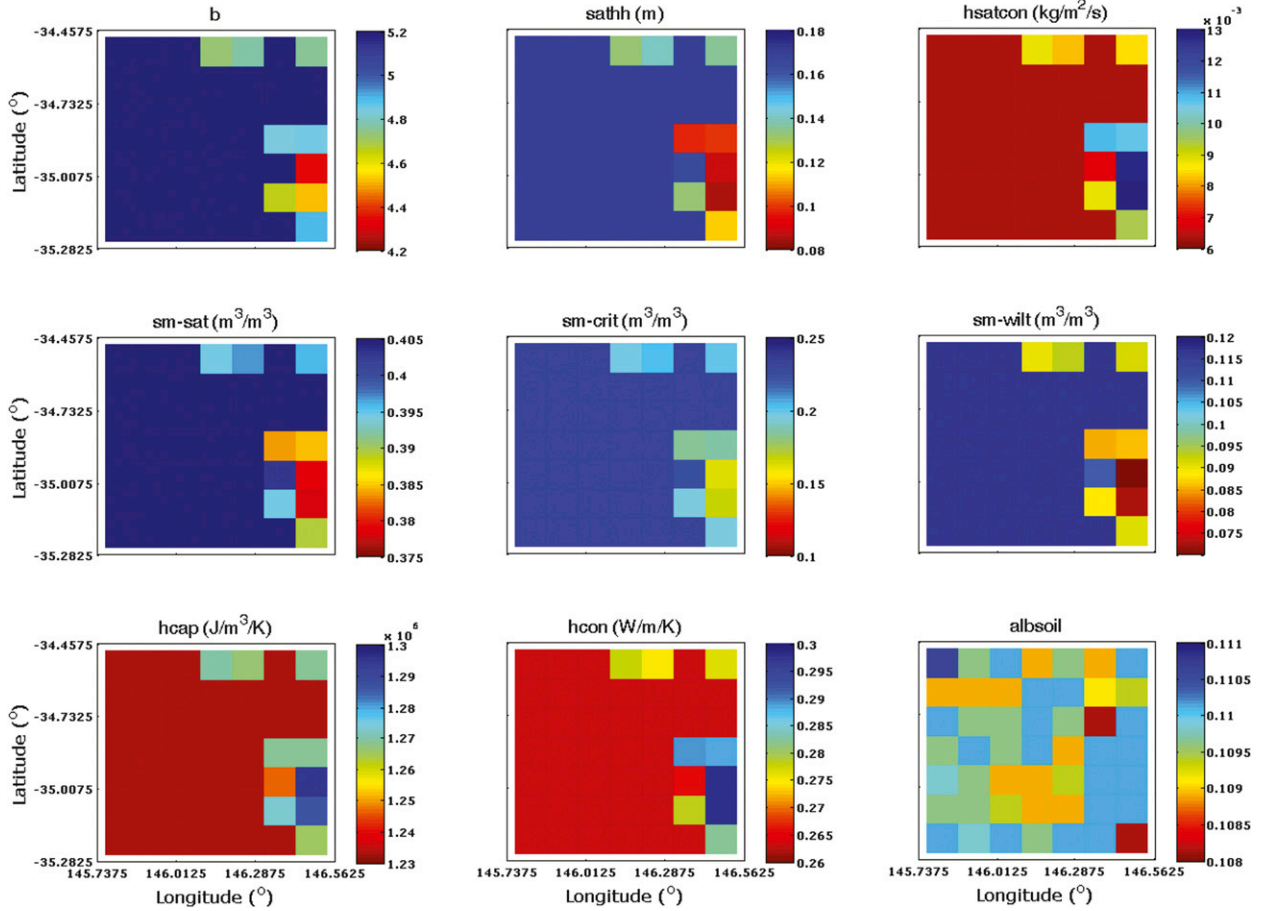


FIG. 6. (Continued)

from the ensemble members, but these uncertainties were constrained to the lower and upper bounds specified in Table 1. The soil moisture error that comes along with the SMOS level 2 soil moisture was used as the observation uncertainty. It is noted that the error in the SMOS level 2 soil moisture is only associated with model inversion error and not the sensor error. In the EnKF, the model error and the variance of parameter values were determined from the ensemble members. The time-variant model error in the EDA was estimated adaptively using the background error estimation procedure outlined in section 2c. The initial population was generated by using the model parameter bounds and the forcing data uncertainties in Table 1. As in the standard NSGA-II procedure, the crossover probability of 0.8 and a mutation probability of $1/m$ (where m is the number of variables) were used in the variation and recombination procedures.

The assimilation was run at a daily interval with a 200-member ensemble from January to July 2010. To evaluate 200 members, the EDA divides the ensemble size of

200 into smaller populations of 40 members (i.e., $2n = 40$), where they were evolved through five generations. A subset of $n = 20$ members was determined from the final population for each assimilation time step to represent the updated members. As a result, the number of updated members was 200 for the EnKF, and 20 for the EDA method for each assimilation time step.

3. Results and discussion

a. Preliminary test of convergence for the EDA method

Prior to presenting the modeling outputs with actual SMOS data, a preliminary experiment was undertaken to illustrate the capability of the EDA procedure to retrieve model state and parameter values. The experiment is synthetic in a way that “truth” values of initial model state and parameters were generated randomly based on the parameter bounds in Table 1 and were subsequently applied into the JULES model to obtain

TABLE 2. Converged model parameter intervals from the EDA method, represented by the largest membership clusters. The definition of their coverage of parameter space is given by their centroids and lower and upper bounds. The coverage is presented as a fraction, where a maximum value of unity represents a perfectly converged cluster and a value close to zero represents a sensitive cluster.

