Hydrology and Earth System Sciences

Simultaneous estimation of land surface scheme states and parameters using the ensemble Kalman filter: identical twin experiments

S. Nie[1], J. Zhu[2], and Y. Luo[1]

[1]National Climate Center, China Meteorological Administration, Beijing, China
[2]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing China

Abstract. The performance of the ensemble Kalman filter (EnKF) in soil moisture assimilation applications is investigated in the context of simultaneous state-parameter estimation in the presence of uncertainties from model parameters, soil moisture initial condition and atmospheric forcing. A physically based land surface model is used for this purpose. Using a series of identical twin experiments in two kinds of initial parameter distribution (IPD) scenarios, the narrow IPD (NIPD) scenario and the wide IPD (WIPD) scenario, model-generated near surface soil moisture observations are assimilated to estimate soil moisture state and three hydraulic parameters (the saturated hydraulic conductivity, the saturated soil moisture suction and a soil texture empirical parameter) in the model. The estimation of single imperfect parameter is successful with the ensemble mean value of all three estimated parameters converging to their true values respectively in both NIPD and WIPD scenarios. Increasing the number of imperfect parameters leads to a decline in the estimation performance. A wide initial distribution of estimated parameters can produce improved simultaneous multi-parameter estimation performances compared to that of the NIPD scenario. However, when the number of estimated parameters increased to three, not all parameters were estimated successfully for both NIPD and WIPD scenarios. By introducing constraints between estimated hydraulic parameters, the performance of the constrained three-parameter estimation was successful, even if temporally sparse observations were available for assimilation. The constrained estimation method can reduce RMSE much more in soil moisture forecasting compared to the non-constrained estimation method and traditional non-parameter-estimation assimila-

tion method. The benefit of this method in estimating all imperfect parameters simultaneously can be fully demonstrated when the corresponding non-constrained estimation method displays a relatively poor parameter estimation performance. Because all these constraints between parameters were obtained in a statistical sense, this constrained state-parameter estimation scheme is likely suitable for other land surface models even with more imperfect parameters estimated in soil moisture assimilation applications.

1 Introduction

Soil moisture is a key state variable controlling the partitioning of water and energy fluxes at the land surface. It influences the surface water cycle, and, consequently, the latent heat flux and surface energy balance. As a numerical simulation to the realistic land surface state, the land surface model (LSM) is a popular tool providing proper soil moisture initial conditions for numerical weather prediction models and climate models. However, given its simplified physical and mathematical processes, LSM only approximates actual processes in nature. Uncertainties in hydrodynamic processes, model variables and model parameters lead to large errors in the simulation of soil moisture condition. The proper initialization of soil moisture conditions in LSMs remains an open issue in meteorological and hydrological research.

Modern data assimilation technique effectively accounts for this issue. Merging information from uncertain soil moisture observations and uncertain land model predictions optimally, this technique can improve the estimation of the soil moisture state in LSMs (Houser et al., 1998; Reichle et al., 2001a, b). As a Monte Carlo approximation to the traditional Kalman filter (Kalman and Bucy, 1961), the ensemble

Correspondence to: S. Nie
(niesp@cma.gov.cn)

Kalman filter (EnKF) method, first introduced by Evensen (1994), has received an increasing attention and been widely used in recent years. By propagating an ensemble of state vectors in parallel, such that each state vector represents a particular realization of generated model replicates, it provides a flow-dependent background error covariance obtained at each update and adjusts the background optimally to newly available observations. In recent years, the EnKF has been successfully applied to different soil moisture assimilation problems (Walker and Houser, 2001; Reichle et al., 2002a, b, 2008; Reichle and Koster, 2005; Crow and Van Loon, 2006; Crow and van den Berg, 2010; Ni-Meister et al., 2006; Zhang et al., 2010). In most of these studies, however, the EnKF was only used for estimating time-varying state variables under the presumption that model parameters were specified in advance by calibrations. In common calibration methods, model parameters are adjusted by a historical batch of measurements so that the behavior of model approximates, as closely and consistently as possible, the observed behavior of the real land system over some period of time (Niyogi et al., 2002; Xia et al., 2002; Coudert et al., 2006). There exist two main weaknesses in these calibration approaches: (i) they cannot include information from new observations, and (ii) as commonly practiced, they ignore errors from initial condition and atmospheric forcing data. Because the EnKF accounts for a wide range of possible model errors easily (Evensen, 2003), it has the potential to overcome these two drawbacks by explicitly accounting for all sources of uncertainty and developing a simultaneous treatment of state and parameter estimation to refine its assimilation performance.

By means of the state augmentation technique (Anderson, 2001), model parameter estimation can easily be included in the framework of the EnKF. The principle of state augmentation is that model parameters can be considered as "pseudo" model states along with conventional state variables, and then the error covariance sampled by ensemble members can be used directly to update those model parameters in exactly the same manner as for the conventional state variables. Recently, the state-parameter estimation approach was successfully applied in atmospheric (Aksoy et al., 2006), oceanic (Annan et al., 2005), and ecological (Chen et al., 2008) assimilation fields. In the hydrological field, the idea of state-parameter estimation was first introduced by Moradkhani et al. (2005a, b), using the EnKF and particle filter. The results were promising, and nearly all of the parameters were well estimated. Subsequently, this approach was widely used in many hydrological studies (Franssen and Kinzelbach, 2008; DeChant and Moradkhani, 2010; Leisenring and Moradkhani, 2010; Wang et al., 2009; Montzka et al., 2011). However, similar simultaneous estimations of soil moisture and hydraulic parameters are few at present, except for certain studies available using the particle filter method (Qin et al., 2009; Montzka et al., 2011). Hence, the applicability of simultaneous state-parameter estimation approach

in soil moisture assimilation under the EnKF framework requires careful study. As soil hydraulic parameters in many LSMs are generally difficult to measure at regional scales, they are usually estimated according to soil texture class by pedotransfer functions (Dickinson et al., 1993; Wosten et al., 2001). Because such functions are obtained from experimental data under specific conditions, uncertainty arises from the extrapolation of these functions to other regional or global LSM scales (Cornelis et al., 2001). According to the study of Montaldo et al. (2007), when key LSM soil hydraulic parameter are estimated poorly, large errors in these parameters can result in a persistent bias in soil moisture prediction and cause soil moisture assimilation approaches to fail. Therefore, this paper first investigates the capability of the EnKF in reducing errors in posterior values of hydraulic parameters by simultaneous state-parameter estimation approach in soil moisture assimilation.

Despite promising results obtained from the applications of the EnKF in parameter estimation, some deficiencies still exist in these studies. The noticeable decline of estimation performance occurs when multiple imperfect parameters are estimated simultaneously. When the number of estimated parameters increases to a certain extent, it is difficult for all estimated parameters to converge to their "true" values entirely even with long enough estimation periods (Moradkhani et al., 2005a; Aksoy et al., 2006; Jung et al., 2010; Montzka et al., 2011). One possible reason is that constraints among different parameters, arising from different physical relationships, are often neglected in the data assimilation framework (Wang et al., 2009). Because most of studies applied for simultaneous state-parameter estimation in hydrologic field are based on unconstrained assimilation methods (Moradkhani et al., 2005a, b; Franssen and Kinzelbach, 2008; DeChant and Moradkhani, 2010; Leisenring and Moradkhani, 2010; Montzka et al., 2011), this problem remains a challenge in the application of multiple parameter estimation in hydrologic data assimilation. Currently, only a few studies have considered constraints in hydrologic data assimilation field. Pan and Wood (2006) used a two-step EnKF approach to deal with the water balance constraint in the estimation of the terrestrial water budget. Wang et al. (2009) compared three methods that deal with inequality constraints in the EnKF framework by a conceptual hydrologic model for state estimation and sequential parameter learning. However, related studies for the application of equality constraints between hydrologic parameters using the EnKF assimilation approach based on a physical LSM are few in number at present. Therefore, this paper also discusses the applicability of a new constrained parameter estimation procedure for simultaneous multi-parameter estimation in soil moisture data assimilation under the EnKF framework.

2 Land surface model and parameter estimation framework

2.1 Land surface model

The land surface model used is the Atmosphere-Vegetation Interaction Model (AVIM) (Ji, 1995), which contains a physical process mode and a vegetation biological process mode. Detailed descriptions of this model are given by Ji and Hu (1989) and Ji (1995). The version used in this study only considers the physical process mode, a typical soil-vegetation-atmosphere (SVAT) type model developed by Ji and Hu (1989). This model includes three soil layers with thicknesses of 0.1, 0.9, and 3.6 m from ground. The layer-averaged soil moisture is modeled for each of the three soil layers. The free drainage assumption is used for the bottom layer. The change of soil moisture in the near surface layer and root zone layer over a time step is controlled by the change in water flux over these two layers. Richards' equation for unsaturated flow is used for the simulation of this flux, expressed as:

$$F(z) = -k(z)\frac{d\psi(z)}{dz}\bigg|_z + k(z)z \neq 0 \tag{1}$$

where z is the depth and $F(z)$ is the soil water flux. The unsaturated hydraulic conductivity $k(z)$ and unsaturated soil water suction $\psi(z)$ are defined in Clapp and Hornberger (1978):

$$k(z) = k_{sat}(\frac{\delta}{\delta_{sat}})^{2b+3} \tag{2}$$

$$\psi(z) = \psi_{sat}(\frac{\delta}{\delta_{sat}})^{-b} \tag{3}$$

where δ and δ_{sat} are the unsaturated and saturated soil moisture; k_{sat} and ψ_{sat} are the saturated hydraulic conductivity and soil moisture suction respectively; and b is a soil texture empirical parameter. In this paper, parameters k_{sat}, ψ_{sat} and b were chosen for estimation.

2.2 Parameter estimation framework using the EnKF

Parameter estimation frameworks used in this paper are based on the EnKF. The comprehensive presentation of the standard EnKF is given by Evensen (2003), and this subsection represents primarily the modifications to the standard EnKF after considering simultaneous state-parameter estimation in its framework.

