Particle Filter-based assimilation algorithms for improved estimation of root-zone soil moisture under dynamic vegetation conditions

Karthik Nagarajan [a,*], Jasmeet Judge [a], Wendy D. Graham [b], Alejandro Monsivais-Huertero [c]

[a] Center for Remote Sensing, Agricultural and Biological Engineering Department, The University of Florida, P.O. Box 110570, Gainesville, FL 32611, USA
[b] Water Institute, The University of Florida, Gainesville, FL 32611, USA
[c] ESIME Unidad Ticoman, Insitituto Polictenico Nacional, Mexico City, Mexico

ABSTRACT

In this study, we implement Particle Filter (PF)-based assimilation algorithms to improve root-zone soil moisture (RZSM) estimates from a coupled SVAT-vegetation model during a growing season of sweet corn in North Central Florida. The results from four different PF algorithms were compared with those from the Ensemble Kalman Filter (EnKF) when near-surface soil moisture was assimilated every 3 days using both synthetic and field observations. In the synthetic case, the PF algorithm with the best performance used residual resampling of the states and obtained resampled parameters from a uniform distribution and provided reductions of 76% in root mean square error (RMSE) over the openloop estimates. The EnKF provided the RZSM and parameter estimates that were closer to the truth than the PF with an 84% reduction in RMSE. When field observations were assimilated, the PF algorithm that maintained maximum parameter diversity offered the largest reduction of 16% in root mean square difference (RMSD) over the openloop estimates. Minimal differences were observed in the overall performance of the EnKF and PF using field observations since errors in model physics affected both the filters in a similar manner, with maximum reductions in RMSD compared to the openloop during the mid and reproductive stages.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Root-zone soil moisture (RZSM) is an important hydrologic state variable that governs moisture and energy fluxes at the land–atmospheric interface. Accurate knowledge of RZSM is essential for near-term climate predictions, hydrologic and agricultural research [24], and for effective water resources management. Typically, Soil Vegetation Atmosphere Transfer (SVAT)-vegetation growth models simulate energy and moisture transport in soil and vegetation, and estimate these fluxes at the land surface and in the root-zone over a growing season [7,40]. However, these coupled models exhibit large uncertainties in RZSM estimates due to errors in conceptualization, computation, initialization, forcings, and model parameters. Such uncertainties can be significantly reduced by assimilating in situ and/or remotely sensed observations of soil moisture [41,39,46,23]. Ensemble-based techniques such as the Ensemble Kalman Filter (EnKF) [17] and the Particle Filter (PF) [20] have been used for data assimilation in hydrology.

Both the EnKF and the PF techniques involve representing states and parameter distributions as ensembles or set of particles and are particularly suitable for estimating hydrologic states whose evolution over time are described by non-linear models. However, the EnKF uses innovations to update the prior states and parameters, while the PF uses the innovations to assign posterior weights to resample prior states and parameters. The EnKF also uses only second-order statistics for updating the states and parameters, while the PF utilizes the entire probability density function (PDF) of the states given the observations in computing the posterior weights, employing a Bayesian approach. Thus, the PF algorithms are expected to provide estimates closer to the observations than the EnKF when non-Gaussian distributions are involved or non-linear relationships exist between the estimated state and observed data. The PFs also require larger number of particles than the EnKF-based algorithms because of the need to estimate the entire PDF and are computationally more intensive. However, recent advances in high-performance computing allow further investigations into the implementation of different PF-based algorithms.

Significant research has been conducted on EnKF-based assimilation for RZSM estimation [42,46,12,37]. However, only a few studies have been performed using PF [38,45]. Also, only few studies have been conducted that compare the two techniques [54,52,23] for hydrology applications. Weerts and de Vries [52] compared the performance of the two techniques for rainfall-runoff estimation using synthetic and field observations and found that the EnKF provided discharge estimates closer to the truth than

* Corresponding author.
  E-mail address: nagkart@ufl.edu (K. Nagarajan).

the PF when field observations were assimilated. For studies in RZSM, Zhou et al. [54] found the EnKF to perform better than the PF, while Han and Li [23] found that both the EnKF and PF offered similar performances. Both the studies assimilated empirically-derived near-surface soil moisture from remotely sensed microwave observations, where the physical model is assumed to capture the biophysics perfectly, typical for synthetic studies. However, for field studies, additional errors may be introduced due to imperfect model physics and if the error distributions do not represent the actual errors (typically not known) in the model physics, the weights computed in the PF generate estimates that are further from the truth. Since the PF is more sensitive to model errors than the EnKF [52], studies comparing the performance of the EnKF and the PF using both synthetic and field observations are necessary to understand the impact of uncertainties in model biophysics, particularly under dynamic vegetation conditions.

The PF-based assimilations typically include resampling algorithms to regenerate values of states and parameters from the posterior distributions. Resampling is desired because PFs implemented without it, such as the Sequential Importance Sampling (SIS), can potentially lead to severe depletion of samples resulting in sample impoverishment [16]. Studies that used PF for RZSM estimation have incorporated the Sampling Importance Resampling (SIR) [38,23] algorithm. However, Liu and Chen [33] proposed residual resampling (RR) as an improvement to the SIR algorithm as it is computationally more efficient and provides lower variance of the posterior estimates. Weerts and El Serafy [52] compared the SIR and RR algorithms in rainfall-runoff modeling and found that the RR algorithm offered lower RMSEs. To date, the RR algorithm has not been applied in RZSM studies.

As mentioned earlier, for synthetic studies, the model error is assumed to be zero. However, when field observations are assimilated, additional bias may be introduced, particularly under dynamic land-surface conditions. Simultaneous estimation of parameters with states [9] allows for addressing bias in the model [14]. Applying the SIR or the RR algorithm without parameter perturbation may lead to repeated resampling of a few highly likely parameters resulting in sample impoverishment and inaccurate RZSM predictions after several updates. Hence, the resampling algorithms in PFs should be augmented by techniques that allow parameter values to maintain parameter diversity to obtain accurate estimates of RZSM in such cases. Moradkhani et al. [39] used a random walk model to perturb parameter estimates and maintain diversity. Because accurate estimation of model parameters play a significant role in accurate estimation of RZSM, an analyses of the impact of different parameter resampling algorithms on RZSM estimates is needed.

According to the hypothesis of equifinality [2], estimating states and parameters from observations related to RZSM imposes an ill-posed problem, wherein multiple parameter combinations may offer similar RZSM estimates. Both the EnKF and the PF (with resampling) narrow the PDF of parameters during assimilation. In this study, a comparison using parameter values estimated by different resampling methods within PF, including an algorithm with no resampling, is performed to investigate equifinality in RZSM estimation. The PF algorithm implemented without resampling would resemble the Generalized Likelihood Uncertainty Estimation (GLUE) method which has been used to study equifinality [2].

The goal of this study is to understand the impact of different parameter resampling algorithms in PF on the RZSM estimates while simultaneously updating states and parameters under dynamic vegetation conditions. A coupled SVAT-vegetation model is used to estimate RZSM during a growing season of sweet corn in North Central Florida. The estimates from the PF using four different algorithms are compared with those from the EnKF when synthetic and field-based observations of near-surface (0–5 cm) soil moisture are assimilated every 3 days. The four algorithms include, a PF without resampling or memory (PF-NRNM), a PF implemented using the RR algorithm (PF-RR), and PFs implemented using the RR algorithm with Gaussian (RR2) and uniform (RR3) parameter perturbations. Convergence of the RZSM estimates are investigated using two PF algorithms, one that includes resampling and the other that does not, to determine the number of particles needed in the PF for our application. Parameter convergence and root mean square errors (RMSE) in RZSM estimates are analyzed over the whole season and also during different growth stages to understand seasonal differences in algorithm performance.

In the next section, we briefly describe the field observations, the coupled SVAT-Vegetation model, and the PF and the EnKF algorithms used in this study.

2. Experiment, model, and assimilation

2.1. MicroWEX-2

The Second Microwave, Water and Energy Balance Experiment (MicroWEX-2) was conducted in North Central Florida (29.41 N, 82.18 W) during a growing season of sweet corn from Day of Year (DoY) 78 (March 18) to DoY 154 (June 2) in 2004 [28]. The experimental site was 3.6 h (9 acre). The soils at the site were primarily sand (89%) and the crop was heavily irrigated. During the MicroWEX-2, observations were conducted for soil moisture and temperature at depths of 2, 4, 8, 32, 64, and 100 cm, every 15 min along with observations of up and downwelling solar and longwave radiation, air temperature, relative humidity, precipitation, and irrigation. Judge et al. [28] provides details of observations conducted during MicroWEX-2 including microwave, soil, and vegetation observations. Table 1 shows the different growth stages of corn and their associated vegetation characteristics.

2.2. LSP-DSSAT model

The SVAT model used in this study is the Land Surface process (LSP) model [30]. It simulates 1-d coupled energy and moisture transport in soil and vegetation using a diffusion type equation, and estimates energy and moisture fluxes at the land surface and in the root zone. The model is forced with micrometeorological parameters such as air temperature, relative humidity, downwelling solar and longwave radiation, irrigation/precipitation, and windspeed. The model has been rigorously tested [27] and extended to wheat-stubble [29] and brome-grass [31] in the Great Plains, prairie wetlands in Florida [53], to tundra in the Arctic [10], and to growing crop [7]. The LSP model includes 16 parameters as shown in Table 2. The vegetation energy balance is calculated using the model developed by Verseghy et al. [51] for the water drainage from canopy, the bulk transfer approach for the sensible heat flux from Trenberth [49], and the latent heat flux following [43]. The coupled energy and water balance in soil is calculated from Philip and de Vries [44,15]. A block-centered finite-difference scheme is employed to solve the coupled governing equations at an adaptive time step (seconds/minutes) in response to the forcings [7].

The LSP model was coupled to a vegetation growth model, viz. Decision Support System for Agrotechnology Transfer (DSSAT) model to provide the flux estimates during dynamic vegetation conditions [7]. The DSSAT simulates crop growth and development at a daily step using modules for soil, soil–plant–atmosphere, weather, management, including irrigation and fertilization [26]. The DSSAT model includes modules for over 25 types of crops, including corn, soybeans, wheat, cotton, and different grass types for pasture. The model has been extensively tested in different

Table 1
Vegetation characteristics and growth stages during MicroWEX-2 from Casanova and Judge [7].

Growth stage	Number of days	Canopy height	LAI	Vegetation cover	Characteristics
Early	27 (DoY 78–105)	<17 cm	<0.2	<0.22	Almost bare soil
Mid	30 (DoY 105–135)	17–162 cm	0.2–2.49	0.22–1.00	Maximum vegetative growth stage
Reproductive	19 (DoY 135–154)	162–200 cm	2.49–2.75	1.00	Silking and ear formation, biomass increases primarily due to ear development

Table 2
Parameters included in the LSP model [7]. The values for canopy parameters were from Goudriaan [21] and ranges for soil parameters were from Rossi and Nimmo [47].