Parameter	Centroid	Lower bound	Upper bound	Coverage
Model parameters				
b	0.04160	0.03210	0.0499	0.4419
Sathh	−0.0310	−0.0478	0.0077	0.9795
Hsatcon	−0.0079	−0.0278	0.0299	0.7844
sm-sat	0.00590	−0.0278	0.0299	0.4672
sm-crit	−0.0124	−0.0433	0.0299	0.8395
sm-wilt	−0.0091	−0.0478	0.0277	0.9946
Hcap	−0.0037	−0.0411	0.0344	0.9941
Hcon	0.00920	−0.0278	0.0455	0.9921
Albsoil	0.00250	−0.0278	0.0299	0.9396
Meteorological forcing variables				
SWR	0.03560	−0.0078	0.0499	0.8744
LWR	0.04220	0.03210	0.0499	0.5377
Rain	−0.0079	−0.0278	0.0299	0.5483
Tempr	0.04060	−0.0034	0.0499	0.9174
Wind	0.02930	−0.0167	0.0499	0.9862
Press	−0.0136	−0.0433	0.0299	0.7927
spHum	−0.0217	−0.0478	0.0277	0.9965

a synthetic surface (0–8 cm) soil moisture from January to December 2010. This synthetic soil moisture dataset was then used as the observation to drive a daily time step soil moisture assimilation into the JULES model using the EDA. The EDA employed time variant model state and parameter values such that their updated estimates were obtained for each assimilation time step.

The ensemble estimate of soil moisture from the EDA was compared to the original synthetic soil moisture observation, with a root-mean-square error (RMSE) of about $0.02 \text{ m}^3 \text{ m}^{-3}$ and R^2 of 0.75. Moreover, the updated model parameter estimates from the EDA are compared to the original truth values in Fig. 3. These results illustrate the EDA method to both retrieve model parameter values and accurately estimate soil moisture. The results also raise theoretical challenges such as those identified in Vrugt and Sadegh (2013) and Gupta et al. (2008) for diagnostic model evaluation.

As illustrated in Fig. 3, the model parameter values change between assimilation time steps in response to changes in observation data. It is noted that the changes in model parameter values are not linked to changes in the watershed, but that the changes are constrained to the landscape properties in the study area. The JULES model parameter values used were determined from soil and land cover data, and the interval bounds are physically meaningful in the context of the JULES model. In essence, each ensemble member that has been evaluated is valid in terms of the landscape properties and is not some random scenario to simply obtain a good agreement between observation and model output. Based on these original model parameter values and their intervals (both determined from soil–land cover data), a large population of members (or samples) can be determined. No member chosen from this population should be deemed statistically invalid because it satisfies some conditions at a specific space and time.

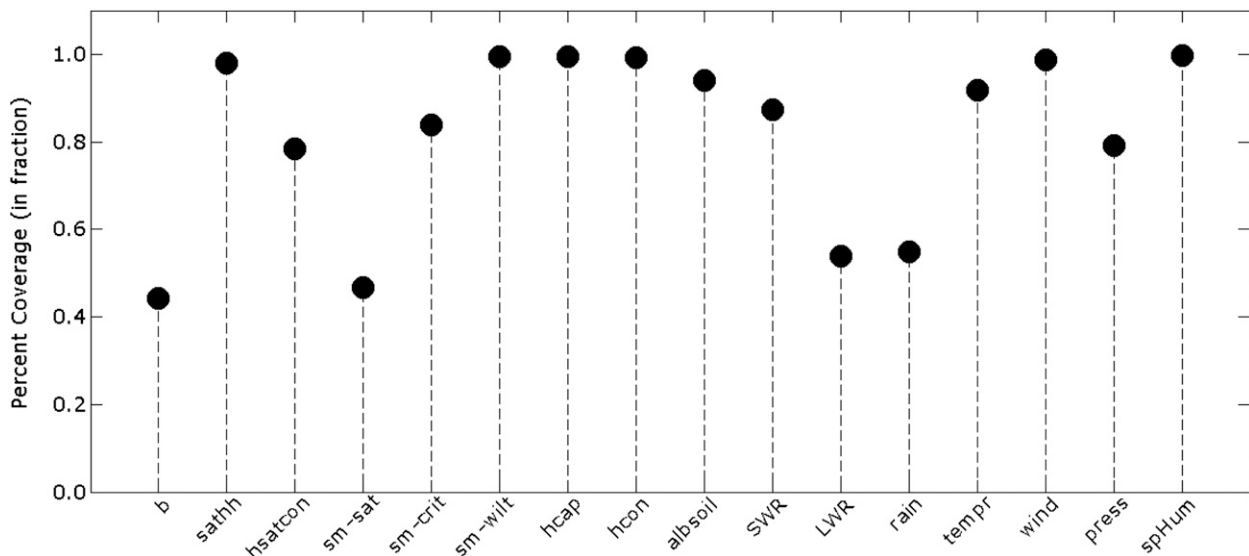


FIG. 7. Proportion of EDA-converged forcing and model parameters as indicated by the coverage (in fraction) of largest membership clusters. The coverage is estimated as the ratio of the number of cluster members to the total number of members across all assimilation time steps.