2.2.1 State-parameter estimation without constraint

To extend the applicability of the EnKF to state-parameter estimation, building an evolution of parameter similar to that of the model state variable is needed. By adding mean-zero

Gaussian random noise $\boldsymbol{\tau}_{t-1}^i$ with covariance $\mathbf{Q}_{t-1}^\theta$ to parameter $\boldsymbol{\theta}^i$, the evolution of parameter can be expressed in the form of:

$$\boldsymbol{\theta}_t^{i-} = \boldsymbol{\theta}_{t-1}^{i+} + \boldsymbol{\tau}_{t-1}^i, \boldsymbol{\tau}_{t-1}^i \sim N(0, \mathbf{Q}_{t-1}^\theta) \tag{4}$$

The superscripts "−" and "+" refer to states in the forecast step and update step, respectively; the superscript "i" indicates the ensemble members. When multiple parameters are estimated simultaneously, perturbations on different parameters are considered as mutually independent without cross-parameter constraint. The covariance of θ_t^{i-} is a diagonal matrix.

With artificially perturbed parameters, the time evolution for each ensemble member of state vector $\boldsymbol{x}^i$ in the EnKF is:

$$\boldsymbol{x}_t^{i-} = \boldsymbol{f}_t(\boldsymbol{x}_{t-1}^{i+}, \boldsymbol{u}_{t-1}^i, \boldsymbol{\theta}_{t-1}^{i+}), i = 1, 2, ..., n. \tag{5}$$

where $\boldsymbol{x}_t^{i-}$ is the i-th forecast ensemble member at time t and $\boldsymbol{x}_{t-1}^{i+}$ is the i-th updated ensemble member at time $t-1$. The nonlinear operator $\boldsymbol{f}(.)$ denotes the land surface model processes containing state vectors $\boldsymbol{x}^i$, forcing data vectors $\boldsymbol{u}^i$, and model parameter vectors $\boldsymbol{\theta}^i$. The forcing data perturbations are made by adding mean-zero Gaussian noise $\boldsymbol{\mu}_{t-1}^i$ with covariance $\mathbf{Q}_{t-1}^u$ to the forcing data at each time step:

$$\boldsymbol{u}_{t-1}^i = \boldsymbol{u}_{t-1} + \boldsymbol{\mu}_{t-1}^i, \boldsymbol{\mu}_{t-1}^i \sim N(0, \mathbf{Q}_{t-1}^u) \tag{6}$$

When observations are available, each ensemble member of state vector and parameter vector is updated as follows:

$$\begin{pmatrix} \boldsymbol{x}_t^{i+} \\ \boldsymbol{\theta}_t^{i+} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_t^{i-} \\ \boldsymbol{\theta}_t^{i-} \end{pmatrix} + \mathbf{K}_t^{\theta,x}(\boldsymbol{y}_t^i - \mathbf{H}_t \boldsymbol{x}_t^{i-}) \tag{7}$$

where $\mathbf{H}_t$ is the measurement operator and $\boldsymbol{y}_t^i$ is the i-th member of observation ensemble generated by adding mean-zero random measurement error $\boldsymbol{\eta}_t^i$ with covariance $\mathbf{Q}_t^y$ to actual observation (Burgers et al., 1998):

$$\boldsymbol{y}_t^i = \boldsymbol{y}_t + \boldsymbol{\eta}_t^i, \boldsymbol{\eta}_t^i \sim N(0, \mathbf{Q}_t^y) \tag{8}$$

$\mathbf{K}_t^{\theta,x}$ is the Kalman gain matrix that considers state-parameter estimation. It is obtained by:

$$\mathbf{K}_t^{\theta,x} = \mathbf{P}_t^{\theta,x,-}\mathbf{H}_t^T(\mathbf{H}_t\mathbf{P}_t^{\theta,x,-}\mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{9}$$

where $\mathbf{P}_t^{\theta,x,-}$ and $\mathbf{R}_t$ are forecast error covariance matrix and observation error covariance matrix respectively. $\mathbf{P}_t^{\theta,x,-}$ is computed as the sample covariance from forecast ensemble of model state variables and parameters. It is an ensemble covariance matrix around the ensemble mean:

$$\mathbf{P}_t^{\theta,x,-} = \frac{1}{n-1}\mathbf{X}_t\mathbf{X}_t^T \tag{10}$$

where, $\mathbf{X}_t = [\boldsymbol{x}_t^{1-} - \bar{\boldsymbol{x}}_t^-, ..., \boldsymbol{x}_t^{n-} - \bar{\boldsymbol{x}}_t^-; \boldsymbol{\theta}_t^{1-} - \overline{\boldsymbol{\theta}}_t^-, ..., \boldsymbol{\theta}_t^{n-} - \overline{\boldsymbol{\theta}}_t^-]$ and $\bar{\boldsymbol{x}}_t^- = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_t^{i-}, \overline{\boldsymbol{\theta}}_t^- = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\theta}_t^{i-}$ denote the ensemble mean of forecast state variables and parameters, respectively.

2.2.2 State-parameter estimation with constraint

In fact, some statistical relationships exist between different model parameters (e.g., Cosby et al., 1984; Rawls et al., 1982; Schaap and Leij, 2000; van Genuchten, 1980; Zhuang et al., 2001). As additional information, these statistical constraints between parameters require assessment in the framework of the EnKF to perform better state-parameter estimation.

In the general case, the constraints are nonlinear, expressed as:

$$g_t^* = G_t(\theta_t^*) \tag{11}$$

where g_t^* denotes nonlinear constraints between different model parameters θ_t^* at time t. Without losing generality, the parameters, which are not contained in g_t^*, are defined as θ''. The constrained update to each ensemble member of state vectors x^i is computed as follows:

$$\begin{pmatrix} x_t^{i+} \\ \theta_t^{i+} \end{pmatrix} = \begin{pmatrix} x_t^{i-} \\ \theta_t^{i-} \end{pmatrix} + \mathbf{K}_t^{\theta'',g^*}(y_t^i - \mathbf{H}_t x_t^{i-}) \tag{12}$$

where $\mathbf{K}_t^{\theta'',g^*}$ is the Kalman gain matrix including constraints. It is obtained as follows:

$$\mathbf{K}_t^{\theta'',g^*} = \mathbf{P}_t^{\theta'',g^*,-}\mathbf{H}_t^T(\mathbf{H}_t \mathbf{P}_t^{\theta'',g^*,-}\mathbf{H}_t^T + \mathbf{R}_t)^{-1} \tag{13}$$

where $\mathbf{P}_t^{\theta'',g^*,-}$ is the constrained error covariance matrix of state ensembles and parameter ensembles. Each ensemble member will satisfy the constraints. When constraints exist among parameters, the definition of $\mathbf{Q}$ in Eq. (4) should have cross-parameter correlations and the generation of perturbations $\boldsymbol{\tau}$ should be taken the constraints into consideration. In Sect. 3 we provide a detailed description of how to generate perturbations according to the specified constraints considered in this paper.

3 Experiments background and approach

This study is based on a series of identical twin experiments taking soil moisture in the top two layers and parameters k_{sat}, ψ_{sat} and b as state variables in the EnKF. The design of the identical twin experiment is similar to that of Crow and Van Loon (2006), with the assumptions that the "true" states are model-generated and the source and magnitude of model errors and observation errors are perfectly known statistically. This approach avoids a number of key complexities facing to assimilate actual soil moisture observations and makes the parameter estimation behavior of the EnKF more transparent. Note that little information about the statistical properties of errors in realistic soil moisture assimilation may degrade the performance of the EnKF in parameter estimation.

Because this study investigates the feasibility of EnKF-based parameter estimation in soil moisture assimilation,

all experiments were conducted at point scale for computational simplification. Jiangji station (the outlet of the Shiguanhe sub-basin in the Huaihe River Basin) of the HUaihe river Basin EXperiment (HUBEX, China's contribution to GEWEX Asian Monsoon Experiment, Fujiyoshi et al., 2006) was chosen as the experiment site due to its comprehensive meteorological forcing data sets. In this station, the soil texture is sandy loam, and the vegetation type is broadleaf shrubs with bare soil. The experiment period covered all of 1998. During the Intensive Observation Period in 1998 (from 21 May to 31 August), hourly gauge-based precipitation, once daily air temperature, humidity, surface pressure and wind speed data sets were available in Jiangji. During other period in the year, daily observations of these meteorological forcing from the Gushi meteorological site (approximately 15 km from Jiangji) was used. Without incoming radiation observation, the radiation forcing data from the NCEP (National Centers for Environmental Prediction) reanalysis dataset version 1 was used as a substitute. All these forcing data sets were used to force the AVIM in Jiangji in all identical twin experiments with a time step of half hours for the model and one-day frequency for assimilating soil moisture "observations".

The "true" soil moisture state was obtained by integrating the AVIM from 1 January 1998 to 31 December 1998 (after a 2-yr spinup period from 1 January 1996 to 31 December 1997) with standard AVIM parameters (Ji and Hu, 1989; Ji, 1995) and the atmospheric forcing data described above. To obtain "prior" state of soil moisture, model error from three sources (i) parameters (k_{sat}, ψ_{sat} and b), (ii) soil moisture initial conditions, and (iii) precipitation and short-wave (long-wave) radiations (Margulis et al., 2002; Reichle et al., 2002b) were considered in prior model integration. Errors in parameters were generated by replacing the "truth states" values with assumed imperfect parameter values. These prior parameter values were distant from the true values within the parameter range of the AVIM for testing the validity of parameter estimation frameworks even when all estimated parameters had large parameter errors compared to their true values. Errors in soil moisture initial condition were generated by replacing the "truth states" values with assumed imperfect values and adding zero mean Gaussian noise with a standard deviation of 50 % of the assumed values. Errors in precipitation and radiation were imposed by adding mean-zero Gaussian random noises (Eq. 6) to the true forcing fields. Specific differences between "true" and "prior" integrations are listed in Table 1. Collectively, these differences in parameter, initial condition and forcing data were considered as "actual errors" and represent our imperfect understanding to the true soil moisture states. In all identical twin experiments, the "actual observation" assimilated was the near surface soil moisture. It was derived from the true state by adding mean-zero Gaussian random errors with a standard deviation of 0.02 cm^3 cm^{-3} once a day. Figure 1 displays the precipitation forcing, the "true" and "prior" soil moisture for

Table 1. Specific differences of soil hydraulic parameters, initial soil moisture condition, and meteorological forcing data between "true" and "prior" model integrations in the identical twin experiments.