	Parameter	Description	Values
CANOPY	z_{ob}	Bare soil roughness length (m)	0.004
	x	Leaf angle distribution parameter	0.819
	σ	Leaf reflectance	0.474
	ϵ_c	Canopy emissivity	0.973
	ϵ_s	Soil emissivity	0.953
	c_d	Canopy drag coefficient	0.328
	i_w	Canopy wind intensity factor	67.9
	l_w	Leaf width (m)	0.0531
	F_b	Base assimilation rate (kg CO_2/ m^2 s)	-0.82×10^{-8}
	ϵ_{photo}	Photosynthetic efficiency (kg CO_2/J)	0.897×10^{-6}
	$soil_a$	Slope parameter for r_s (m^2 s/ kg H_2O)	370
	$soil_b$	Intercept parameter for r_s (m^2 s/ kg H_2O)	-531
Soil (0– 1.7 m)	λ	Pore-size index	0.1–0.9
	ψ_0	Air entry pressure (m H_2O)	0.05–1.0
	K_{sat}	Saturated hydraulic conductivity (m/s)	10^{-5}–10^{-3}
	ϕ	Porosity (m^3/m^3)	0.2–0.55
Soil (1.7– 2.7 m)	λ	Pore-size index	0.05
	ψ_0	Air entry pressure (m H_2O)	0.019
	K_{sat}	Saturated hydraulic conductivity (m/s)	8.93×10^{-5}
	ϕ	Porosity (m^3/m^3)	0.41

hydro-climatic regions [4,3,25,36,13,26,5,34,1,48]. The model was also tested and calibrated for its applicability to North-Central Florida [8] before it was coupled to the LSP model. In the coupled LSP-DSSAT model, the LSP model provides DSSAT with estimates of soil moisture and temperature profiles and ET. The DSSAT model provides LSP with vegetation characteristics that influence heat, moisture, and radiation transfer at the land surface and in the vadose zone [7]. The coupled LSP-DSSAT model has been used to implement the EnKF to understand impacts of simultaneous state-parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation [37].

2.3. Assimilation algorithms

In the following sections, a brief overview of the PF and the EnKF techniques is provided, along with the four different parameter resampling algorithms used in this study.

2.3.1. Particle Filter

In the PF [20], a set of particles representing state vectors is propagated in parallel, such that each state vector represents one realization of a non-linear model. The state equation for each particle in the filter is [19,39]:

$$x_t^{i-} = f(x_{t-1}^{i+}, u_{t-1}^i, \theta_{t-1}^+, t-1) + \eta_{t-1}^i \tag{1}$$

where $f(\cdot)$ is the non-linear model, x_t^{i-} is the state of the ith particle prior to the update at time t, x_{t-1}^{i+} is the posterior state of the ith particle at time $t-1$, u_{t-1}^i represents the meteorological forcings and

other inputs, θ_{t-1}^+ represents the model parameters, and η_{t-1}^i is the model error. In our study, the model physics are assumed to be perfect ($\eta_{t-1}^i = 0$). Eq. (1) also provides the transition probability, $p(x_t^i|x_{t-1}^i)$, of x_t^{i-}.

The observation equation given in Eq. (2) relates the prior state (x^{i-}) to the observations (d^i) through the measurement operator h, perturbed by errors associated with the observations, ϵ, at time t.

$$d_t^i = h x_t^{i-} + \epsilon_t^i \tag{2}$$

where ϵ is the error associated to the observations.

The PF represents the posterior state estimates, x_t^i, as a set of weights, w_t^i, calculated for each particle i. These combined weights construct the posterior PDF as given in Eq. (3). During assimilation at time t, the weight of the ith particle is updated based on the likelihood of the observation, $p(d_t^i|x_t^{i-})$, as given in Eq. (4)

$$p(x_t|d_t) = \sum_{i=0}^{N} w_t^i \delta(x_t - x_t^{i-}) \tag{3}$$

$$w_t^i \propto w_{t-1}^i \frac{p(d_t^i|x_t^{i-})p(x_t^{i-}|x_{t-1}^i)}{q(x_t^{i-}|x_{t-1}^i, d_t^i)} \tag{4}$$

where $\delta(\cdot)$ is the Kronecker delta function and $q(\cdot)$ is the proposal function. Past studies in hydrology have used the state Eq. (1) as the proposal function in Eq. (4), i.e. $q(x_t^{i-}|x_{t-1}^i, d_t^i) = p(x_t^{i-}|x_{t-1}^i)$. Using the assumption, Eq. (4) can be simplified to obtain the normalized weights for N particles as:

$$w_t^i = w_{t-1}^i \frac{p(d_t^i|x_t^{i-})}{\sum_{j=0}^{N} w_{t-1}^j p(d_t^j|x_t^{j-})} \tag{5}$$

Even though, the observation error in the PF need not be Gaussian, typically it is assumed Gaussian [38,23]. Then, the likelihood, $p(d_t^i|x_t^{i-})$, can be represented as $p(d_t^i - h x_t^{i-}|R)$ and

$$p(\epsilon_t^i|R) = \frac{\exp\left(-\frac{1}{2}\epsilon_t^{iT} R^{-1} \epsilon_t^i\right)}{2\pi^{k/2}\sqrt{|R|}} \tag{6}$$

where R is matrix of dimension k representing measurement error variance. Fig. 1 compares and illustrates the key steps involved in the PF. The posterior state estimated from the N particles is given by,

$$x_t = \sum_{i=0}^{N} x_t^{i-} w_t^i \delta(x_t - x_t^{i-}) \tag{7}$$

2.3.2. Resampling algorithms in PF

During the update step in PF, the weights w_t^i are computed and used to resample the PDF for N particles. In this study, we compare four different parameter resampling algorithms within the PF: PF-NRNM, PF-RR, PF-RR2, and PF-RR3. The PF-RR, PF-RR2, and PF-RR3 algorithms employ the RR [33] algorithm to resample the prior state vector.

- *Particle filter without resampling or memory (PF-NRNM)* – In the PF-NRNM algorithm, the states and parameters of the initial particles are not resampled and remain unchanged throughout assimilation similar to the openloop case. The algorithm
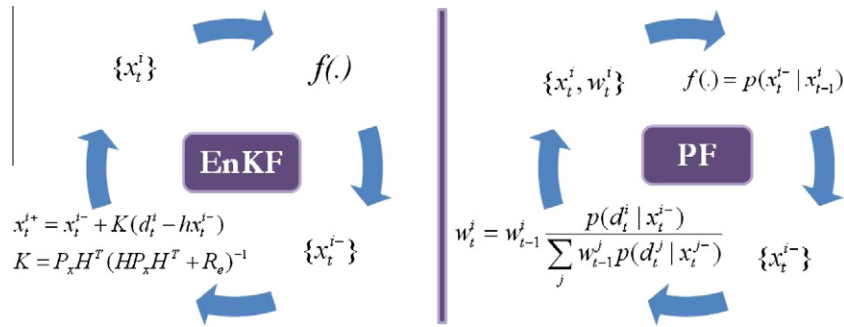
Fig. 1. Implementation of the (a) ensemble Kalman Filter and (b) Particle Filter algorithms.

maintains diversity because it does not retain memory and the weights (w_t^i) are not dependent on the weights computed during the previous update (w_{t-1}^i), as given in Eq. (8). The posterior parameter distribution is estimated from the original uniform parameter distribution during each update. This algorithm is similar to the GLUE methodology [2]

$$w_t^i = \frac{p(d_t^i|x_t^{i-})}{\sum_{j=0}^{N} p(d_t^j|x_t^{j-})} \qquad (8)$$

In contrast, in a non-resampling PF with memory, the presence of memory due to the dependence of weights between consecutive updates, as given in Eq. (5), leads to sample impoverishment.

- *Particle Filter with residual resampling (PF-RR)* – In the PF-RR algorithm, both the model states and parameters are resampled using the RR algorithm [33]. The parameters are not perturbed. For example, if the third and fifth particles out of 10 particles obtained updated weights of 0.2 and 0.8, respectively, the state

and parameter values corresponding to the third and fifth particles are replicated two and eight times, respectively, to obtain 10 updated particles (see Fig. 2).

- *Particle Filter with residual resampling and Gaussian parameter perturbation (PF-RR2)* – In the PF-RR2, the RR algorithm is used for resampling states and parameters as in PF-RR, but the resampled parameters are perturbed by a Gaussian error with zero mean and standard deviation, $s\sigma_t$, to introduce diversity. The σ_t is the standard deviation of the parameters at time t before resampling and s is a tuning parameter which determines the radius around each particle being explored. s was set to 0.1 in this study, following [39].

- *Particle Filter with residual resampling and uniform parameter replacement (PF-RR3)* – Similar to the PF-RR2, the RR algorithm is also used to resample states and parameters in the PF-RR3 algorithm. But, unlike the Gaussian error, every replicated parameter is resampled from a uniform distribution constructed from the minimum and maximum value of parameters
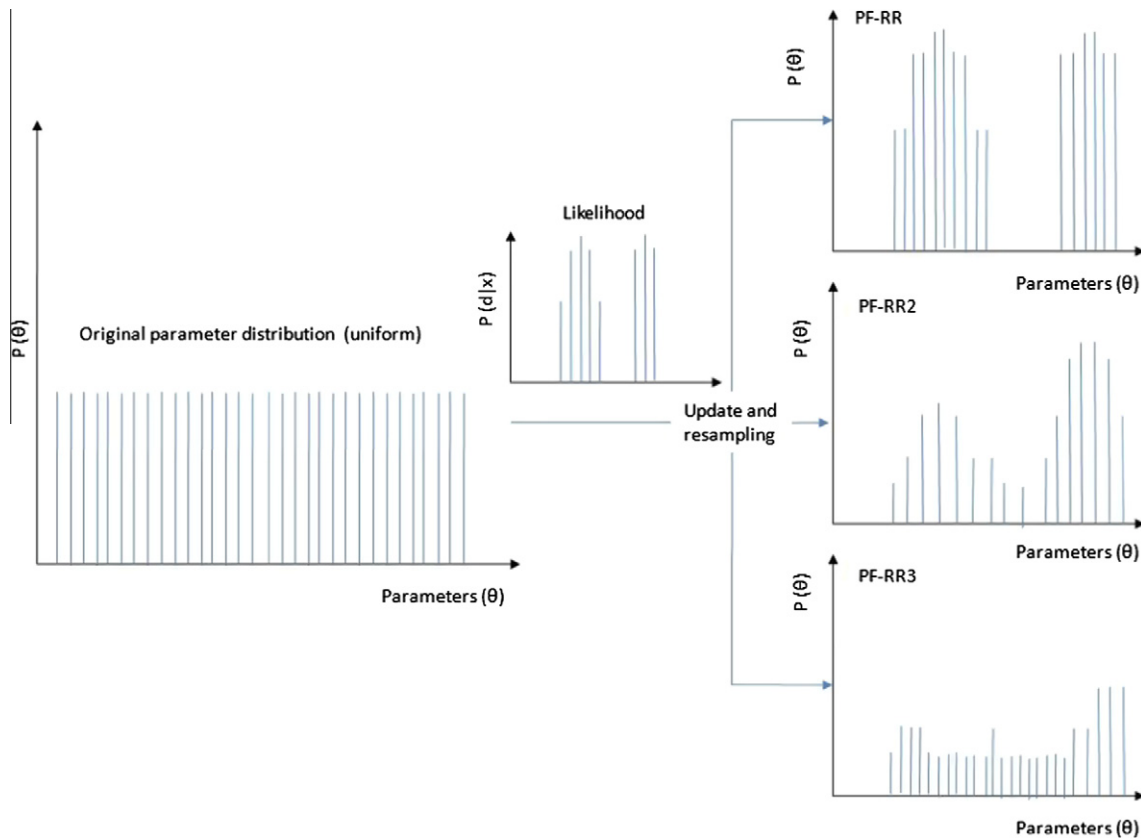

Fig. 2. Resampling algorithms for parameters within the PF used in this study. RR represents residual resampling, RR2 represents residual resampling with Gaussian parameter perturbation, and RR3 represents residual resampling with uniform parameter replacement.

contained in the resampled set (see Fig. 2). Thus this algorithm resamples from a wider distribution than the one in the PF-RR2 algorithm, offering more diversity in parameter values.