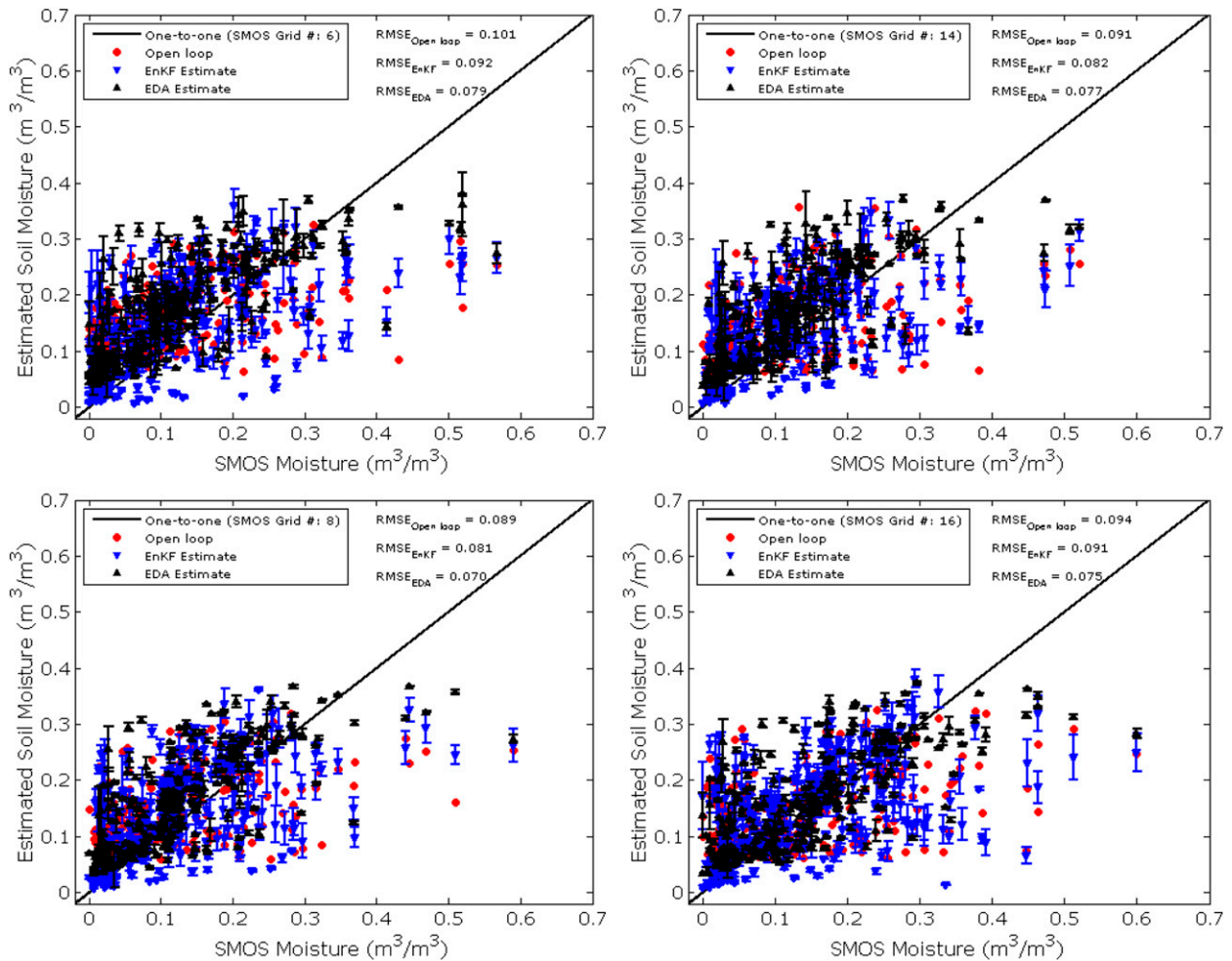


FIG. 8. Evaluation of the open loop estimate of surface soil moisture together with the updated ensemble estimates, based on the mean and standard deviation (for the error bars) of the ensemble from the EnKF and the EDA compared against the SMOS soil moisture for four different SMOS grids: Grid numbers (top left) 6 and (top right) 14; and (bottom left) 8 and (bottom right) 16.

In the EDA procedure, the updated members were chosen from the physically meaningful population because they meet some set conditions in space and time. The degree to which the chosen members represent meaningful parameter values in the JULES model was based on the landscape properties data. This is because the lack of observations for model parameters, together with the problem of nonuniqueness and the inherent complexity in land surface modeling, mean that the closest proof of statistical validity of chosen members is based on proper representation of the landscape properties.

It is noted that more experimentation is needed to diagnose model deficiencies through further evaluation of the updated ensemble members. While the temporal evaluation on a parameter-by-parameter basis is important, the interconnectedness between model parameters is crucial for further examination of model weaknesses. These evaluations have the potential to quantify the

combined temporal changes in model parameters in relation to the changes in observation data.

b. Convergence of model parameter ensemble from the EnKF and the EDA

The convergence of model parameters in the EnKF and the EDA is examined through the distribution of model parameter values for the updated ensemble members across assimilation time steps. The distribution of parameter values for updated members across the assimilation time steps are shown in Fig. 4 for the EnKF method and in Fig. 5 for the EDA method.