Variables & Parameters	Units	True	Prior
Saturated hydraulic conductivity	$\mathrm{m\,s^{-1}}$	5.23×10^{-6}	5.0×10^{-5}
Empirical parameter b	–	4.74	12.0
Saturated soil moisture suction	m	-0.218	-0.7
Initial soil moisture	$\mathrm{cm^3\,cm^{-3}}$	2-yr spinup values	0.12 for both two layers and adding zero mean Gaussian noise with a standard deviation of 50 % of this initial value
Precipitation	$\mathrm{mm\,(day)^{-1}}$	Gauge-based data and NCEP dataset 1	Adding Gaussian noise with mean square deviation of 20 % to the true values once daily and the minimum mean square deviation is limited to $2\,\mathrm{mm\,(day)^{-1}}$
Long- and short-waves radiations	$\mathrm{W\,m^{-2}}$	NCEP dataset 1	Adding Gaussian noise with mean square deviation of 30 % to the true values once daily

top two layers and the "actual observation" of near surface soil moisture used in identical twin experiment. For errors in hydraulic parameters, initial condition and atmospheric forcing data, there were significant deviations from prior states to true states of soil moisture in both layers. Perturbations on forcing data, soil moisture initial conditions and forcing data created large enough ensemble spreads on prior soil moisture for both two layers to maintain the uncertainty in the modeling prediction. In both soil layers, the ensemble spreads narrowed slowly with time and finally maintained an approximate a value of $0.02\,\mathrm{cm^3\,cm^{-3}}$, comparable to the "observed error" in the soil moisture observations.

Given the statistical properties of model errors and observation errors, the EnKF attempts to modify prior state back to the true state by assimilating "actual observations" (Crow and Van Loon, 2006). In the filter, the number of ensemble size was set to 100 in all experiments, achieving a balance between the computational effort of processing a large number of runs and the need for having a sufficiently large set of ensembles to characterize the ensemble distribution. The ensemble of soil moisture initial values was generated by adding zero mean Gaussian noise with a standard deviation of 50 % to the prior values at the first time step. The ensemble of forcing data was generated by perturbing prior forcing data with the same statistical properties as the actual forcing data errors once a day. The random perturbation method was also applied to obtain the ensemble of model parameters. A noticeable issue here was the magnitude of standard deviation of perturbation on model parameters because no straightforward guidance exists for the proper range of estimated parameter deviation. Given the difficulty of accurately estimating prior parameter uncertainty in real soil moisture assimilation settings, two kinds of initial pa-

rameter distribution scenarios (hereafter denoted as IPD scenario), the narrow IPD (NIPD) scenario and the wide IPD (WIPD) scenario, were considered in this study to fully explore the capability of the EnKF in estimating all these chosen hydraulic parameters. In NIPD scenario, initial ensemble spreads of all estimated parameters were small and unable to cover their true values, representing our insufficient knowledge about first guess error of these hydraulic parameters. In WIPD scenario, initial ensemble distributions of all parameters were large enough for covering their true values, indicating our satisfactory knowledge about first guess error of all parameters. The prior ensemble mean, standard deviation and uncertainty range of all these three parameters in both two IPD scenarios are listed in Table 2. Furthermore, because parameters were not dynamical variables, the variances of them were reduced at the update step but remained constant at the forecast step. This caused the variances of parameters to decrease progressively and may lead to filter divergence in parameters. To avoid filter divergence, therefore, a small perturbation was implemented on parameters continually in a similar way as that on forcing data according to Eq. (4) with a time interval of 10 days. The standard deviations of the small perturbations on k_{sat}, ψ_{sat}, and b were chosen as $0.8 \times 10^{-6}\,\mathrm{m\,s^{-1}}$, 0.01 m, and 0.08 respectively, which are much smaller than the orders of parameters themselves. Small standard deviation and sparse perturbation interval on these parameters in the filter processes avoids the behavior of the model being shocked for sharp changes of parameters in model integration.

In a non-constrained parameter estimation framework, perturbations on different parameters were considered mutually independent. In constrained estimation framework, statistical relationships between these hydraulic parameters
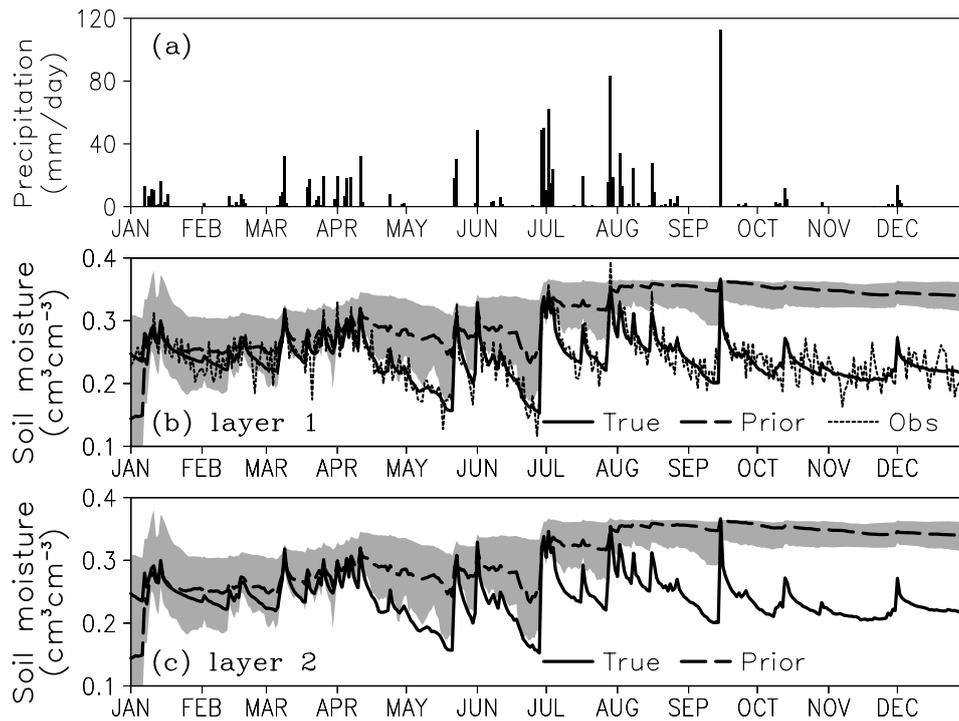
Fig. 1. The precipitation forcing **(a)** and soil moisture evolution of the near surface layer **(b)** and the root zone layer **(c)** in the identical twin experiment during 1998 at Jiangji station. The solid line represents the "true" state, the long dash line represents the "prior" state, the short dash line represents the "actual observation" of near surface soil moisture, and the shaded area represents the ensemble spread of the "prior" soil moisture.

Table 2. Specific values of uncertainty range, initial ensemble mean and standard deviation for parameters k_{sat}, ψ_{sat} and b in both narrow initial parameter distribution (NIPD) scenarios and wide initial parameter distribution (WIPD) scenarios.

Parameter	Units	Min	Max	NIPD scenario		WIPD scenario	
				Mean	Std. dev.	Mean	Std. dev.
k_{sat}	ms^{-1}	1.1×10^{-6}	8.2×10^{-5}	5.0×10^{-5}	8.0×10^{-6}	5.0×10^{-5}	3.3×10^{-5}
ψ_{sat}	m	-0.95	-0.02	-0.7	0.05	-0.7	0.7
b	–	1.4	15.5	12.0	0.8	12.0	8.0

were taken into account in the assimilation processes. All constrained relationships considered in this study were soft equality constraints that exist physically for which the model can run through despite these constraints violations. Although these constraints differed from hard constraints requiring satisfaction in model integration (e.g., inequality constraints in Wang et al., 2009), it might retain benefits for consideration in the multi-parameter estimation processes under ensemble assimilation framework of the EnKF. Given that some studies (e.g., Rawls et al., 1982; Zhuang et al., 2001) did not have all relationships between these three hydraulic parameters and others (e.g., Schaap and Leij, 2000; van Genuchten, 1980) did not have enough soil classifications as that in the AVIM model, the literature of Cosby et al. (1984) was selected in this study for its unified soil samples to obtain statistical relationships between parameters k_{sat}, ψ_{sat} and b. Table 5 in Cosby et al. (1984) can be formulated to explicitly display equality constraints between these three parameters:

$$\Delta k_{sat} = (\beta_1^{\Delta b} - 1)k_{sat} \tag{14}$$

$$\Delta \psi_{sat} = (\beta_2^{\Delta b} - 1)\psi_{sat} \tag{15}$$

where Δk_{sat}, $\Delta \psi_{sat}$ and Δb are the perturbations of parameters k_{sat}, ψ_{sat} and b respectively, and β_1, β_2 are constrained coefficients assigned as 1.2474 and 0.827, respectively according to statistical relationships obtained from Cosby et al. (1984). These two equations are included in the error covariance matrix in Eq. (13) to constrain random perturbations of parameters in constrained estimation experiments.
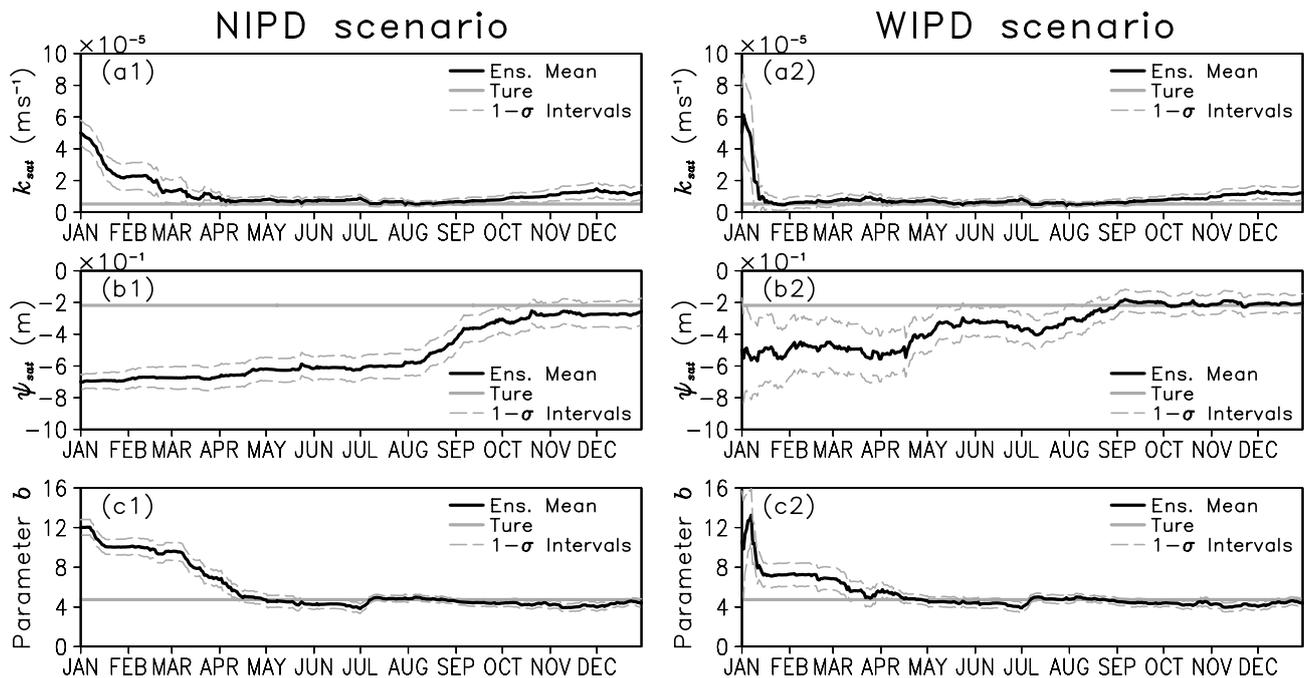
Fig. 2. Time evolution of the ensemble mean parameter values (solid black line) vs. the true parameter values (solid gray line) from single-parameter estimation experiments in NIPD and WIPD scenarios. Estimated parameters are **(a1)**, **(a2)** the saturated hydraulic conductivity; **(b1)**, **(b2)** the saturated soil moisture suction; and **(c1)**, **(c2)** a soil texture empirical parameter b. The area between two dashed gray lines represents the 1-standard deviation (1-σ) intervals of the parameter spread.

4 Results

4.1 Single-parameter estimation results

Results from the individual estimation of these three hydraulic parameters in both NIPD and WIPD scenarios are presented in Figs. 2 and 3. In each experiment, only one such parameter was perturbed around its imperfect mean value, and other parameters were kept unperturbed at their true values.

Figure 2 illustrates that the one year evolution (once-daily analyses) of the ensemble mean parameter values along with the true parameter values remain constant in time. The area between two dash gray lines around the estimated mean parameter value represents the 1-σ (one standard deviation) limits of the parameter ensemble spread. These standard deviation limits were computed by averaging the standard deviations of each 100-member ensembles at each forecast step. Successful parameter estimation should indicate that the error of the estimated parameter is smaller than or very close to the 1-σ limit. The estimated mean parameter values of all three parameters for both NIPD and WIPD scenarios converge to their true values within several months, and the true values remain stable within the 1-σ limit subsequently. When these model parameters were included in the augmented state vectors of the EnKF, the perturbation of parameter led to the update of soil moisture state to a certain

extent. We assumed the general form of the relationship between the perturbations of these parameters and the update of soil moisture as follows:

$$\Delta sm \sim f_1(\Delta k_{sat}) + f_2(\Delta \psi_{sat}) + f_3(\Delta b) \tag{16}$$

where Δk_{sat}, $\Delta \psi_{sat}$ and Δb are the perturbations of parameters k_{sat}, ψ_{sat} and b, and Δsm is the update of soil moisture. Nonlinear operators $f_1(\cdot)$, $f_2(\cdot)$ and $f_3(\cdot)$ denote the sampled relationships between Δsm and Δk_{sat}, $\Delta \psi_{sat}$ and Δb, respectively. In each single-parameter estimation experiment, one of these operators in Eq. (16) is considered in the EnKF update process. In assimilation processes, therefore, the available near surface soil moisture observation information can be transferred by the operator to correct the error in corresponding imperfect parameter and estimate it successfully. Further analysis found that the convergence rate of approach to the true values varied among different parameters and different IPD scenarios. We defined "convergence time" as the time taken for a true parameter value to fall first within the 1-σ limit around the estimated mean parameter. The convergence times of parameters k_{sat} and b were much less than that of parameter ψ_{sat} in both NIPD and WIPD scenarios. Because the convergence time can scale the efficiency of the EnKF to estimate each parameter to a certain extent, this result implies that near surface soil moisture observations contain more useful for parameter k_{sat} and b than for parameter ψ_{sat}, leading to errors in parameter k_{sat} and b that are
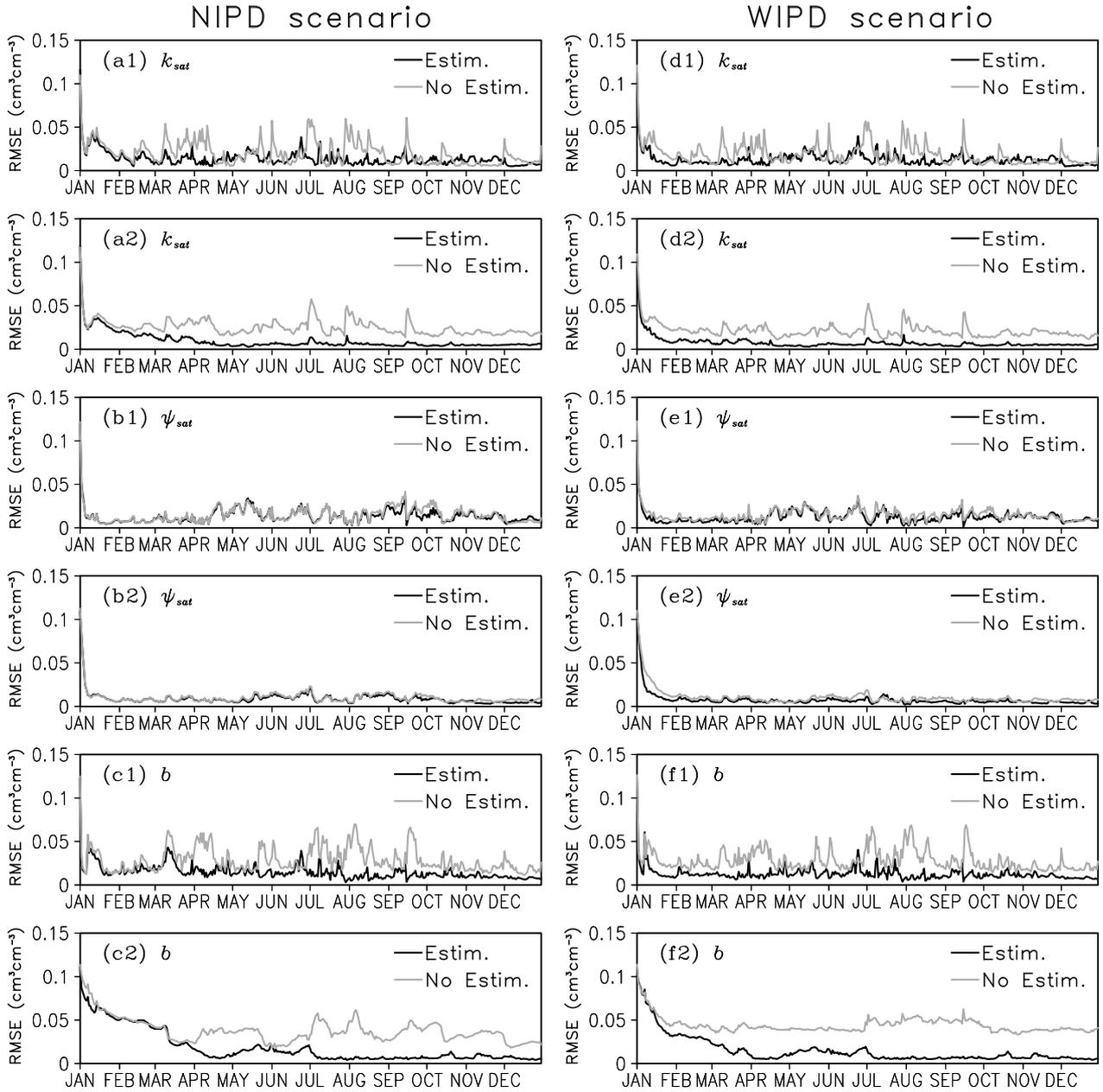
Fig. 3. The time evolution of the root mean squared error (RMSE) of near surface layer soil moisture **(a1–f1)**, and root zone layer soil moisture **(a2–f2)** for soil moisture forecasting from single-parameter estimation experiments (solid black lines) compared with that of non-parameter-estimation benchmark experiments (solid gray lines) in both NIPD and WIPD scenarios. Parameters shown are **(a1)**, **(a2)**, **(d1)**, **(d2)** the saturated hydraulic conductivity; **(b1)**, **(b2)**, **(e1)**, **(e2)** the saturated soil moisture suction; **(c1)**, **(c2)**, **(f1)**, **(f2)** a soil texture empirical parameter b.

easier to correct than those in parameter ψ_{sat} by assimilating the same soil moisture observation in individual parameter estimation experiments. Moreover, for the same estimated parameter, the convergence time in WIPD scenario was less that in NIPD scenario, indicating that a wider prior distribution of imperfect parameters can make estimated parameter converge to its true value easier under the same conditions.