The PF-RR and PF-RR3 algorithms represent two extremes in standard deviations of errors in posterior estimates of parameters. The PF-RR represents a standard deviation of zero, while the PF-RR3 algorithm represents a case of maximum standard deviation.

2.3.3. Ensemble Kalman filter

Both the EnKF and the PF have the same propagation step but have different update steps. As described in Section 2.3.1, the PF uses the entire PDF of the states given the observations to update the posterior states, while the EnKF assumes that the states and parameters are Gaussian in calculating the Kalman gain which is then used to update the posterior states. The posterior state vectors x_t^{i+} for the ith ensemble at time t is computed as a linear combination of the prior estimate x_t^{i-} and the observation d_t^i weighed by the Kalman gain K_t

$$x_t^{i+} = x_t^{i-} + K_t(d_t^i - hx_t^{i-}) \qquad (9)$$

The Kalman gain can be calculated as [18]:

$$K = P_x H^T (H P_x H^T + R_e)^{-1} \qquad (10)$$

where P_x represents the prior state covariance matrix, R_e is the covariance matrix of the observation errors, and H is the observation operator relating the ensemble of perturbed observations to the ensemble of states. The posterior state estimated from the N ensemble members is given by,

$$x_t = \sum_{i=0}^{N} \frac{x_t^{i+}}{N} \qquad (11)$$

2.3.4. Comparison of PF and EnKF

As mentioned in Section 2.3.3, the primary difference between the PF and the EnKF is in the update step. The PF generates the posterior estimates by resampling the prior estimates based upon the weights computed from Eq. (5), while the EnKF computes the posterior estimates as a linear combination of the prior estimates and the innovations weighted by the Kalman gain [Eq. (9)]. The above difference in the update step may lead to parameter impoverishment in the PF when parameters are also updated because resampling limits the posterior estimates to a subset of the prior values. In the EnKF, the parameter values does not suffer from diversity issues since the posterior estimates from Eq. (9) are not limited to a subset of the prior values.

Another difference between the PF and the EnKF algorithms is their effect on the update of unobserved states and parameters in the augmented state vector. In the PF, the posterior weights, w_t, used for updating the unobserved states and parameters are obtained based only upon the likelihood of the observed states, $p(d_t^i|x_t^{i-})$, as shown in Eq. (5). Since the proposal function in Eq. (4) is assumed to be the state equation, $f(\cdot)$, in Eq. (1), the relation between the observed and unobserved states and parameters are not embedded within the computations of the posterior weights. However, in the EnKF, the state covariance matrix, P_x, includes correlations between unobserved states and parameters and the observed states and thus regulates the updates of the unobserved states and parameters. Therefore, while performing simultaneous update of states and parameters, augmenting parameters that are not significantly sensitive to the observed states may impact the performance of the PF more adversely than the EnKF.

The third significant difference between the PF and the EnKF is the effect of number of particles or ensemble members on the performance of the filters. Since the PF needs to resolve higher-order distribution properties that are ignored by the EnKF, more particles

may be required in the implementation of the PFs for reliable comparisons with the EnKF. In a recent study, 500 particles were chosen for assimilation to achieve reliable comparisons between the PF and the EnKF for rainfall runoff [50]. Zhou et al. [54] found that particle sizes greater than 800 resulted in nearly the same RMSEs between the PF and the EnKF for root zone soil moisture. In this study, we analyze the convergence of the RZSM estimates to determine the number of particles needed in the PF for our application.

3. Methodology

In this section, we describe the LSP-DSSAT simulations and implementation of the assimilation algorithms in this study.

3.1. LSP-DSSAT simulation

At the MicroWEX-2 site, the first 1.7 m of soil was primarily sandy, with 89.4% sand by volume and the second layer (1.7–2.7 m) constituted 40.5% sand by volume. In the LSP model, the soil was discretized into 35 computational blocks (nodes) in the two layers. The blocks increased in thickness exponentially, with 4 blocks in the top 5 cm of the soil. The coupled LSP-DSSAT simulations were conducted from planting on DoY 78, to harvest, on DoY 153, during a growing season of sweet corn in 2004. Micrometeorological and vegetation forcings for the simulations were obtained from the MicroWEX-2 and DSSAT, respectively. Initial conditions were not known during the MicroWEX-2 because sensor installation was completed 7 days after planting. The first values observed by the soil moisture and temperature sensors were used as the initial moisture and temperature values for the DoY 78.

3.2. Implementation of assimilation algorithms

In this study, the non-linear propagator $f(\cdot)$ in Eq. (1) represents the coupled LSP-DSSAT model; x is the state vector consisting of volumetric soil moisture (VSM) at different depths in the soil estimated by the LSP-DSSAT model; u_t is the vector of meteorological forcings at time t, and θ are the time-invariant model parameters in the LSP model, shown in Table 2.

The VSM observations at 0–5 cm for assimilation were obtained from MicroWEX-2 observations at the depths of 2 and 4 cm and assumed to be unbiased, following Lorenc and Hammon [35]. The augmented state vector technique [9] is used in this study to simultaneously estimate states and parameters. The augmented state vector x_i (see Eq. (12)) includes soil moisture at the 35 computational nodes and the imperfectly known parameters, namely, porosity (ϕ), air entry pressure (ψ_0), pore-size index (λ), and saturated hydraulic conductivity (K_{sat}) describing the first soil layer, assumed homogeneous. These four parameters were found to be the most influential for RZSM estimation, based upon time-dependent correlations of the RZSM estimates to the 16 parameters in the LSP model [37]. In this study, the second layer (1.7–2.7 m) of soil was assumed to have known constitutive properties, as shown in Table 2 obtained from Casanova and Judge [7]. The ith particle is therefore expressed as,

$$x_i = \begin{bmatrix} VSM_{i1} \\ VSM_{i2} \\ \vdots \\ VSM_{ik} \\ \phi_i \\ \psi_{0_i} \\ K_{sat_i} \\ \lambda_i \end{bmatrix} \qquad (12)$$

where k represents the number of nodes of the LSP model.

The RZSM was calculated from the LSP-DSSAT model predictions using the following equation

$$\text{RZSM} = \sum_{i=1}^{m} \text{VSM}_i \Delta z_i \tag{13}$$

where m indicates the total number of nodes (blocks) within the root zone (0–1.2 m), Δz_i the thickness of the ith node, and VSM_i the volumetric soil moisture at ith node.

Soil moisture observations (both synthetic and from MicroWEX-2) at depths of 2 and 4 cm were assimilated into the LSP-DSSAT model every 3 days using the PF and EnKF. The assimilation time was chosen to be 6 a.m.EST for this study, corresponding to the current and near-future availability of remotely sensed soil moisture from microwave observations [32].

The RZSM estimates and their standard deviations obtained from the model were compared with the truth for both the PF and EnKF. The RMSE between the mean estimates and the truth along with the time-average ensemble standard deviation (ASD) were computed to quantify algorithm performance. The root mean square differences (RMSD) between the estimates and field observations along with the ASD was used to quantify the filter performances in the MicroWEX-2 case. While the RMSE/D is a metric of accuracy, the ASD is a metric of uncertainty and can be used as an error metric of the RMSE. An optimal estimate of RZSM would have low values of both the RMSE/D and ASD

$$\text{RMSE} = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} (\text{RZSM}_t^{mean} - \text{RZSM}_t^{truth})^2} \tag{14}$$

$$\text{RMSD} = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} (\text{RZSM}_t^{mean} - \text{RZSM}_t^{obs})^2} \tag{15}$$

$$\text{ASD} = \frac{1}{N} \sum_{t=1}^{N_t} \frac{1}{N} \sum_{i=1}^{N} (\text{RZSM}_t^i - \text{RZSM}_t^{mean})^2 \tag{16}$$

where N_t is the number of RZSM estimates over time and RZSM_t^{mean} is the mean RZSM at time t computed from the N particles. The number of particles in the filters, N, were determined based upon convergence of RZSM using two PF algorithms implemented with and without resampling. Fig. 3 shows the RMSEs in the RZSM estimates obtained by the PF-NRNM and PF-RR2 algorithms for particle sizes varying from 50 to 2000, when synthetic observations of VSM were assimilated. Significant reductions in the RMSE values were observed when the particle size was increased from 50 to 500, but further increases in the particles did not result in continued improvements in RMSE. In this study, 500 (N = 500) particles were used for assimilation in both the PF and the EnKF algorithms.
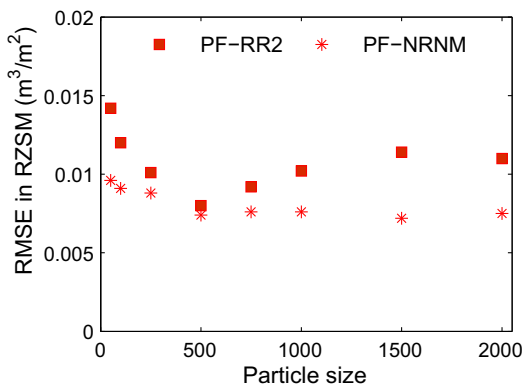

Fig. 3. RMSEs in the RZSM estimates for various particle sizes when synthetic observations of VSM at 0–5 cm were assimilated using the PF-NRNM and PF-RR2 algorithms.