In the EnKF output, the parameter values are distributed between the lower and upper bounds of the defined model parameter range. This pattern is consistent across all model parameters and from one assimilation time step to the next. The near-uniform distribution of model parameter values show that the ensemble

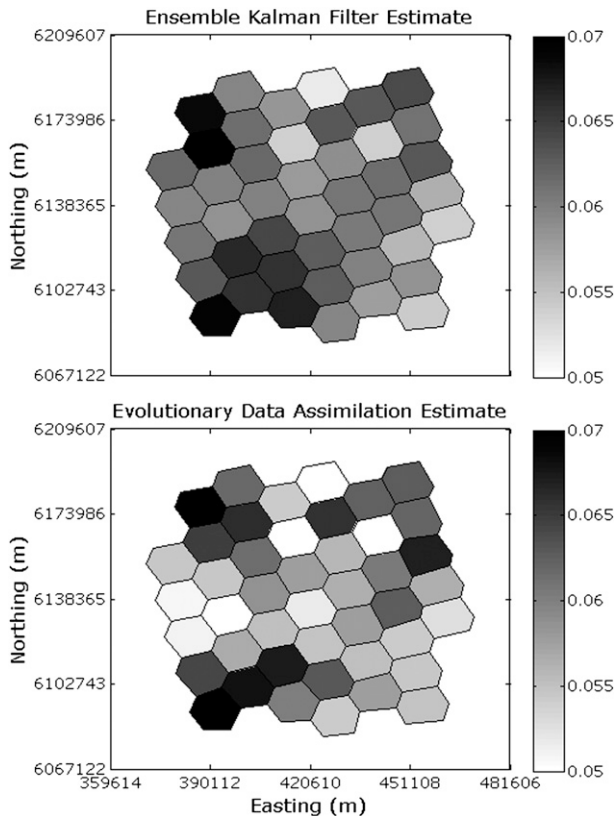


FIG. 9. The overall absolute difference between the SMOS soil moisture ($\text{m}^3 \text{m}^{-3}$) and the updated ensemble estimations of soil moisture from (top) the EnKF and (bottom) the EDA method across all assimilation time steps.

predictions were obtained from values within the entire range for each model parameter. In other words, the ensemble predictions can account for the uncertainty of model parameters based on the entire interval between the lower and upper bounds. As a result, the parameter distributions for updated members do not converge in the EnKF approach, and the test of clustering according to Thorndike (1953) fails. It is noted that the independence of the errors associated with the state variables and model parameters may partly influence the non-convergence found in the EnKF. The complex dynamics between model state and parameter values, and the associated lack of observation for model parameters, mean that statistical validation of the output is difficult. Thus, the physical meaning of the chosen members is preferable, with the capability to account for the landscape properties.

The EDA output shows a clustering of the parameter values and, more importantly, a convergence of model parameters. Overall, the pattern of parameter clusters is consistent across assimilation time steps in a way that cluster locations are almost predictable between assimilation

time steps. The recurrence of these cluster groups show that the converged parameter values for one assimilation time step are consistent and applicable to other assimilation time steps. In the EDA, these updated members represent model parameter values with the optimal compromise between ensemble predictions and observations.

The spatial distribution of the model parameter values across the 12-km model grids is shown in Fig. 6 for the median ensemble member from the EnKF and EDA methods at the end of the July 2010 assimilation time step. A comparison of this spatial distribution to the soil texture groups in Fig. 1 shows that the model parameter values from the EDA are fairly consistent to soil texture classes spatially across the model grids, whereas there is a poor agreement with those derived from the EnKF.

To quantify the level of convergence of model parameters from the EDA, a clustering analysis was performed to evaluate the persistence of cluster groups across all assimilation steps for each model parameter. The clustering analysis was conducted on the ensemble parameter values where the appropriate number of clusters was determined using the “knee” procedure in Thorndike (1953). The number of cluster groups examined when determining the appropriate number of clusters typically varied between four and eight. The cluster with the largest membership was determined along with its coverage of the parameter space, with the centroid and the lower and upper bounds representing the converged parameter space with the largest weight. The largest membership cluster for each model parameter across all assimilation time steps is shown in Table 2. The coverage of parameter space represents the proportion of members in the largest membership cluster in relation to the total number of members across all assimilation time steps. The coverage, therefore, quantifies the weight of the cluster with largest membership and accounts for variability of cluster memberships due to different cluster groupings. The coverage representing the level of convergence for each model parameter across all assimilation time steps is shown in Fig. 7.

The convergence of model parameter values shown in Table 2 and Fig. 7 for the EDA output is significant. Across all assimilation time steps the uncertainty applied to rainfall was found to be between -2.8% and 3% with a coverage of about 55%. About 12 model parameters out of 16 converged to within about 80% coverage across all assimilation time steps. The evaluation also identified model parameters b and sm-sat , and forcing variables LWR and rain, as having the least convergence and therefore the most sensitivity across different observation/assimilation time steps. The high level of convergence for the 12 model parameters means that their clustered intervals are consistently reliable