In addition to the mean parameter values, the root mean squared error (RMSE) of soil moisture in each forecast step compared to soil moisture "true state" was calculated to describe errors in soil moisture forecasting:

$$\text{RMSE}_t = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(sm_t^{i-} - sm_t^{\text{true}})^2} \qquad (17)$$
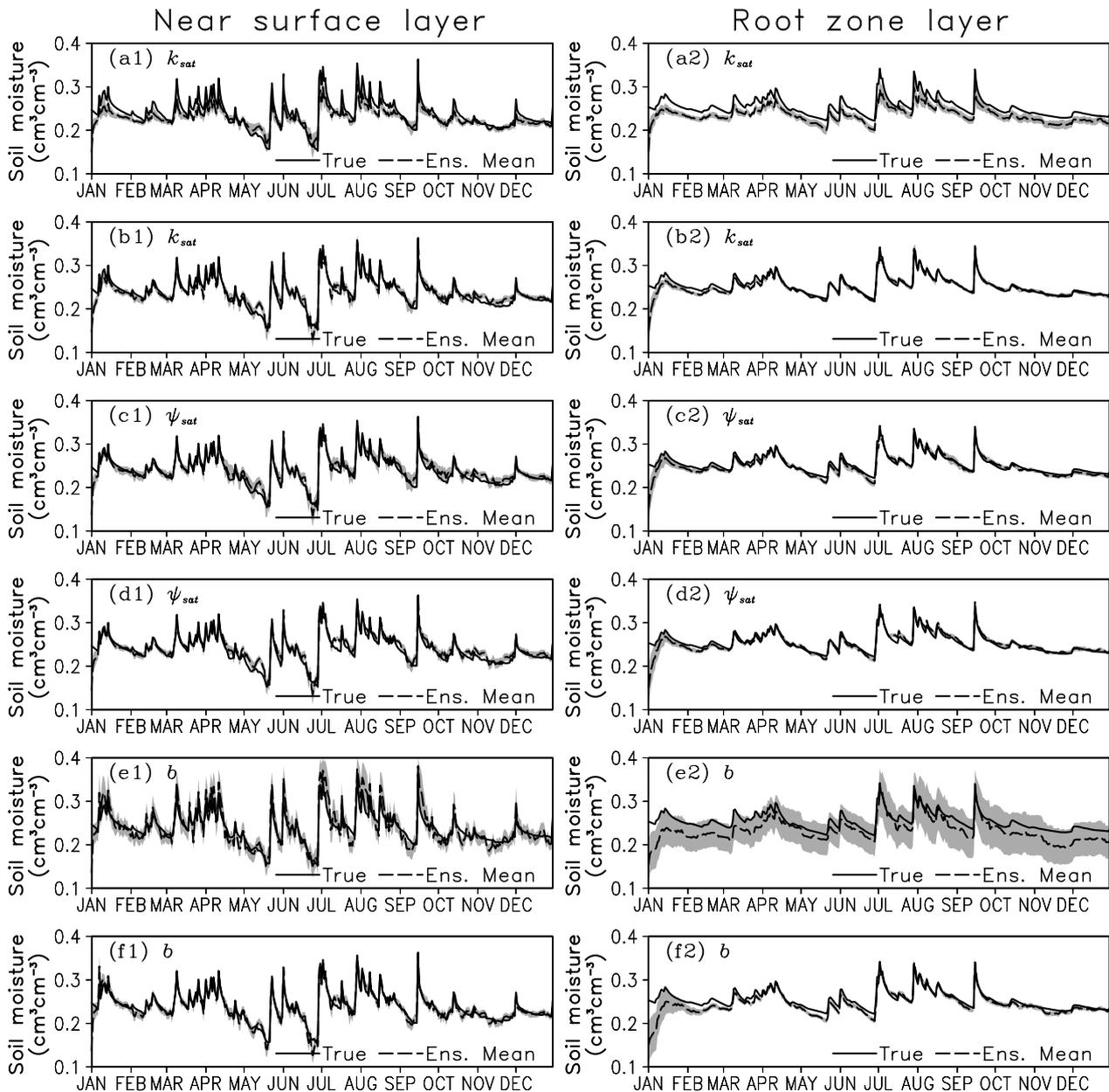
Fig. 4. The time evolution of ensemble soil moisture forecasting of near surface layer **(a1–f1)**, and root zone layer **(a2–f2)** in WIPD scenario. **(a1)**, **(a2)**, **(c1)**, **(c2)**, **(e1)**, **(e2)** are for non-parameter-estimation benchmark experiments and **(b1)**, **(b2)**, **(d1)**, **(d2)**, **(f1)**, **(f2)** are for single-parameter estimation experiments. The estimated parameters shown are **(a1)**, **(a2)**, **(b1)**, (b2) the saturated hydraulic conductivity; **(c1)**, **(c2)**, **(d1)**, **(d2)** the saturated soil moisture suction; **(e1)**, **(e2)**, **(f1)**, **(f2)** a soil texture empirical parameter b. The solid line represents the "true" values, the long dash line represents the ensemble mean values, and the shaded area represents the ensemble spread of soil moisture forecasting.

where the subscript "t" represents time step (every day in the experiment period), n is the ensemble member, sm_t^{i-} is the soil moisture forecast states of the t time step, based on posterior soil moisture states and parameters of the last assimilation time step with perturbed precipitation and radiations. sm_t^{true} is the "true state" of soil moisture for the t time step. The evolution of RMSE for soil moisture forecasting in top

two soil layers for both NIPD and WIPD scenarios are displayed in Fig. 3. In all panels, the RMSE from respective estimation experiments are plotted along with the RMSE from corresponding non-parameter-estimation benchmark experiments. The results of benchmark experiments were obtained by considering imperfect parameters but no parameter estimation in assimilation processes. In both NIPD and WIPD

Table 3. Summary of time average root mean squared error (RMSE) and relative root mean squared error (RRE) of soil moisture in near surface layer (SM1) and root zone layer (SM2) in three single-parameter estimation experiments and corresponding non-parameter-estimation experiments in both NIPD and WIPD scenarios with individual imperfect parameters of the saturated hydraulic conductivity k_{sat}, the saturated soil moisture suction ψ_{sat}, and a soil texture empirical parameter b, respectively.

Scenarios	Imperfect parameter	Estimation RMSE (cm^3 cm^{-3})		Non-estimation RMSE (cm^3 cm^{-3})		RRE	
		SM1	SM2	SM1	SM2	SM1	SM2
NIPD	k_{sat}	0.014	0.009	0.020	0.025	30.0 %	64.0 %
	ψ_{sat}	0.014	0.009	0.015	0.010	6.7 %	10.0 %
	b	0.015	0.018	0.022	0.037	31.8 %	51.4 %
WIPD	k_{sat}	0.013	0.007	0.019	0.022	31.6 %	68.2 %
	ψ_{sat}	0.012	0.007	0.015	0.010	20.0 %	30.0 %
	b	0.013	0.015	0.028	0.048	53.6 %	68.8 %

scenarios, the RMSE of both two soil layers in parameter estimation experiments were lower than that in non-parameter-estimation benchmark experiments for all three parameters. In these identical twin experiments, parameter error was considered one of the main error sources of soil moisture simulation. With parameter errors reduced by estimation processes in the EnKF, fewer error contributions from imperfect parameters can refine the performance of soil moisture forecasting in each estimation experiment.

Figure 3 also depicts that the behavior of RMSE evaluation of soil moisture forecasting in parameter ψ_{sat} estimation case is less identifiable than that of parameter k_{sat} and b in both scenarios. Taking WIPD scenario as an example, Fig. 4 compares the time evolution of soil moisture forecasting and its ensemble uncertainty in these three single parameter estimation experiments. Because the root zone layer is less affected by atmosphere forcing and soil moisture observation, the uncertainty of soil moisture forecasting in this layer can reflect the impact of parameter errors to a large extent. The uncertainty of soil moisture forecasting in near surface layer primarily reflects the impact of errors of atmosphere forcing. Errors associated with imperfect parameter k_{sat} and b are reflected in soil moisture forecasting more significantly than those associated with parameter ψ_{sat} (Fig. 4a2, e2), which leads to large biases in soil moisture ensemble forecasting values compared to their true values. This result implies that errors in these two parameters are comparable to those in soil moisture initial condition and forcing data. For parameter ψ_{sat}, because the uncertainty of soil moisture forecasting caused by parameter error (Fig. 4c2) was much smaller than that by errors in initial condition (the first two months in Fig. 4c2) and forcing data (Fig. 4c1), the RMSE evaluation of soil moisture forecasting was less correlated with the error in parameter ψ_{sat}. Therefore, even when parameter ψ_{sat} did not converge to its true value (before April in Fig. 2b2), the RMSE evolution of soil moisture forecasting was reduced

reasonably (Fig. 4e1, e2) by assimilating near surface soil moisture observation using the EnKF.