3.3. Uncertainty in forcings, observations, and parameters

In this study, the errors in both synthetic and MicroWEX-2 observations were assumed Gaussian with zero mean. The standard deviation of the errors was 2% by volume [6]. The four uncertain soil parameters were randomly generated from a uniform distribution with literature-based upper and lower bounds (see Table 2). The use of a uniform distribution avoids the generation of negative parameters and assumes that all values are equally likely. Among all the inputs/forcings to the LSP-DSSAT model, precipitation/irrigation observations typically have the highest errors compared to other micrometeorological parameters. The errors in precipitation/irrigation observations can range between 3 and 12%, depending on the duration and intensity of rainfall [11,22]. In this study, a Gaussian observation error with standard deviation equal to 12% of the observed value of precipitation/irrigation was introduced during events. No errors were introduced in the absence of the events [37].

3.4. Synthetic and field observations

The synthetic truth was assumed to be one of the particles during an openloop simulation of the LSP-DSSAT using perturbed precipitation forcings and perturbed parameters. A zero-mean Gaussian error with 2% standard deviation was added to the synthetic truth to generate synthetic observations. The synthetic truth was not included in the 500 particles used during the assimilation. The field observations were obtained from the MicroWEX-2 experiment, described in Section 2.1. As done in the synthetic case, a zero-mean Gaussian error with 2% standard deviation [6] was also added to the field observations.

4. Results and discussion

4.1. Synthetic experiment

4.1.1. State estimation

Fig. 4(a) shows the absolute errors in RZSM estimates (ΔRZSM) obtained using the four PF algorithms for the synthetic case. The mean estimates using the PF algorithms are closer to the truth over the entire season (see Fig. 5(a)) compared to the mean estimates from an ensemble of 500 realizations without assimilation (openloop). Table 3 shows the RMSE and ASD obtained for the PF algorithms when synthetic observations are assimilated. The RMSEs in the RZSM estimates are reduced by 76% (0.006 m³/m²), 72% (0.007 m³/m²), and 72% (0.007 m³/m²) using the PF-RR3, PF-RR2, and PF-NRNM algorithms, respectively, compared to those obtained by the openloop. Since the PF-RR2 and PF-RR3 algorithms narrow the parameter space due to resampling, their ASD values are lower (0.008 m³/m² and 0.01 m³/m², respectively) compared to the PF-NRNM algorithm. As mentioned earlier in Section 2.3.2, the PF-NRNM algorithm overcomes the issue of sample impoverishment by restricting the dependence of weights between consecutive updates and results in the highest ASD value of 0.018 m³/m². Although the PF-RR algorithm provides the lowest ASD of 0.003 m³/m², the reduction in the RMSE of 64% compared to the openloop is the lowest among the PF algorithms indicating the algorithm is over-confident of its estimates. This low reduction in the PF-RR algorithm is due to the lack of parameter diversity and convergence early during the season (DoY 119) to parameter values that are farther from the truth (see Section 4.1.2). While the early convergence decreases the RMSEs during the early stage of crop growth, they result in RZSM estimates that drift away from the truth as vegetation grows (see Table 4). The PF-NRNM, PR-RR2, and PF-RR3 algorithms maintain parameter diversity and ex-
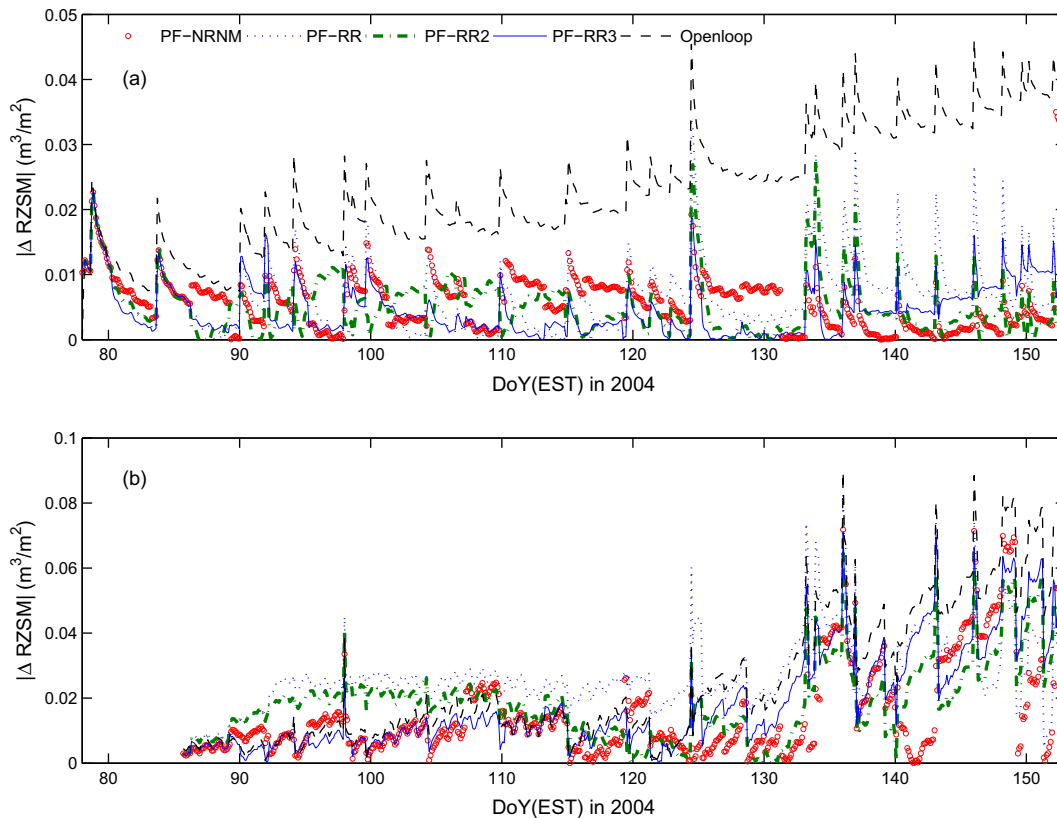
Fig. 4. Absolute errors/differences in mean estimates of RZSM estimated by the LSP-DSSAT model and the truth (synthetic)/observation (MicroWEX-2) when (a) synthetic and (b) MicroWEX-2 observations of VSM at 0–5 cm were assimilated using four different PF algorithms.

hibit larger RMSE reductions during the reproductive stage, with the PF-RR2 algorithm providing the largest reduction of 85.7% during the stage.

4.1.2. Parameter convergence

Since model physics is assumed perfectly known (i.e., $\eta = 0$) in the synthetic experiment, parameters that are sensitive to RZSM are expected to converge to the true values. Performance of the PF algorithms can hence be gauged based on the differences between their parameter estimates and the true values. In a previous study conducted using the LSP-DSSAT model, Monsivais et al. [37] inferred that while ϕ was highly correlated to RZSM with a season-average correlation coefficient of 0.8, λ, K_{sat}, and ψ_0 were only moderately correlated with coefficients varying between -0.25 and 0.1. All parameters, except for ϕ, also displayed temporal variations in correlation to RZSM with highest correlation during precipitation events and lower correlations during subsequent dry-down periods. Because in the implementation of the PF algorithms, the parameters are resampled based upon weights computed from the likelihood of the observed states without considering their sensitivity to the observed states, unless assimilation is performed at times of highest sensitivity, PFs may lead to parameter estimates that are far from the truth. The PF-NRNM and PF-RR3 algorithms maintain a wider parameter diversity and therefore converge to parameter means that are closer to the truth but with higher uncertainty compared to the PF-RR and PF-RR2 algorithms. Since the PF-RR algorithm does not involve parameter perturbation, it leads to convergence early during the season (DoY 119) with ASD values of zero. Table 6 gives the means and standard deviations of ϕ, ψ_0, λ, and K_{sat} at the end of the season when synthetic observations were assimilated.

Porosity is highly correlated to RZSM throughout the season, and all the PFs lead to ϕ values that are close to the truth (see Table 6). The maximum difference between the estimates and the true value was approximately 0.05 for PF-RR3. The PF-RR algorithm is the fastest converging and converges to a ϕ value of 0.2912 by DoY 119 (see Fig. 7). The PF-RR2 algorithm is able to estimate a mean ϕ value of 0.3217 with a SD of 0.0475 due to the Gaussian perturbation during resampling (see Fig. 7).

K_{sat} highly influences moisture transport immediately after precipitation events and showcases strongest sensitivity to RZSM immediately following such events. The estimated K_{sat} is lower than the true value for all PF algorithms (see Fig. 8). Since the first four assimilation updates (i.e., DoY 80, DoY 83, DoY 86, and DoY 89) were performed during times of low sensitivity to RZSM, the PF-RR and PF-RR2 algorithms converge to K_{sat} values farther from the truth with low predicted uncertainty and continue to resample lower values after DoY 90. This results in poor RZSM estimates during dry-down periods throughout the season, as seen in Fig. 5(a). The PF-NRNM and PF-RR3 algorithms maintain a wider parameter diversity throughout assimilation and the mean estimates of K_{sat} and RZSM estimates are closer to the truth, but with higher predicted uncertainty. Consequently, the ASDs in RZSM estimates increase significantly for all PFs immediately after precipitation events (e.g., DoY 110, DoY 115, and DoY 124) and continue to remain high during the initial dry-down phase as seen in Fig. 5(b).

Similar to the K_{sat}, λ and ψ_0 are also sensitive to precipitation events and are estimated with larger uncertainties around such events (see Figs. 9 and 10). While λ is moderately correlated to RZSM during the early and reproductive stages of the season, it is relatively uncorrelated during the mid-stage. None of the PF algorithms show strong convergence to the true value of λ. The
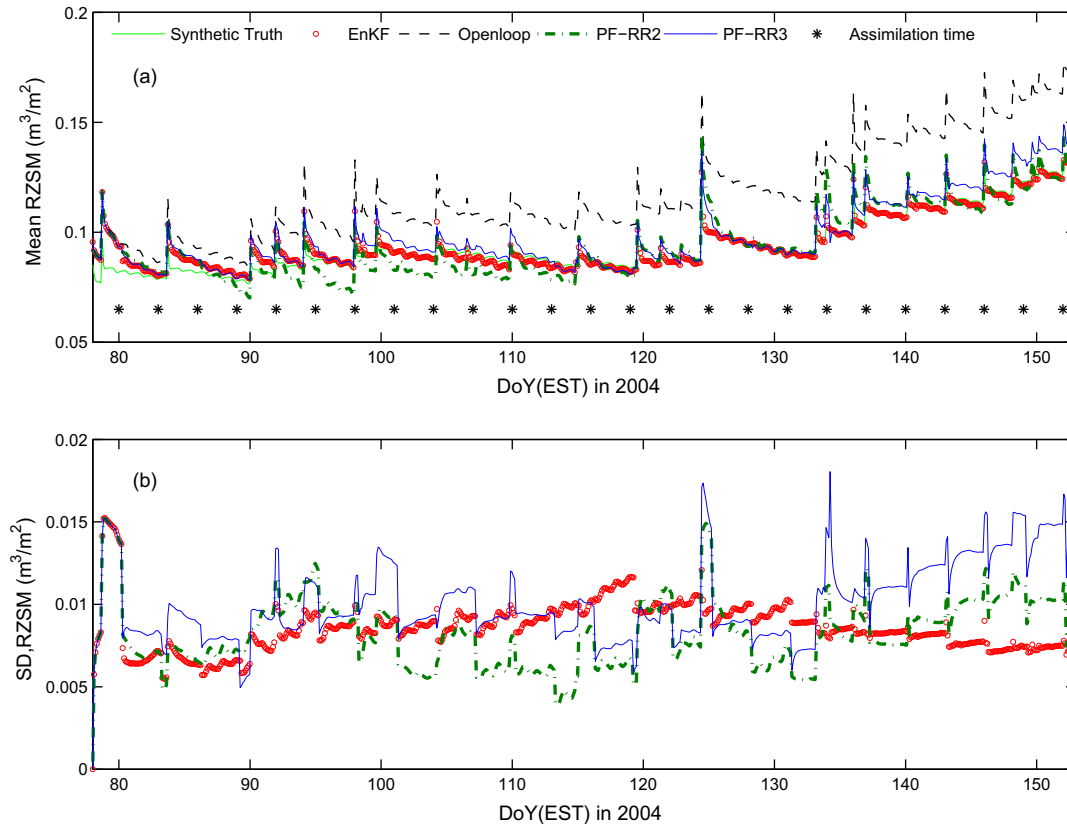
Fig. 5. (a) Means and (b) ensemble standard deviations (SD) of posterior RZSM estimates when synthetic observations of VSM at 0–5 cm were assimilated using EnKF, PF-RR2 and PF-RR3 algorithms.