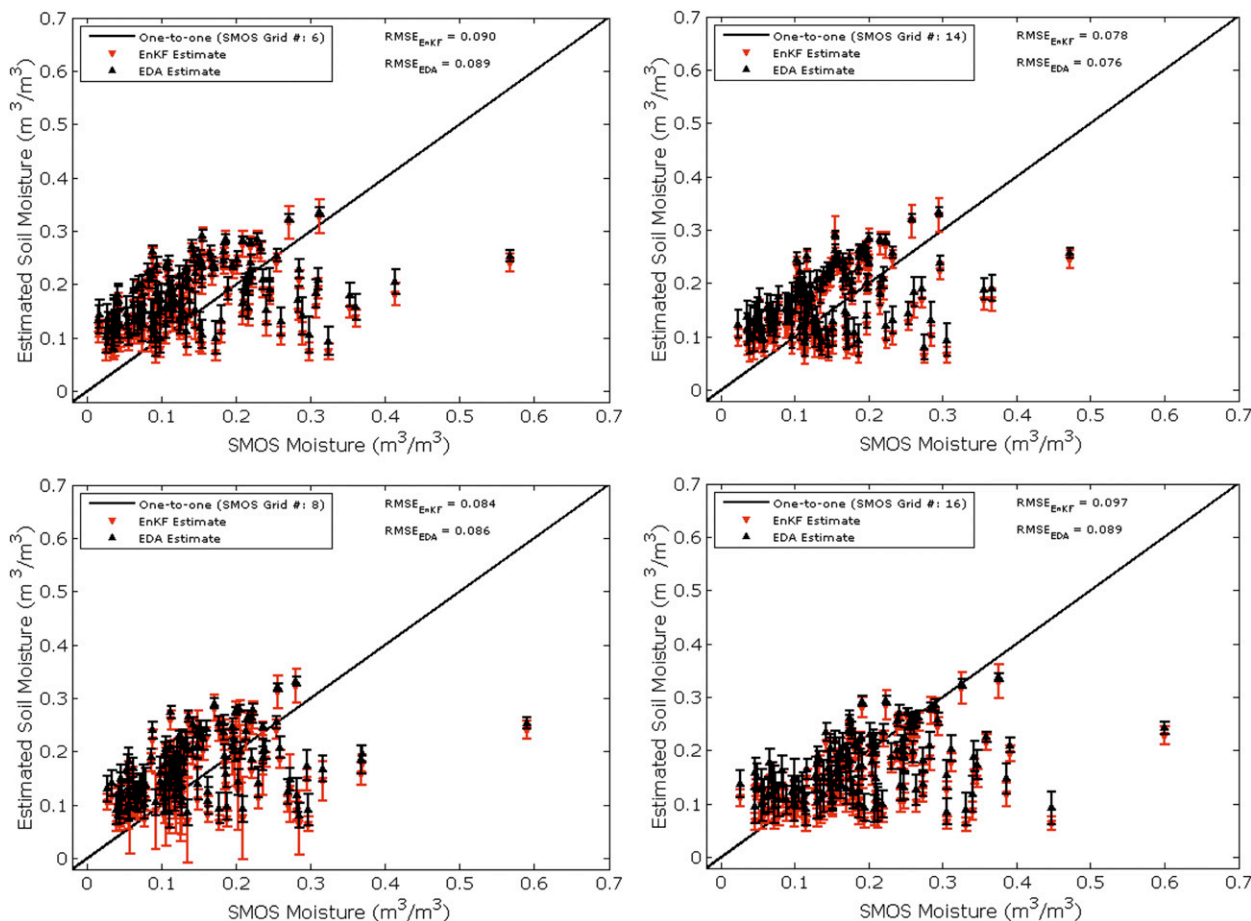


FIG. 10. As in Fig. 8, but for the forward estimations. The forward estimations from EnKF and EDA were determined by using the updated members at the end of July 2010 into the JULES model to estimate soil moisture from August to December 2010.

across the assimilation time steps. The significance of these findings is that the convergence of parameter values across different observation–assimilation time steps is valuable in the retrieval of variables that are not explicitly observed. This illustrates the potential of the EDA approach for examining the convergence of model parameters and their associated clusters through time, in order to determine their relationships, sensitivities, and their responses to changes in observation and forcing data.

c. Soil moisture estimations from the EnKF and the EDA

The surface soil moisture estimations from the EnKF and the EDA were compared in two stages: (i) an evaluation using the SMOS soil moisture for the assimilation time period and (ii) a subsequent validation using SMOS soil moisture for an independent estimation time period outside the assimilation. The open loop estimates and the updated ensemble estimates based on the ensemble mean

from EnKF and EDA for the assimilation period are compared to the SMOS soil moisture in Fig. 8. The SMOS grids shown in these comparisons were randomly chosen without bias toward any of the methods. It is noteworthy that, in this case, the open loop is the soil moisture estimated from ensemble members that were not updated. Both the EnKF and the EDA show an improvement in estimation accuracy compared to the open loop estimates across all SMOS grids.

The EDA has a higher estimation accuracy than the EnKF based on their comparison to the SMOS soil moisture using the estimated RMSE. The high performance of the EDA is consistent across SMOS grids, demonstrated by its low residual between the estimated soil moisture and the SMOS soil moisture in Fig. 9. The soil moisture residuals shown in this case were obtained as the overall absolute difference between the SMOS soil moisture and the ensemble estimates of soil moisture from EnKF and EDA methods across assimilation time steps for all grids.

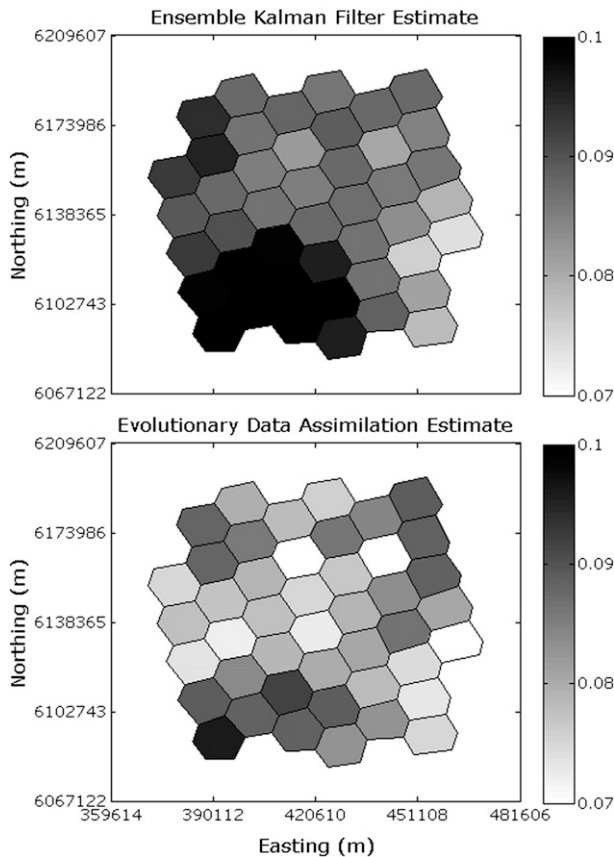


FIG. 11. As in Fig. 9, but for the overall RMS difference for the evaluation time period.