Further analysis revealed that the decrease of RMSE from non-parameter-estimation experiments to estimation experiments varied among parameters. To quantify relative estimation performance, we defined the "relative root mean squared error", as follows:

$$\mathrm{RRE} = \frac{\langle \mathrm{RMSE_{No\text{-}Estimation}} \rangle - \langle \mathrm{RMSE_{Estimation}} \rangle}{\langle \mathrm{RMSE_{No\text{-}Estimation}} \rangle} \times 100\% \quad (18)$$

where the operation $\langle \cdot \rangle$ denotes time average over the entire experiment period. The RRE was a relative measure of how much error has been reduced by parameter estimation compared to non-parameter-estimation benchmark experiments. The results of time average RMSE and RRE for single-parameter estimation experiments in both NIPD and WIPD scenarios are summarized in Table 3. The estimation experiments in the WIPD scenario had smaller RMSE and larger RRE than those in the NIPD scenario. One possible reason might be that estimation experiments in the WIPD scenario all have less convergence time than that in the NIPD scenario, causing more parameter error reduced by the EnKF estimation processes. The smallest RRE was exhibited by parameter ψ_{sat} in both NIPD scenario (6.7 %, 10 %) and WIPD scenario (20 %, 30 %). The RRE of parameter k_{sat} and b were much larger than that of parameter ψ_{sat}, which were both over 30 % for the near surface layer and even more than 60 % for the root zone layer. This result implies that the soil moisture forecasting is more sensitive to the error in parameter k_{sat} and b than that in parameter ψ_{sat}, consistent with the sensitivity analysis of Wen et al. (1998). The RRE for the root zone layer, in which no soil moisture observations were assimilated in the EnKF, was larger than that for the near surface layer for all three parameters and IPD scenarios. As only near surface soil moisture observations were used in these experiments, observation error in near surface
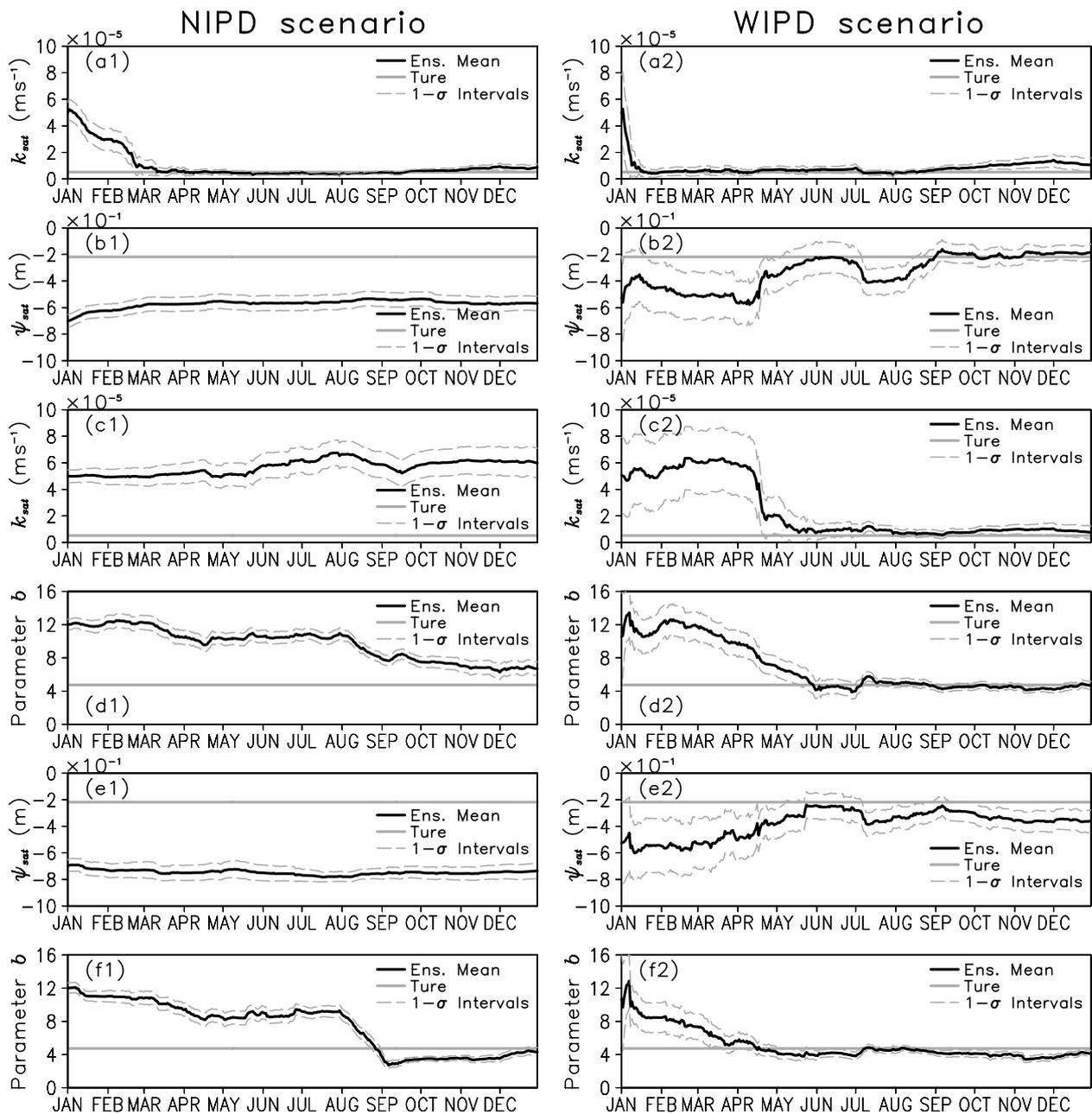
Fig. 5. Same as in Fig. 2 but for three sets of dual-parameter estimation experiments, which are **(a1)**, **(a2)**, **(b1)**, **(b2)** for k_{sat} and ψ_{sat} set; **(c1)**, **(c2)**, **(d1)**, **(d2)** for k_{sat} and b set; and **(e1)**, **(e2)**, **(f1)**, **(f2)** for ψ_{sat} and b set, in both NIPD and WIPD scenarios with mutually independent parameter perturbations.

soil layer had less effect on soil moisture forecasting in the root zone layer. Therefore, with the error in the imperfect parameter reduced by the EnKF, soil moisture forecasting in no observation soil layer can be improved more significantly in any corresponding parameter estimation processes. Combining the previous analyses of ensemble mean parameter values and RMSE, the EnKF-based single-parameter estimation performed successfully in soil moisture assimilation.

4.2 Multi-parameter estimation results

To obtain a comprehensive picture of the EnKF's capability and limits in parameter estimation when multiple imperfect parameters are involved, the results from three sets of dual-parameter estimation experiments and subsequently the three-parameter estimation experiments are presented here.

The dual-parameter experiments were performed with ψ_{sat} and b, k_{sat} and ψ_{sat}, k_{sat} and b as the three sets of

Table 4. Summary of the time average root mean squared error (RMSE) and relative root mean squared error (RRE) of soil moisture in near surface layer (SM1) and root zone layer (SM2) in three sets of dual-parameter estimation (DPE) experiments (with ψ_{sat} and b, k_{sat} and ψ_{sat}, and k_{sat} and b as imperfect parameters respectively) and corresponding non-parameter-estimation (NPE) experiments in both NIPD and WIPD scenarios.

Scenarios	Imperfect parameter set	DPE RMSE (cm^3 cm^{-3})		NPE RMSE (cm^3 cm^{-3})		RRE	
		SM1	SM2	SM1	SM2	SM1	SM2
NIPD	ψ_{sat}, b	0.018	0.024	0.025	0.032	28.0 %	25.0 %
	k_{sat}, ψ_{sat}	0.018	0.015	0.024	0.034	25.0 %	55.9 %
	k_{sat}, b	0.022	0.016	0.023	0.020	4.3 %	20.0 %
	Set average	0.019	0.018	0.024	0.029	19.1 %	33.6 %
WIPD	ψ_{sat}, b	0.012	0.012	0.024	0.034	50.0 %	64.7 %
	k_{sat}, ψ_{sat}	0.013	0.01	0.022	0.026	40.9 %	61.5 %
	k_{sat}, b	0.014	0.012	0.034	0.054	58.8 %	77.8 %
	Set average	0.013	0.011	0.027	0.038	49.9 %	68.0 %

imperfect parameters, respectively. Figure 5 shows the evolution of ensemble mean parameter values from these dual-parameter experiments for both NIPD and WIPD scenarios. The performances of simultaneous dual parameter estimation have some deterioration compared to that of individual estimations. In the NIPD scenario, only one parameter was estimated to the true value in sets ψ_{sat}, b (Fig. 5a1) and k_{sat}, ψ_{sat} (Fig. 5f1). For the k_{sat} and b set, both imperfect parameters failed estimation, although some convergence of the estimated value of parameter b to its true value was still noticeable. The estimation performance in the WIPD scenario was better than that in the NIPD scenario. Except for parameter ψ_{sat} in the ψ_{sat}, b set (Fig. 5e2), all other imperfect parameters in these three dual-parameter sets were estimated successfully. Unlike single-parameter estimation case, sampled relationships of two parameters in Eq. (16) need to be considered simultaneously in the EnKF for an increase of imperfect parameters here. Therefore, the degree of freedom in assimilation processes increased and made the dual-parameter estimation unstable and intractable. The comparative results between NIPD and WIPD scenarios imply that wide enough initial perturbation on estimated parameters has positive influences on the performance of estimation. In other words, more ensemble sampling information content in the wide prior distribution of estimated parameters is used to identify the parameters properly.

Table 4 illustrates the time average RMSE and RRE of these three dual-parameter estimation sets. Similar to the single parameter estimation experiments, the RMSE of all three dual-parameter estimation experiment sets in both NIPD and WIPD scenarios were lower than of non-parameter-estimation benchmark experiments, despite not all incorrect parameters converging to their true values in some cases (Fig. 5). For more error associated with incorrect parameters being reduced by estimation processes, the RRE in WIPD

scenario (49.9 % and 68 % for two layers on set average) was also much larger than that in NIPD scenario (19.1 % and 33.6 % on set average). In general, although some estimated parameters had worse estimations compared to single-parameter estimation cases, the estimation with two imperfect parameters using the EnKF was still beneficial to reduce the parameter-related model error in soil moisture forecasting, especially when the prior distribution of estimated parameters was wide enough to cover true values.