PF-NRNM and PF-RR3 algorithms have mean estimates that are closer to the truth. In the PF-RR2 algorithm, the λ estimates converge towards the truth until DoY 105 and then move farther from the truth because of low correlation to RZSM during the mid stage.

4.1.3. Comparison of PF with EnKF

When synthetic observations of VSM are assimilated, the EnKF offers the best overall performance with a 84% reduction in RMSE (0.004 m^3/m^2) over the openloop estimates (see Fig. 5(a)), compared to 76% offered by the best PF-RR3 algorithm. In a synthetic experiment, the differences in the performance of the PF and EnKF algorithms can be gauged based on the deviation of parameter estimates from their true values.

Since porosity, ϕ, is highly sensitive to RZSM, both the PF and EnKF algorithms converge to values that are close to the truth (see Table 6). The standard deviation in the estimates of ϕ vary between the PF and EnKF algorithms with the PF-RR algorithm providing the least standard deviation by converging around DoY115 (Fig. 7). In the EnKF, parameters, K_{sat}, λ, and ψ_0, showcase equifinality (Figs. 8–10) by maintaining large predicted uncertainty and not converging to the true values with no adverse effects on the RZSM estimates. On the contrary, in PF algorithms that maintain the least diversity, namely the PF-RR and PF-RR2 algorithms, the parameters converge to values far away from the truth, providing reduced improvements to RZSM estimates. The PF-NRNM and PF-RR3 algorithms provide parameter estimates that are closer to those estimated by the EnKF and also provide the lowest RMSEs among the PF algorithms over the entire season.

Seasonal variations in RMSEs are also observed between the PF and EnKF algorithms. Significant improvements in RMSEs are not observed in the PF algorithms in comparison to the EnKF as the

season progresses (see Table 4). Among the PF algorithms, the PF-RR2 algorithm provided the maximum improvement with a 53.3% RMSE reduction during the early stage to 85.7% during the reproductive stage. In the EnKF, the reductions in RMSE increases from 60% during the early stage to 97.1% during the reproductive stage (see Table 4). The reduced improvements obtained in the PF algorithms is attributed to the suboptimal computation of the weights, which are used for resampling the parameters, based only upon the likelihood of the observed states, as discussed in Section 2.3.4.

4.2. MicroWEX-2 experiment

4.2.1. State estimation

When MicroWEX-2 observations were assimilated, additional bias in the form of errors are introduced due to imperfect knowledge of the model physics. As mentioned earlier, a common practice to deal with the bias in the model is using the augmented state vector with added uncertainties in the sensitive parameters. Fig. 4(b) shows the absolute differences in RZSM estimates (ΔRZSM) obtained using the four PF algorithms for the MicroW-EX-2 case. The PF-RR algorithm performs poorly due to lack of parameter diversity and the RZSM estimates deviate significantly from the observations. The PF-NRNM algorithm results in a 16% (0.021 m^3/m^2) RMSD reduction over openloop estimates, and the PF-RR2 and PF-RR3 algorithms result in slightly lower but similar improvements of 8% (0.023 m^3/m^2) and 4% (0.024 m^3/m^2) RMSD reductions, respectively (see Table 3). This suggests that parameter diversity should be maintained to obtain lower RMSD values when real observations are assimilated. Under dynamic land surface conditions, when the sensitivities of parameters to RZSM vary over time, maintaining the diversity in parameters may be essential
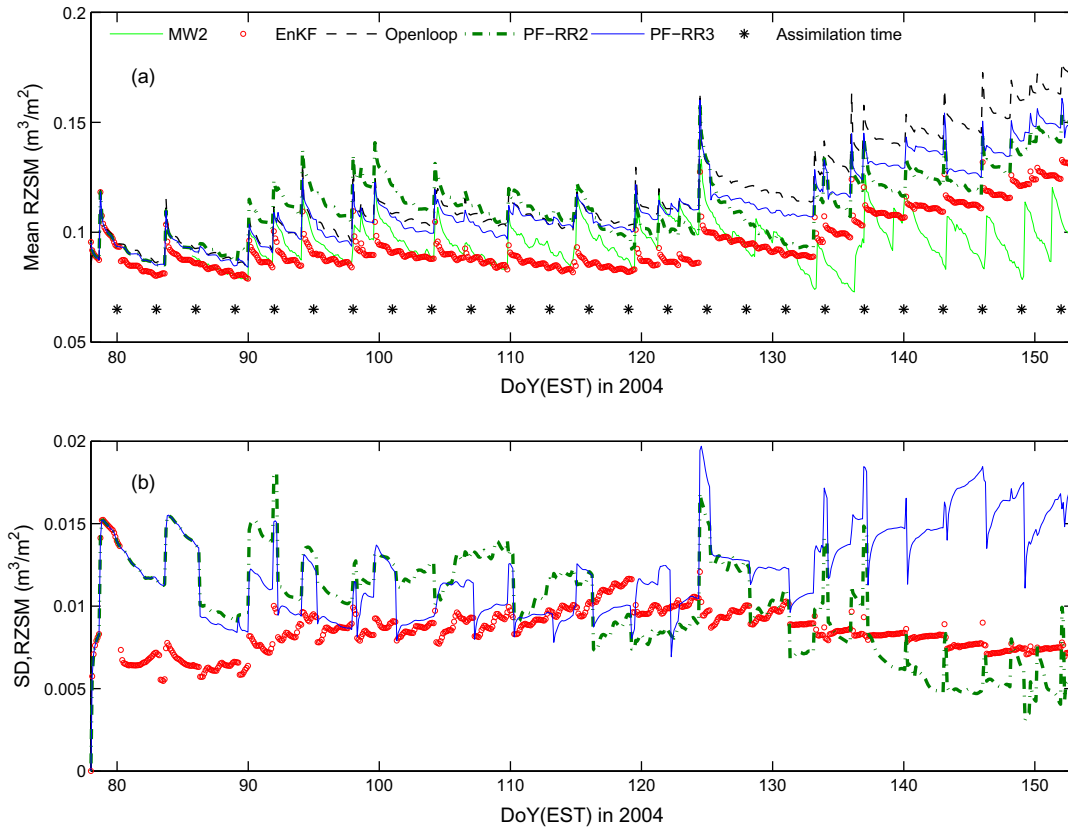
Fig. 6. (a) Means and (b) ensemble standard deviations (SD) of posterior RZSM estimates when VSM observations from MicroWEX-2 at 0–5 cm were assimilated using EnKF, PF-RR2, and PF-RR3 algorithms.

Table 3
Root mean square errors and differences (RMSE/D) and average standard deviations (ASD) of the RZSM (m^3/m^2) averaged over the whole season assimilating synthetic and MicroWEX-2 observations of VSM at 0–5 cm, every 3 days using the EnKF and PF algorithms.

Scenario	Synthetic		MicroWEX-2	
	RMSE	ASD	RMSD	ASD
EnKF	0.004	0.009	0.019	0.009
PF-NRNM	0.007	0.018	0.021	0.017
PF-RR	0.009	0.003	0.029	0.008
PF-RR2	0.007	0.008	0.023	0.010
PF-RR3	0.006	0.010	0.024	0.012
Openloop	0.025	0.024	0.032	0.025

Table 4
RMSE and ASD of the RZSM (m^3/m^2) averaged over different growth stages of corn when assimilating synthetic observations of VSM at 0–5 cm every 3 days using the EnKF and PF algorithms.

Scenario	Early stage		Mid stage		Reproductive stage	
	RMSE	ASD	RMSE	ASD	RMSE	ASD
EnKF	0.006	0.008	0.002	0.010	0.001	0.008
PF-NRNM	0.008	0.015	0.007	0.019	0.007	0.020
PF-RR	0.008	0.005	0.008	0.002	0.011	0.001
PF-RR2	0.007	0.008	0.007	0.007	0.005	0.009
PF-RR3	0.007	0.009	0.004	0.009	0.008	0.013
Openloop	0.016	0.015	0.024	0.026	0.035	0.034

for improved estimation rather than allowing them to converge earlier during assimilation as happened in the case of the PF-RR algorithm.

Table 5
RMSD and ASD of the RZSM (m^3/m^2) averaged over different growth stages of corn when assimilating MicroWEX-2 observations of VSM at 0–5 cm every 3 days using the EnKF and PF algorithms.