An evaluation procedure was undertaken to examine the differences in soil moisture estimates from using the updated model parameters from both methods for future time periods beyond the assimilation time period. The updated ensemble of model states and parameters from the EnKF and EDA, which were obtained on 31 July 2010 (i.e., end of assimilation period), were applied into the JULES model to estimate soil moisture forward in time from August to December 2010. It is noteworthy that 20 members were used in the EDA, as these represent the updated members at each assimilation time step; this is unlike the EnKF, where all 200 members were updated. The forward estimates from the EnKF and EDA are compared to the observed SMOS soil moisture in Fig. 10. The comparison between the forward estimations shows that the EDA (with RMSE of $0.089 \text{ m}^3 \text{ m}^{-3}$) has a slightly higher accuracy than the EnKF (with RMSE of $0.097 \text{ m}^3 \text{ m}^{-3}$ for SMOS grid 16). The RMSE for all the SMOS grids are shown in Fig. 11. The high accuracy in the EDA members illustrates the contribution of model parameter convergence for improved forward estimations. These results show that the

convergence of model parameters can contribute positively at the evaluation stage and at subsequent validation of forward estimates. However, contributions of convergence of model parameters at the time of update do not have a significant impact on the EnKF method.

Additional validation was undertaken to compare the ensemble estimates from EnKF and EDA to the in situ OzNet data (Smith et al. 2012). It is noteworthy that the in situ OzNet data are point observations, with overall daily variability during the modeling period between 0.036 and $0.116 \text{ m}^3 \text{ m}^{-3}$ across stations in the Yanco area. Accordingly, the validation results are described in recognition of the spatial-temporal variation of soil moisture, and the OzNet observation is, spatially, a fractional subset of the updated ensemble estimations at the 12-km SMOS DGG. The soil moisture comparisons are shown in Fig. 12 for stations Y3, Y5, Y8, and Y12.

The validation results show that the updated ensemble estimations from EnKF and EDA had higher accuracy than the SMOS soil moisture, and that both assimilation procedures had improved soil moisture estimation. The comparison between the EnKF and the EDA showed that there is an improved accuracy in the EDA output across the OzNet monitoring stations. Overall, the soil moisture estimation accuracy from the EnKF and the EDA was equivalent based on the SMOS evaluation, with the EDA showing an improved accuracy in the context of the OzNet validation data.

4. Conclusions

This study has examined the contributions of model parameter convergence for the EnKF and the EDA methods in soil moisture estimation using the JULES model. The SMOS level 2 soil moisture has been assimilated into the JULES model using the EnKF and the EDA methods, with their updated members examined for convergence of model parameter values along with accuracy of soil moisture estimation.

The results showed that convergence of model parameters can be obtained through the EDA approach. The level of convergence has been quantified for each model parameter using clustered intervals within which the recurrence of parameter values is highest. Aside from information on convergence, the EDA provides information on the level of sensitivity of model parameters across assimilation time steps. The ensemble parameter values from the EDA provide the potential for further investigation into the dynamics of model structure and identification of model weaknesses. The EDA approach is also shown to be robust for model parameter estimation across different observation scenarios.