When the uncertainty extended to the entire set of three parameters, some significant deterioration of estimation performance is occurred (Fig. 6). The estimated ensemble mean values of all three parameters did not converge to their true values throughout the entire experiment period in the NIPD scenario. When the prior distribution of parameter was narrow, small parameter uncertainty reflected by ensemble sampling was not enough for the EnKF to build reasonable ensemble statistical relationships between the ensembles of soil moisture observation and all estimated parameters simultaneously. In the WIPD scenario, as a wider prior distribution of estimated parameters led to more appropriate sampling information included in parameter ensembles, better estimation performances with k_{sat} and b converging to their respective true values successfully were obtained compared to that in the NIPD scenario. However, independent random perturbations added to different estimated parameters during the update process of assimilation may still break the complicated relationships between hydraulic parameters and soil moisture observations easily and may cause the failure in simultaneous estimation of all incorrect parameters. For dealing with all imperfect parameters properly, constraints in Eqs. (14) and (15) (obtained from Cosby et al., 1984) were taken into account for further improving estimation performance.

Next, we discuss the three-parameter estimation case in the NIPD and WIPD scenarios with constrained parameter
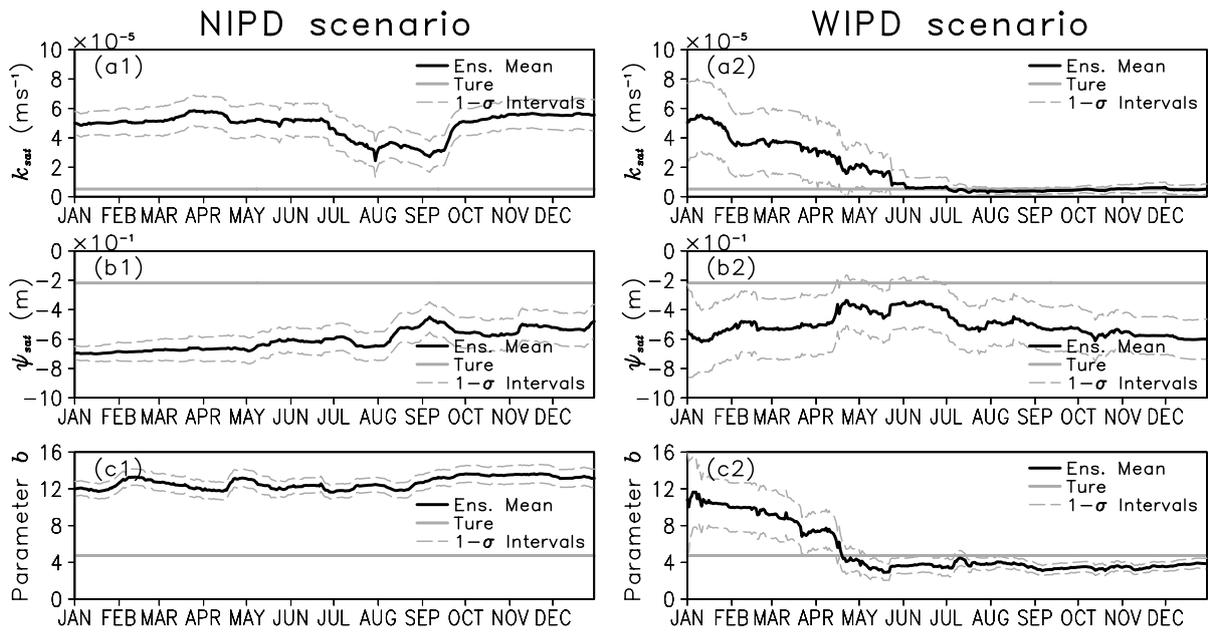
Fig. 6. Same as in Fig. 2 but for the three-parameter estimation experiments in both NIPD and WIPD scenarios with mutually independent parameter perturbations.
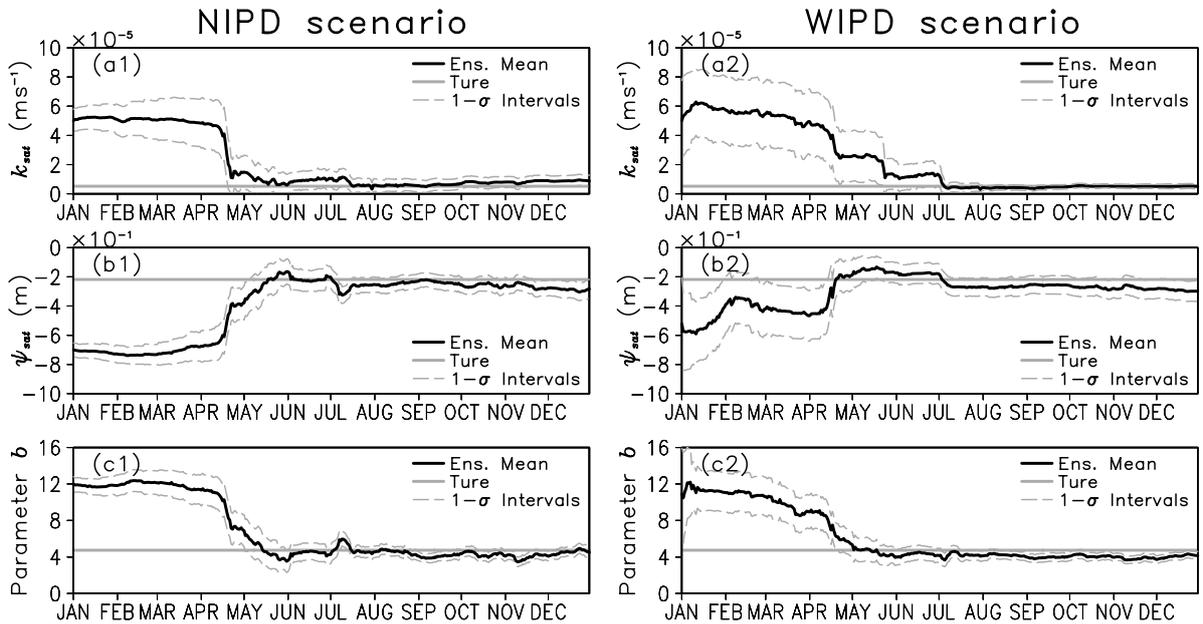
Fig. 7. Same as in Fig. 2 but for the three-parameter estimation experiments in both NIPD and WIPD scenarios with constrained parameter perturbations.

perturbation. The evolutions of the ensemble mean parameter values are shown in Fig. 7. The estimated ensemble mean values for all three parameters converge to their true values successfully in both NIPD and WIPD scenarios with almost the same convergence time of five months. From the viewpoint of adding information, imposing constraints between model parameters also adds new information to the EnKF

assimilation system in addition to the available soil moisture observation. Moreover, constraints also transfer observation information to all estimated parameters in a balanced way and correct these parameters in a coordinated and consistent way during the whole update process of assimilation. Therefore, even if more imperfect parameters require estimation, these equality constraints that represent additional

correlation between parameters effectively offset the decline in the estimation power of the EnKF and make all parameters converge to their true values effectively.

The RMSE evolution of constrained estimation, non-constrained estimation and non-parameter-estimation experiments are displayed in Fig. 8. The smallest RMSE for two soil layers are exhibited by constrained estimation experiments in both NIPD and WIPD scenarios. In the NIPD scenario, the RMSE of soil moisture forecasting were almost the same for those three experiments before May. The reason for coincident RMSE was attributed to the key role of soil moisture observation in updating of soil moisture states in this scenario. The evaluation of posterior soil moisture ensembles (Fig. 9) displayed that observation information was enough for the EnKF reducing most of error in soil moisture forecasting in this assimilation period, even if all estimated parameters had imperfect values (Fig. 9a1, a2). Whereas after the convergence time of all parameters in the constrained estimation experiment (Fig. 7a1, b1, c1), because all three parameters failed to be estimated in non-constrained estimation experiment in this period (after June in Fig. 6a1, b1, c1), the cumulative effects of parameter error caused positive biases in the ensemble mean of soil moisture forecasting in the drying down periods (Fig. 9a1, a2). The constrained estimation case had more benefits in RMSE evolution compared to its non-constrained counterpart (Fig. 8a1, a2). In the WIPD scenario, because more parameter prior uncertainties were introduced and kept in the whole EnKF assimilation processes, much large uncertainties were contained in the posterior ensembles of soil moisture forecasting in the non-parameter-estimation experiment (Fig. 9d1, d2). The RMSE in non-estimation experiments (Fig. 8b1, b2) were much larger than that of the NIPD scenario (Fig. 8a1, a2). With larger prior parameter error in this scenario, the soil moisture forecasting errors (combined errors of initial condition, parameter and atmosphere forcing) were much larger than the error in the near surface soil moisture observations at the early stage of the EnKF assimilation processes (Fig. 9e1, e2, f1, f2). In this case, although all three parameters did not converge to their true values in both constrained and non-constrained estimation experiments before April (Figs. 6 and 7), the RMSE of soil moisture forecasting in these two experiments can also converge to an equivalent low level as observation error within several assimilation cycles (Fig. 8b1, b2). After June, as all three parameters in constrained estimation experiment (Fig. 7a2, b2, c2) and two key parameter k_{sat} and b in non-constrained estimation experiment (Fig. 6a2, c2) was successfully estimated by the EnKF, the performance of constrained estimation experiment actually benefited little from these additional constraints between parameters in the last half year in the WIPD scenario. Figures 8b1, b2 indicate that the RMSE of constrained estimation experiment was slightly smaller compared to corresponding non-constrained estimation experiment, although it was still significantly smaller than that of non-parameter-estimation experiment.

4.3 Sparse observation assimilation results

Actual conventional in situ soil moisture observations in China and other areas in the world were sparse in time with a measurement interval of about 10 days (Robock et al., 2000; Nie et al., 2008). For the future application of the constrained multi-parameter estimation method in assimilating actual in situ soil moisture, its performance with temporally sparse observation conditions were tested and are discussed in this subsection.