Scenario	Early stage		Mid stage		Reproductive stage	
	RMSD	ASD	RMSD	ASD	RMSD	ASD
EnKF	0.012	0.008	0.015	0.009	0.029	0.009
PF-NRNM	0.009	0.014	0.013	0.017	0.035	0.019
PF-RR	0.021	0.011	0.028	0.008	0.037	0.003
PF-RR2	0.018	0.012	0.016	0.010	0.034	0.006
PF-RR3	0.007	0.010	0.016	0.011	0.040	0.015
Openloop	0.009	0.017	0.021	0.026	0.055	0.034

4.2.2. Parameter convergence

The means and standard deviations of the parameters estimated using the four PF algorithms and the EnKF are given in Table 7. Similar to the synthetic case, since parameter diversity is not implemented via resampling, the PF-RR algorithm converges on DoY 137 leading to unrealistic zero ASD values. As mentioned earlier, among the four parameters, porosity is most strongly correlated to RZSM. Correlations between K_{sat} and RZSM are sensitive to precipitation events with stronger correlations during precipitation events and weaker correlations during dry-down periods. The PR-NRNM and EnKF algorithms converge to similar ϕ and K_{sat} values (see Figs. 11 and 12) and also provide similar overall improvements to RZSM estimates. The PF-NRNM and PF-RR2 algorithms also converge to similar porosity values of 0.2842 and 0.2514 in the reproductive stage offering RMSDs of 0.035 and 0.034 m^3/m^2, respectively. Although the PF-RR2 algorithm converges to smaller values of K_{sat} compared to the PF-NRNM algorithm, weaker correlation between K_{sat} and RZSM compared to porosity has lesser

Table 6

Means and standard deviations (Std. dev.) of the four soil parameters in the LSP-DSSAT model when assimilating synthetic observations of VSM at 0–5 cm every 3 days using the EnKF and PF algorithms at the end of season.

| Scenario	ϕ		ψ_0		λ		K_{sat}	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
EnKF	0.2892	0.0383	0.4927	0.1863	0.6138	0.1566	5.6×10^{-4}	2.5×10^{-4}
PF-NRNM	0.3057	0.0867	0.4833	0.2359	0.5559	0.2331	4.9×10^{-4}	2.7×10^{-4}
PF-RR	0.2912	0.0000	0.1931	0.0000	0.4073	0.0000	1.2×10^{-4}	0.0000
PF-RR2	0.3217	0.0475	0.1788	0.0884	0.4753	0.0844	2.7×10^{-4}	7.5×10^{-5}
PF-RR3	0.3318	0.0604	0.5178	0.2089	0.5802	0.1521	4.6×10^{-4}	2.3×10^{-4}
Truth	0.2760		0.6554		0.7572		5.8×10^{-4}	
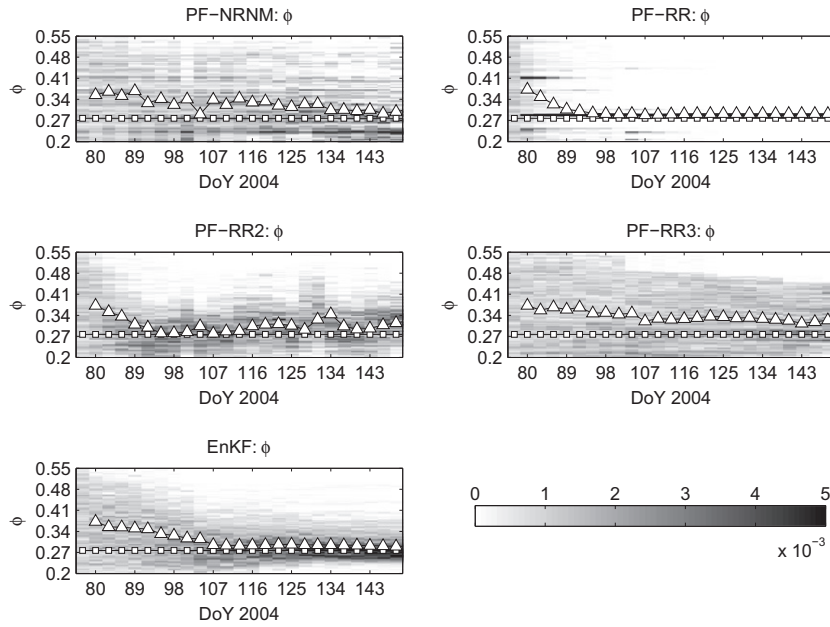
Fig. 7. Posterior distributions of porosity (ϕ) during the growing season for the four PF algorithms and EnKF when synthetic observations of VSM were assimilated. The Square and triangle markers represent the true and mean values, respectively. Grayscale bar represents the probability values of ϕ.
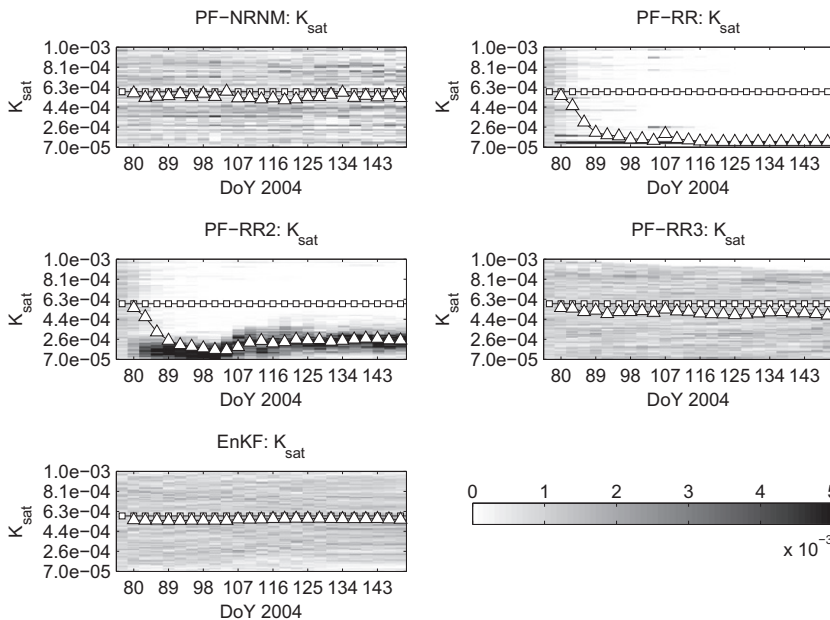

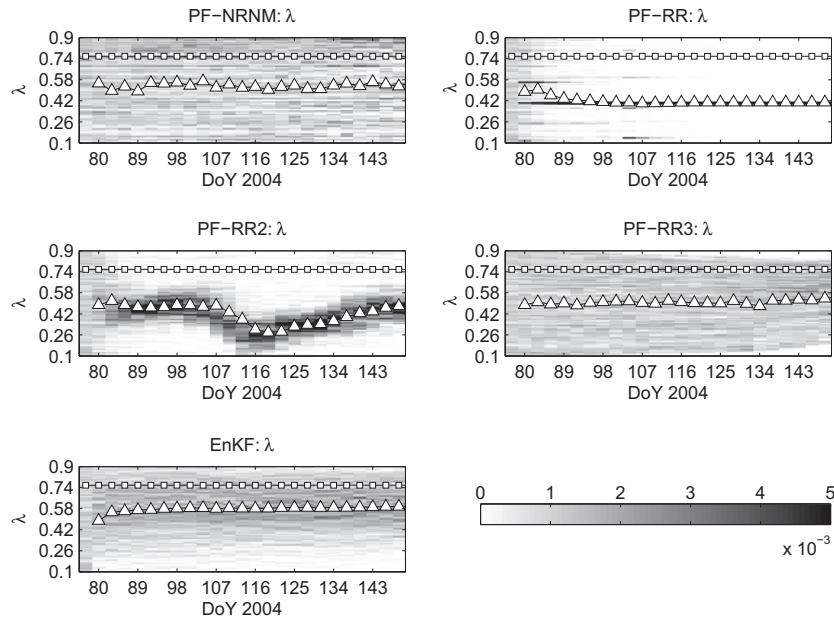
Fig. 8. Posterior distributions of saturated hydraulic conductivity (K_{sat}) during the growing season for the four PF algorithms and EnKF when synthetic observations of VSM were assimilated. The square and triangle markers represent the true and mean values, respectively. Grayscale bar represents the probability values of K_{sat}.
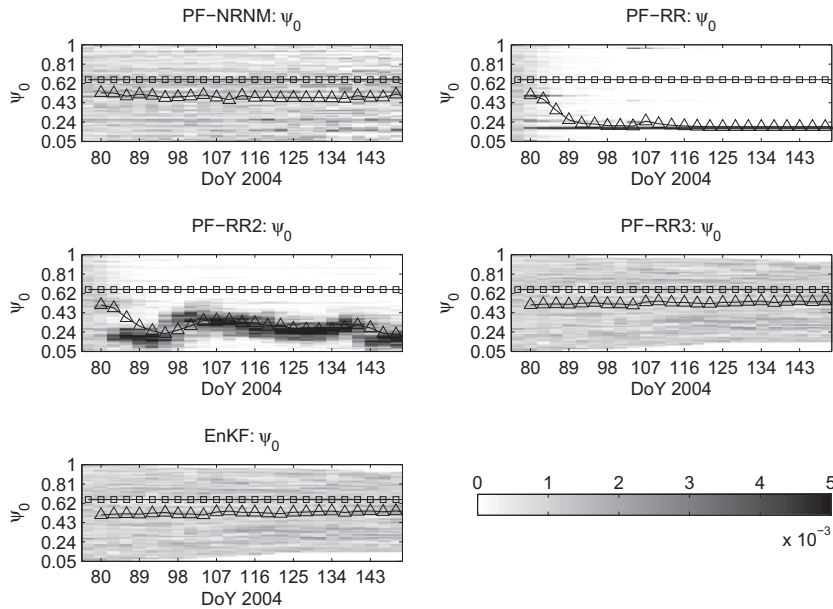
Fig. 9. Posterior distributions of pore-size index (λ) during the growing season for the four PF algorithms and EnKF when synthetic observations of VSM were assimilated. The square and triangle markers reprezent the true and mean values, respectively. Grayscale bar represents the probability values of λ.

Fig. 10. Posterior distributions of air-entry pressure (ψ_0) during the growing season for the four PF algorithms and EnKF when synthetic observations of VSM were assimilated. The square and triangle markers represent the true and mean values, respectively. Grayscale bar represents the probability values of ψ_0.

impacts on RZSM estimates. The ϕ values estimated by the PF-RR and PF-RR3 (see Fig. 11) algorithms are significantly different from the other algorithms and offer reduced improvements, especially in the reproductive stage, where correlations between porosity and RZSM are the strongest. The estimated λ and ψ_0 parameters significantly vary between algorithms (see Figs. 13 and 14) since they are the least correlated to RZSM.

4.2.3. Comparison of PF with EnKF

The EnKF offers a better performance over the PF with a 24% RMSD reduction (0.019 m³/m²) over openloop estimates, similar to the synthetic case. Fig. 6 shows the means and standard deviations (SDs) of the RZSM estimates obtained from the posterior

distributions using the PF-RR2, PF-RR3, and EnKF algorithms when MicroWEX-2 observations are assimilated. Unlike the synthetic case, the differences in the performance of the EnKF and the PF are minimal in the MicroWEX-2 case because the imperfect knowledge of model physics affects both the filtering techniques in a similar manner. The RMSDs were slightly higher for the PF algorithms because they are more sensitive to model and observation errors than EnKF.

While the EnKF exhibits the lowest RMSD over the entire season, variations within the season can be observed in the performance. During the early stage, the assimilation decreases the uncertainty (ASD) of estimates, but does not reduce the RMSDs. While RMSD denotes the degree of accuracy between estimates

Table 7
Means and standard deviations (Std. dev.) of the four soil parameters in the LSP-DSSAT model when assimilating MicroWEX-2 observations of VSM at 0–5 cm every 3 days using the EnKF and PF algorithms at the end of season.