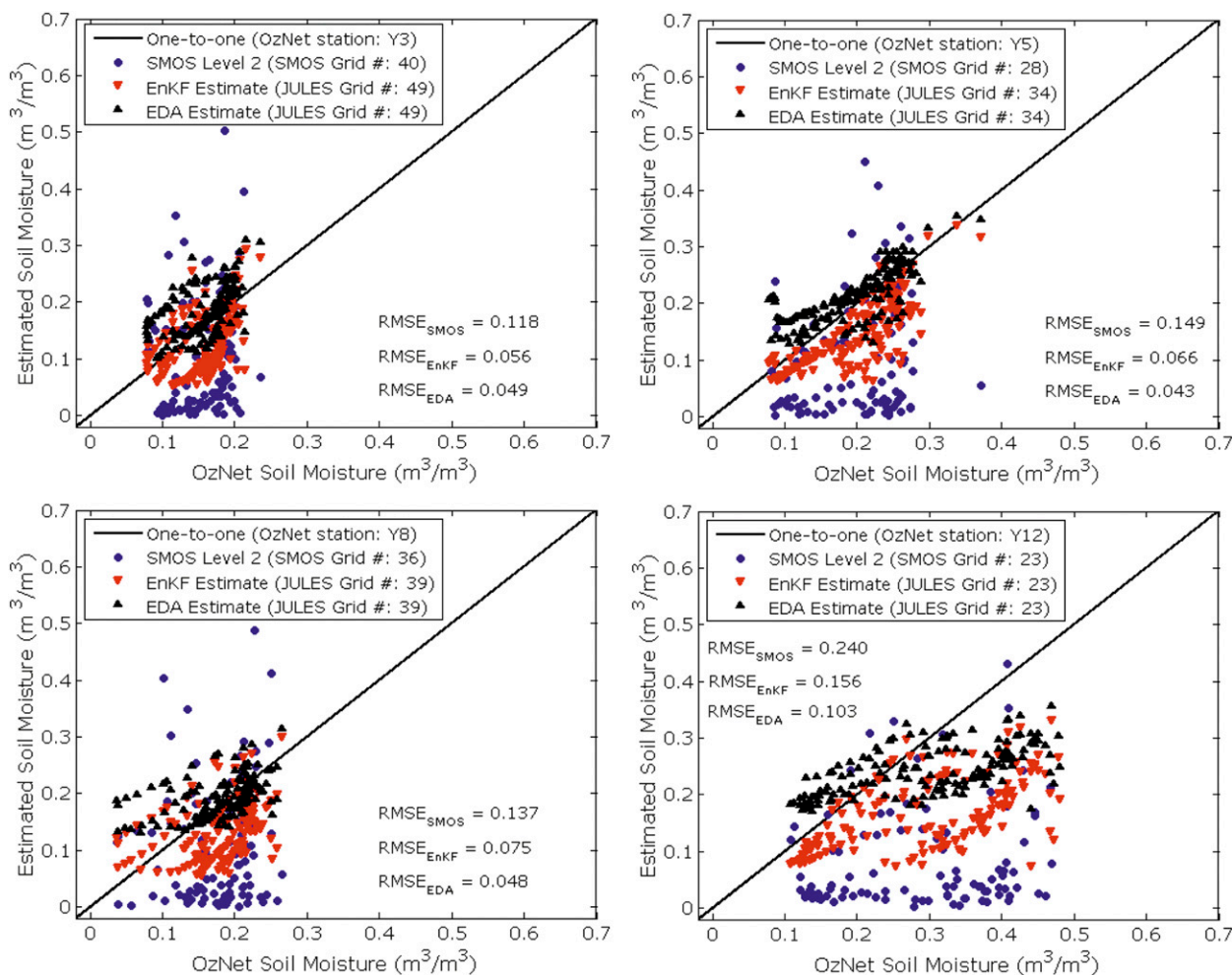


FIG. 12. Evaluation of SMOS soil moisture and the updated ensemble estimation from the EnKF and the EDA against in situ OzNet soil moisture observations at monitoring stations (top left) Y3 and (top right) Y5; and (bottom left) Y8 and (bottom right) Y12.

The soil moisture estimation accuracies from both EnKF and EDA have been shown to be higher than the open loop estimates in the evaluation procedure. At the evaluation stage, the EDA was shown to have a higher estimation accuracy compared to the EnKF across assimilation time steps and SMOS grids. However, validation of the forward soil moisture estimates from the two methods showed comparable results, with the EDA having a slightly higher accuracy. These findings showed that the convergence of model parameters does not significantly influence the estimation accuracy in the EnKF method. However, it has been found that the EDA approach simultaneously provided converged model parameters, along with slightly superior accuracy of soil moisture estimates.

It is noted that not all model parameters/states have converged and that further experimentation is needed to examine the interconnectedness between parameters/states. Though convergence may not be necessary in all

cases, an improved understanding of these model parameters is needed. The temporal changes observed in model states and parameter values were simply linked to the complex nature of the landscape (soil and land cover) properties. It is important to separate the model state–parameter changes associated with the spatial variability of the landscape from the temporal changes associated with the physical makeup of the landscape. The change associated with spatial variation in soil and land cover is what has been examined in this study and constitutes the bedrock to examining the dynamic changes linked with landscape makeup. These investigations are needed in future studies to better diagnose model weaknesses, with the potential to improve our understanding of the landscape process representation in models.

Acknowledgments. This research was supported by funding from the Australian Research Council (DP0879212).

The authors wish to thank both anonymous reviewers for their comments.