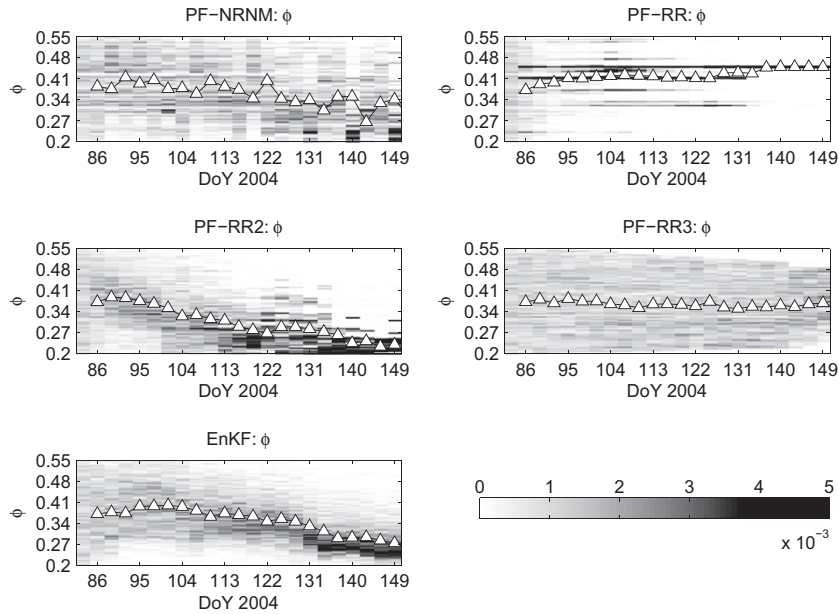| Scenario	ϕ		ψ_0		λ		K_{sat}	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
EnKF	0.2662	0.0484	0.3722	0.1892	0.4174	0.1518	4.1×10^{-4}	2.3×10^{-4}
PF-NRNM	0.2842	0.0755	0.4685	0.2519	0.5852	0.2294	5.0×10^{-4}	2.9×10^{-4}
PF-RR	0.4494	0.0000	0.1010	0.0000	0.5101	0.0000	6.9×10^{-4}	0.0000
PF-RR2	0.2514	0.0349	0.2822	0.0610	0.2076	0.0520	2.2×10^{-4}	1.4×10^{-5}
PF-RR3	0.3820	0.0583	0.4744	0.2164	0.6615	0.1027	5.0×10^{-4}	2.5×10^{-4}

Fig. 11. Posterior distributions of porosity (ϕ) during the growing season for the four PF algorithms and EnKF when MicroWEX-2 observations of VSM were assimilated. The triangle markers represent the mean values. Grayscale bar represents the probability values of ϕ.
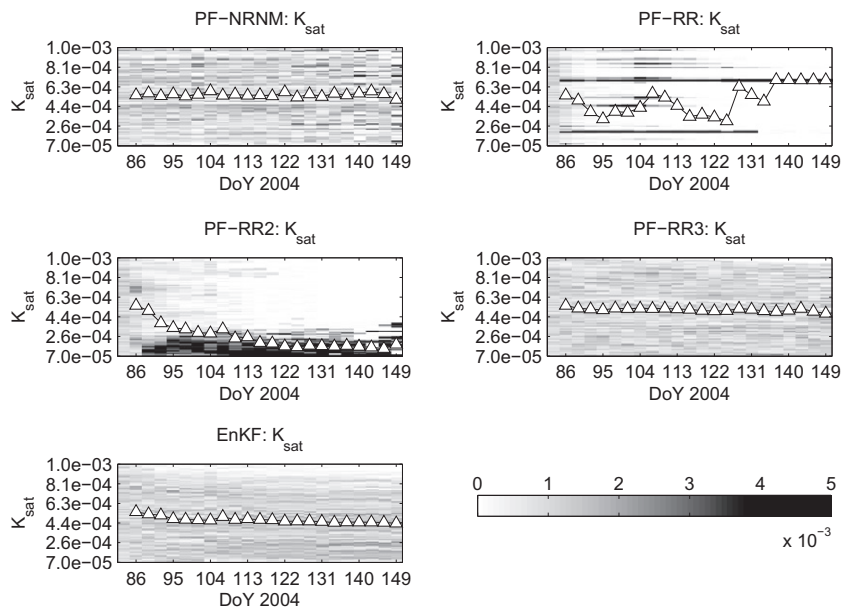

Fig. 12. Posterior distributions of saturated hydraulic conductivity (K_{sat}) during the growing season for the four PF algorithms and EnKF when MicroWEX-2 observations of VSM were assimilated. The triangle markers represent the mean values. Grayscale bar represents the probability values of K_{sat}.

and the observation, ASD represents the uncertainty in the estimates. The ASDs of RZSM estimates are similar for the EnKF, PF-RR2, and PF-RR3 (0.008–0.011 m^3/m^2) and are the highest for the PF-NRNM (0.016 m^3/m^2). During the mid and reproductive stages, all the algorithms, except for the PF-RR, provide significant reductions in both the RMSDs and ASDs over openloop estimates (see Table 5). Higher RMSD values in the reproductive stage compared to those obtained in the synthetic experiment are likely due to the imperfect implementation of soil and plant characterization, such as interactions between the roots and soil in the model compared to field measurements. Typically, land is ploughed and disced up to 30 cm at the beginning of the growing season. This may results in heterogeneity of soil properties. Assumptions of homogeneous soil in the LSP-DSSAT model in this study may not be sufficient to capture reality. In addition, the DSSAT model may need to be updated for canopy and root distribution parameters, along with the soil parameters in the LSP model. These parameters have maximum impact during the reproductive stage and the RMSDs are relatively higher during this stage.

5. Summary and conclusions

In this study, we compared the performance of four PF-based algorithms with the EnKF in improving RZSM estimates from assimilating synthetic and field observations of soil moisture under dynamic vegetation. Both the filtering techniques offered significant reductions in RMSE/RMSD and ASD in the synthetic and MicroWEX-2 cases compared to the openloop.
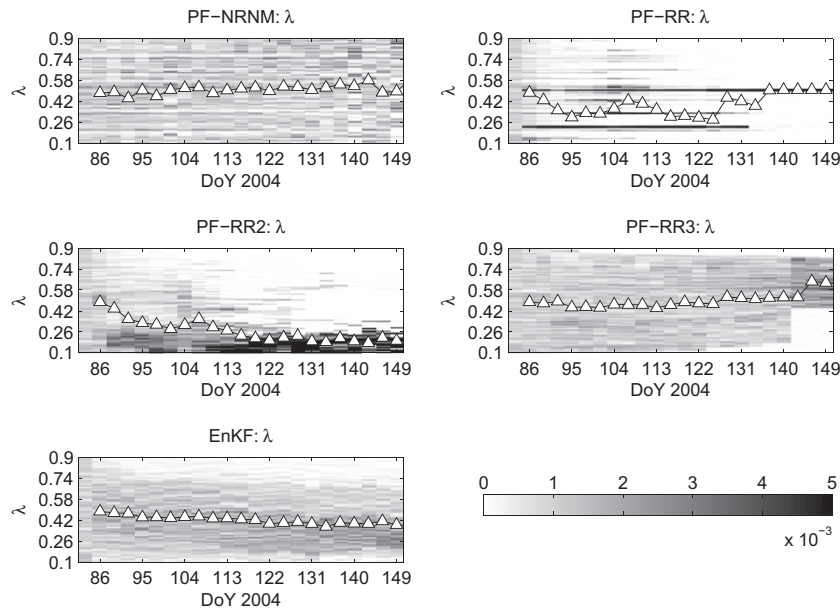

Fig. 13. Posterior distributions of pore-size index (λ) during the growing season for the four PF algorithms and EnKF when MicroWEX-2 observations of VSM were assimilated. The triangle markers represent the mean values. Grayscale bar represents the probability values of λ.
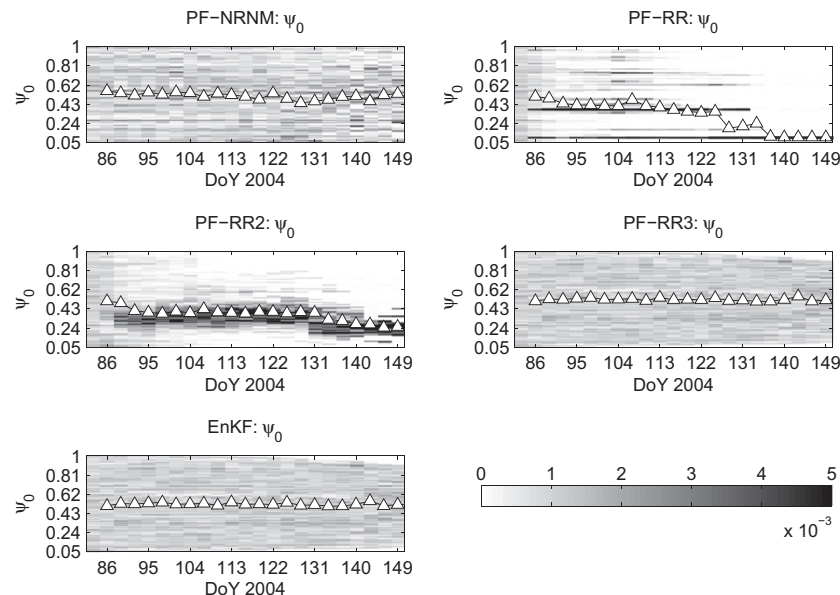

Fig. 14. Posterior distributions of air-entry pressure (ψ_0) during the growing season for the four PF algorithms and EnKF when MicroWEX-2 observations of VSM were assimilated. The triangle markers represent the mean values. Grayscale bar represents the probability values of ψ_0.

In the synthetic experiment, the PF-RR2, PF-RR3, and the PF-NRNM algorithms that retained parameter diversity offered similar reductions in RMSEs (72–76%). The EnKF offered the best RZSM estimate with a 84% reduction in RMSE compared to openloop estimates when synthetic observations were assimilated. The differences in RZSM estimates between the PFs and the EnKF were due to the differences in the update mechanisms of the unobserved states and parameters. In the PF, the posterior weights used for updating the unobserved states and parameters are obtained based only upon the likelihood of the observed states. If the proposal function in the PF is assumed to be the state equation, the relation between the observed and unobserved states and parameters are not embedded within the computations of the posterior weights. However, in the EnKF, the state covariance matrix includes correlations between unobserved states and parameters and the observed states and thus regulates the updates of the unobserved states and parameters. Therefore, unless assimilation is performed at times of highest sensitivity, PFs may lead to parameter estimates that are farther from the truth. This behavior was observed with K_{sat}, ψ_0, and λ estimates because assimilation was performed during times of lowest sensitivity. Since porosity, ϕ, was highly sensitive to soil moisture throughout the season, all algorithms produced estimates that were close to the true value. While performing simultaneous update of states and parameters, augmenting parameters that are not significantly sensitive to the observed states may impact the performance of the PF more adversely than the EnKF.

Unlike the synthetic case, the differences in the performance of the PFs and the EnKF were minimal in the MicroWEX-2 case because imperfect knowledge of model physics affected both the filtering techniques in a similar manner. The PF-NRNM and EnKF algorithms offer 16% and 20% reductions in RMSDs respectively over the openloop simulations. Both the algorithms converged to similar values of ϕ and K_{sat}, parameters most correlated to soil moisture, and therefore offered similar improvements to RZSM estimates. Although assimilation in the MicroWEX-2 case offered improvements in RZSM estimation, they were not as high as compared to the synthetic experiment (<24% RMSD reduction in MicroWEX-2). This is because, in this study, only the errors in soil and precipitation/irrigation forcings were considered. The observed differences in the filter performance between synthetic and field observations may indicate errors in model biophysics that were not considered here, such as heterogeneous soil conditions or errors in predictions of root distribution parameters. Presence of unaccounted biophysical errors in the model affects PF performance more than the EnKF leading to higher RMSDs in the reproductive stage for the PF estimates.

Our synthetic and field experiments reveal that the efficiency of PFs in representing non-Gaussian distributions and capturing non-stationary soil–plant interactions under dynamic vegetation largely depend on model error formulation. Our experiments suggest that both the EnKF and the PF can offer similar improvements over openloop estimates, but recommendation for either filtering strategy can only be made when their performances are assessed for various other applications involving coupled models and when non-linear observations are assimilated.

Acknowledgements

References

[1] Alagarswamy G, Boote K, Allen L, Jones J. Evaluating the cropgro-soybean model ability to simulate photosynthesis response to carbon dioxide levels. J Agron 2006;93:34–42.
[2] Beven K, Freer J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. J Hydrol 2001;249(1–4):11–29.
[3] Boote K, Jones J. Simulation of crop growth, 1998.
[4] Boote K, Jones J, Hoogenboom G, Wilkerson G. Evaluation of the CROPGRO-Soybean model over a wide range of experiments. Boston, MA, 1997.
[5] Braga R, Jones J. Using optimization to estimate soil inputs of crop models for use in site-specific management. Trans ASAE 2004;47(5):1821–31.
[6] Campbell Scientific, Cs616 and cs625 water content reflectometers. In: Tech Report, Logan, Utah, 2006. Available at <http://www.campbellsci.com/documents/manuals/cs616.pdf>.
[7] Casanova J, Judge J. Estimation of energy and moisture fluxes for dynamic vegetation using coupled SVAT and crop-growth models. Water Resour Res 2008;44:W07415. doi:10.1029/2007WR006503.
[8] Casanova J, Judge J, Jones J. Calibration of the ceres-maize model for linkage with a microwave remote sensing model. Trans Am Soc Agric Eng 2006;49(3).
[9] Chen T, Morris J, Martin E. Particle filters for state and parameter estimation in batch processes. J Process Control 2005;15:655–73.
[10] Chung Y. A snow-soil–vegetation–atmosphere-transfer/radiobrightness model for wet snow. PhD thesis, The University of Michigan, 2007.
[11] Ciach G. Local random errors in tipping-bucket rain gauge measurements. J Atmos Ocean Technol 2003;20:752–9.
[12] Crow W, Kustas W, Prueger J. Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. Remote Sens Environ 2008;112:1268–81.
[13] Dardanelli J, Calmon M, Jones J, Andriani J, Diaz M, Collino D. Use of a crop model to evaluate soil impedance and root clumping effects on soil water extraction in three argentine soils. Trans ASAE 2003;46(4):1265–75.
[14] De Lannoy G, Reichle R, Houser P, Pauwels V, Verhoest N. Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour Res 2007;43(W09410). doi:10.1029/2006WR005449.
[15] de Vries D. Simultaneous transfer of heat and moisture in porous media. Trans Am Geophys Union 1958;39(5):909–16.
[16] Doucet A, de Freitas N, Gordon N. Sequential monte carlo methods in practice. New York: Springer; 2008. p. 1–581.
[17] Evensen G. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 1994;99(C5):10143–62.
[18] Evensen G. The ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynam 2003;53(4):343–67.
[19] Gelb A. Applied optimal estimation. The MIT Press; 1974 [Chapter 4, p. 102–55].
[20] Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-Gaussian bayesian state estimation. IEE Proc Radar Signal Process 1993;140(2):107–13.
[21] Goudriaan J. Crop micrometeorology: a simulation study. first ed. Wageningen, The Netherlands: Centre for Agricultural Publishing and Documentation; 1977.
[22] Habib E, Krajewski W, Kruger A. Sampling errors of tipping-bucket rain gauge measurements. J Hydrol Eng 2001;6(2):159–66.
[23] Han X, Li X. An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation. Remote Sens Environ 2008;112:1434–49.
[24] Hanson J, Rojas K, Schaffer M. Calibrating the root zone water quality model. Agron J 1999;91:171–7.
[25] Jagtap S, Jones J. Adaptation and evaluation of the cropgro-soybean model to predict regional yield and production. Agric Ecosyst Environ 2002;93:73–85.
[26] Jones J, Hoogenboom G, Poter C, Boote K, Hunt W, Wilkens P, et al. The DSSAT cropping system model. Eur J Agron 2003;18(3–4):235–65.
[27] Judge J, Abriola L, England A. Numerical validation of the land surface process component of an LSP/R model. Adv Water Res 2003;26(7):733–46.
[28] Judge J, Casanova JJ, Lin T, Tien KC, Jang M, Lanni O, et al. Field observations during the second microwave, water, and energy balance experiment (MicroWEX-2): from March 17 through June 3, 2004. Circular No. 1480. Tech. rep., Center for Remote Sensing, University of Florida, 2005. Available at <http://edis.ifas.ufl.edu/AE360>.
[29] Judge J, England A, Crosson C, Hornbuckle B, Boprie D, Kim E, et al. A growing season land surface process/radiobrightness model for wheat-stubble in the southern great plains. IEEE Trans Geosci Remote Sens 1999;37(5):2152–8.
[30] Judge J, England A, Metcalfe J, McNichol D, Goodison B. Calibration of an integrated land surface process and radiobrightness (LSP/R) model during summertime. Adv Water Res 2007:189–202.
[31] Judge J, England A, Metcalfe J, McNichol D, Goodison B. Calibration of an integrated land surface process and radiobrightness (LSP/R) model during summertime. Adv Water Resour 2008;31(1):189–202.
[32] Kerr Y, Waldteufel P, Wigneron J, Martinuzzi J, Font J, Berger M. Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission. IEEE Trans Geosci Remote Sens 2001;39(8):1729–35.
[33] Liu J, Chen R. Sequential monte carlo methods for dynamic systems. J Am Statist Assoc 1998;93(443):1032–44.

[34] Lizaso J, Batchelor W, Boote K, Westgate M, Rochette P, Moreno-Sotomayor A. Evaluating a leaf-level canopy assimilation model linked to ceres-maize. J Agron 2005;97:734–40.

[35] Lorenc A, Hammon O. Objective quality control of observations using Bayesian methods: theory, and a practical implementation. Quart J Roy Meteorol Soc 1988;114:515–43.

[36] Mavromatis T, Boote K, James W, Wilkerson G, Hoogenboom G. Repeatability of model genetic coefficients derived from soybean performance trials across different states. Crop Sci 2002;42:76–89.

[37] Monsivais-Huertero A, Graham W, Judge J, Agrawal D. Effect of simultaneous state-parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using EnKF. Adv Water Resour 2010;33(4): 468–84.

[38] Moradkhani H. Hydrologic remote sensing and land surface data assimilation. Sensors 2008;8:2986–3004.

[39] Moradkhani H, Sorooshina S, Gupta H, Houser P. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Res 2005;28(2):135–47.

[40] Munoz-Sabatier J, Rudiger C, Clavet J, Fritz N, Jarlan L, Kerr Y. Joint assimilation of surface soil moisture and LAI observations into a land surface model. Agric Forest Meteorol 2008;148(8–9):1362–73.

[41] Njoku E, Jackson T, Lakshmi V, Chan T, Nghiem S. Soil moisture retrieval from AMSR-E. IEEE Trans Geosci Remote Sens 2003;41(2):215–29.

[42] Pauwels V, Verhoest N, De Lannoy G, Guissard V, Lucau C, Defourny P. Optimization of a coupled hydrology-crop growth model through the assimilation of observed soil moisture and leaf area index using an ensemble Kalman filter. Water Resour Res 2007;43:W04421. doi:10.1029/2006WR004942.

[43] Peixoto J, Oort A. Physics of climate. New York: American Institute of Physics; 1992.

[44] Philip J, de Vries D. Moisture movement in porous materials under temperature gradients. Trans Am Geophys Union 1957;38(2):222–32.

[45] Qin J, Liang S, Yang K, Kaihotsu I, Liu R, Koike T. Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal. J Geophys Res 2009:114.

[46] Reichle R, Koster R, Liu P, Mahanama S, Njoku E, Owe M. Comparison and assimilation of global soil moisture retrievals from the advanced microwave scanning radiometer for the earth observing system (AMSR-E) and the scanning multichannel microwave radiometer (SSMR). J Geophys Res 2007;112(D09108). doi:10.1029/2006JD008033.

[47] Rossi C, Nimmo J. Modeling of soil water retention from saturation to oven dryness. Water Resour Res 1994;30:701–8.

[48] Timsina J, Boote K, Duffield S. Evaluating the cropgro soybean model for predicting impacts of insect defoliation and depodding. J Agron 2007;99: 148–57.

[49] Trenberth K. Climate system modeling. New York: Cambridge University Press; 1995.

[50] van Delft G, El Serafy G, Heemink A. The ensemble particle filter (EnPF) in rainfall-runoff models. Stoch Environ Res Risk Assess 2009;1. doi:10.1007/s00477-008-0301-z.

[51] Verseghy S, McFarlane N, Lazare M. CLASS-A Canadian land surface scheme for GCMs, II. Vegetation model and coupled runs. Int J Climatol 1993;13:347–70.

[52] Weerts A, El Serafy Y. Particle filtering and ensemble Kalman Filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resour Res 2006;42. doi:10.1029/2005WR004093.

[53] Whitfield B, Jacobs J, Judge J. Intercomparison study of the land surface process model and the common land model for a prairie wetland in Florida. J Hydrometeorol 2006;7(6):1247–58.

[54] Zhou Y, McLaughlin D, Entekhabi D. Assessing the performance of the ensemble Kalman filter for land surface data assimilation. Mon Weather Rev 2005;134:2128–42.