REFERENCES

- Andreadis, K. M., D. Liang, L. Tsang, D. P. Lettenmaier, and E. G. Josberger, 2008: Characterization of errors in a coupled snow hydrology–microwave emission model. *J. Hydrometeorol.*, **9**, 149–164, doi:10.1175/2007JHM885.1.
- Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), Model description—Part 1: Energy and water fluxes. *Geosci. Model Dev. Discuss.*, **4**, 595–640, doi:10.5194/gmdd-4-595-2011.
- Bureau of Meteorology, 2010: Operational implementation of the access numerical weather prediction systems. NMOC Operations Bull. 83, 35 pp.
- Burgers, T., P. Jan Van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.
- Clark, M., D. Rupp, R. Woods, X. Zheng, R. Ibbitt, A. Slater, J. Schmidt, and M. Uddstrom, 2008: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.*, **31**, 1309–1324, doi:10.1016/j.advwatres.2008.06.005.
- Confesor, R. B., and G. W. Whittaker, 2007: Automatic calibration of hydrologic models with multi-objective evolutionary algorithm and Pareto optimization. *J. Amer. Water Resour. Assoc.*, **43**, 981–989, doi:10.1111/j.1752-1688.2007.00080.x.
- Dumedah, G., 2012: Formulation of the evolutionary-based data assimilation, and its practical implementation. *Water Resour. Manage.*, **26**, 3853–3870, doi:10.1007/s11269-012-0107-0.
- , and P. Coulilaly, 2012: Evolutionary-based data assimilation: New prospects for hydrologic forecasting. *Proc. 10th Int. Conf. on Hydroinformatics—HIC 2012*, Hamburg, Germany, Hamburg University of Technology, HYA00318-00591.
- , and —, 2013a: Evolutionary assimilation of streamflow in distributed hydrologic modeling using in-situ soil moisture data. *Adv. Water Resour.*, **53**, 231–241, doi:10.1016/j.advwatres.2012.07.012.
- , and —, 2013b: Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation. *Adv. Water Resour.*, **60**, 47–63, doi:10.1016/j.advwatres.2013.07.007.
- , and —, 2014a: Examining the differences in streamflow estimation for gauged and ungauged watersheds using the evolutionary data assimilation. *J. Hydroinf.*, in press.
- , and —, 2014b: Integration of evolutionary algorithm into ensemble Kalman filter and particle filter for hydrologic data assimilation. *J. Hydroinf.*, in press.
- , A. A. Berg, and M. Wineberg, 2011: An integrated framework for a joint assimilation of brightness temperature and soil moisture using the Nondominated Sorting Genetic Algorithm-II. *J. Hydrometeorol.*, **12**, 1596–1609, doi:10.1175/JHM-D-10-05029.1.
- , —, and —, 2012: Evaluating autoselection methods used for choosing solutions from Pareto-optimal set: Does nondominance persist from calibration to validation phase? *J. Hydrol. Eng.*, **17**, 150–159, doi:10.1061/(ASCE)HE.1943-5584.0000389.
- , J. P. Walker, and C. Rüdiger, 2014: Can SMOS data be used directly on the 15-km discrete global grid? *IEEE Trans. Geosci. Remote Sens.*, doi:10.1109/TGRS.2013.2262501, in press.
- Efstratiadis, A., and D. Koutsoyiannis, 2010: One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrol. Sci. J.*, **55**, 58–78, doi:10.1080/02666660903526292.
- Evensen, G., 1994: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143–10 162, doi:10.1029/94JC00572.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.
- Gupta, H. V., T. Wagener, and Y. Liu, 2008: Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrol. Processes*, **22**, 3802–3813, doi:10.1002/hyp.6989.
- He, M., T. S. Hogue, S. A. Margulis, and K. J. Franz, 2012: An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions. *Hydrol. Earth Syst. Sci.*, **16**, 815–831, doi:10.5194/hess-16-815-2012.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Jones, D. A., W. Wang, and R. Fawcett, 2007: Climate data for the Australian Water Availability Project. Final Milestone Rep., Bureau of Meteorology, Melbourne, Australia, 37 pp. [Available online at <http://143.188.17.20/data/warehouse/brsShop/data/awapfinalreport200710.pdf>.]
- , —, and —, 2009: High-quality spatial climate data-sets for Australia. *Aust. Meteor. Oceanogr. J.*, **58**, 233–248.
- Lymburner, L., and Coauthors, 2011: The national dynamic land cover dataset. Tech. Rep., Geoscience Australia Record 2011/31, 95 pp. [Available online at <http://www.ga.gov.au/earth-observation/landcover.html>.]
- McKenzie, N. J., and J. Hook, 1992: Interpretations of the atlas of Australian soils. Consulting report to the Environmental Resources Information Network (ERIN), Tech. Rep. 94, CSIRO Division of Soils, Canberra, Australia, 20 pp.
- , D. W. Jacquier, L. J. Ashton, and H. P. Cresswell, 2000: Estimation of soil properties using the Atlas of Australian Soils. Tech. Rep. 11/00, CSIRO Land and Water, Canberra, Australia, 24 pp. [Available online from <http://www.clw.csiro.au/publications/technical2000/tr11-00.pdf>.]
- Mironov, V. L., M. C. Dobson, V. Kaupp, S. A. Komarov, and V. N. Kleshchenko, 2004: Generalized refractive mixing dielectric model for moist soils. *IEEE Trans. Geosci. Remote Sens.*, **42**, 773–785, doi:10.1109/TGRS.2003.823288.
- Moradkhani, H., and K. Hsu, 2005: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.*, **41**, W05012, doi:10.1029/2004WR003604.
- , S. Sorooshian, H. V. Gupta, and R. Paul Houser, 2005: Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.*, **28**, 135–147, doi:10.1016/j.advwatres.2004.09.002.
- Pipunic, R., K. McColl, D. Ryu, and J. Walker, 2011: Can assimilating remotely-sensed surface soil moisture data improve root-zone soil moisture predictions in the CABLE land surface model? *MODSIM2011: 19th International Congress on*

- Modelling and Simulation*, F. Chan, D. Marinova, and R. Anderssen, Eds., Modelling and Simulation Society of Australia and New Zealand, 1994–2001.
- Smith, A. B., and Coauthors, 2012: The Murrumbidgee soil moisture monitoring network data set. *Water Resour. Res.*, **48**, W07701, doi:10.1029/2012WR011976.
- Su, H., Z.-L. Yang, G.-Y. Niu, and C. R. Wilson, 2011: Parameter estimation in ensemble based snow data assimilation: A synthetic study. *Adv. Water Resour.*, **34**, 407–416, doi:10.1016/j.advwatres.2010.12.002.
- Tang, Y., P. Reed, and T. Wagener, 2006: How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrol. Earth Syst. Sci.*, **10**, 289–307, doi:10.5194/hess-10-289-2006.
- Thorndike, R. L., 1953: Who belongs in the family? *Psychometrika*, **18**, 267–276, doi:10.1007/BF02289263.
- Vrugt, J. A., and M. Sadegh, 2013: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. *Water Resour. Res.*, **49**, 4335–4345, doi:10.1002/wrcr.20354.
- , C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten, 2005a: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.*, **41**, W01017, doi:10.1029/2004WR003059.
- , B. A. Robinson, and V. V. Vesselinov, 2005b: Improved inverse modeling for flow and transport in subsurface media: Combined parameter and state estimation. *Geophys. Res. Lett.*, **32**, L18408, doi:10.1029/2005GL023940.
- Weerts, A. H., and G. Y. El Serafy, 2006: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resour. Res.*, **42**, W09403, doi:10.1029/2005WR004093.
- , —, S. Hummel, J. Dhondia, and H. Gerritsen, 2010: Application of generic data assimilation tools (DATools) for flood forecasting purposes. *Comput. Geosci.*, **36**, 453–463, doi:10.1016/j.cageo.2009.07.009.
- Xie, X., and D. Zhang, 2010: Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Adv. Water Resour.*, **33**, 678–690, doi:10.1016/j.advwatres.2010.03.012.