Results from experiments with 10-day assimilation interval are shown in Figs. 10 and 11. Figure 10 illustrates that the performances of ensemble mean values for all parameters were similar to that of one-day assimilation interval experiments, especially in the NIPD scenario, even if much less soil moisture observation information was assimilated in this case. In constrained estimation experiments, all three parameters successfully converged to true values in both NIPD and WIPD scenarios. However, the situation was not satisfied for the non-constrained experiments, in which the estimations for most of estimated parameters failed. Less available soil moisture observation information assimilated in the EnKF weakened the role of the EnKF on reducing uncertainties in wide prior parameter distributions. Therefore, the 1-σ intervals in the WIPD scenario here (e.g., Fig. 10a2, d2) were larger than those of corresponding one-day assimilation interval experiments. The RMSE evolution in Fig. 11 depicts that the smallest RMSE were exhibited by constrained estimation experiments in both NIPD and WIPD scenarios as that in one-day assimilation interval case. With much less soil moisture observation information used to reduce model errors and improve the performance of the EnKF, the RMSE of soil moisture forecasting was always larger than that in one-day assimilation interval case.

Further analyses with 20-, 30- and 40-day assimilation intervals illustrated similar results as that of the 10-day intervals. Table 5 provides a summary of RMSE and RRE for simultaneous three-parameter estimation experiments with different assimilation intervals. In general, the constrained estimations in both NIPD and WIPD scenarios produce consistently improved performances relative to non-constrained-estimation experiments and non-parameter-estimation benchmark experiments. For all assimilation intervals, the RMSE of the constrained estimation experiments in both two soil layers were significantly smaller than that of the non-parameter-estimation experiments. Meanwhile, the RRE of the constrained estimation experiments were always larger than that of the non-constrained-estimation experiments for more error reduced by proper parameter estimation. However, the increases of RRE from non-constrained estimation experiment to constrained estimation experiment were different between the NIPD and WIPD scenarios. In the NIPD scenario, because all parameters of non-constrained estimation experiments lacked estimates, the constrained estimation method fully demonstrated its benefits of estimating
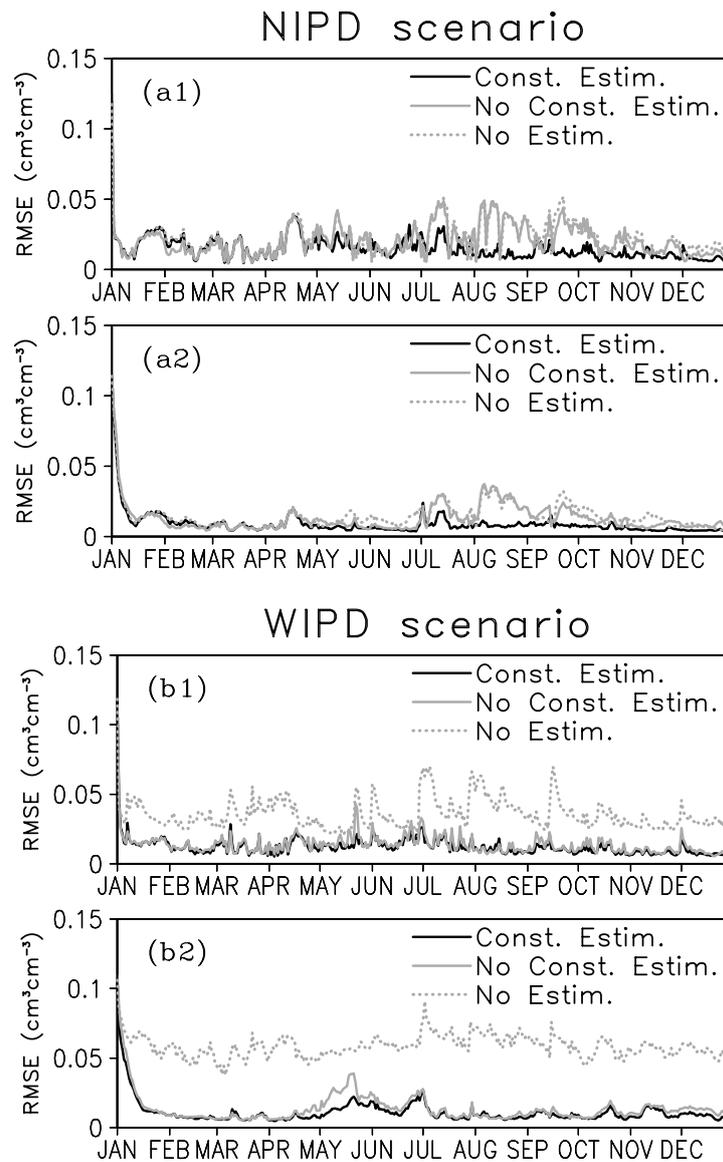
Fig. 8. Same as in Fig. 3 but for the time evolution of RMSE of near surface layer soil moisture **(a1)**, **(b1)** and root zone layer soil moisture **(a2)**, **(b2)** from the constrained (solid black lines) and non-constrained (solid gray lines) three-parameter estimation experiments in both NIPD and WIPD scenarios corresponding to Figs. 5 and 6, respectively. Short dashed gray lines represent RMSE from non-parameter-estimation benchmark experiment.

all imperfect parameters simultaneously, causing significant increases of RRE from the non-constrained estimation experiment (less than 10 %) to the constrained estimation experiment (more than 30 %). In the WIPD scenario, as the non-constrained estimation experiments already have good estimation performance (e.g., Fig. 10e2, f2), the advantage of the constrained estimation method compared to unconstrained method in the RMSE of soil moisture forecasting was not as significant as that in the NIPD scenario. In this scenario, the RRE of constrained estimation experiments were slightly larger than that of non-constrained estimation experiments for all assimilation intervals.

Table 5 also shows that the changes of RRE with assimilation intervals in constrained estimation experiments were different between the NIPD and WIPD scenarios. In the NIPD scenario, the largest RRE was exhibited by the 10- and 20-day assimilation intervals for near surface layer (39.1 %) and root zone layer (41.9 %), respectively. When assimilation interval was as small as one day, frequent corrections from soil moisture observations to model state may have weakened the effects of proper parameter estimation in soil moisture forecasting. Whereas in the extremely sparse observation cases ($\geq$30 days here), too little information was available to update model state from soil moisture observations over a
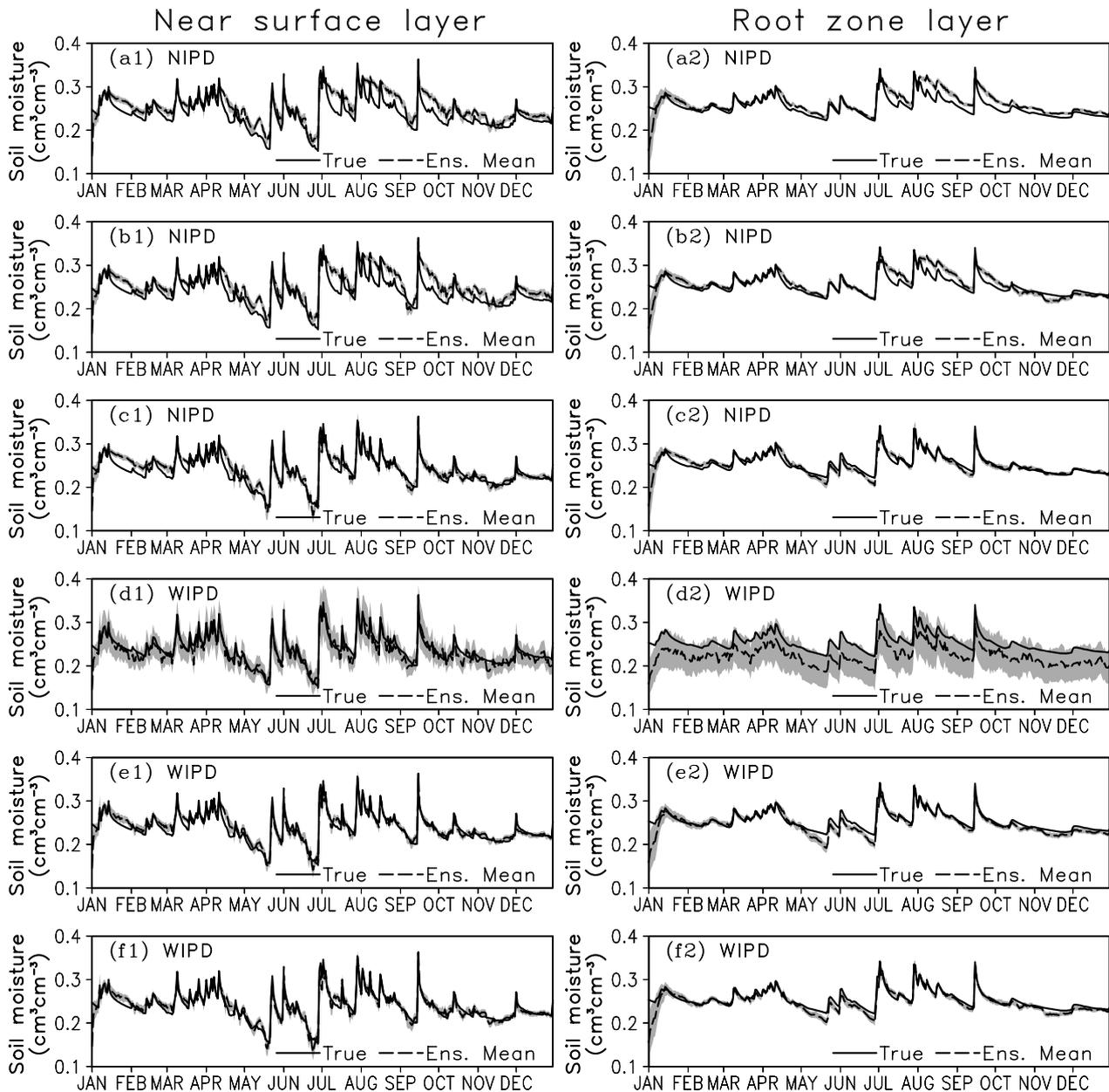
Fig. 9. The time evolution of ensemble soil moisture forecasting of near surface layer **(a1–f1)**, and root zone layer **(a2–f2)** in both NIPD **(a–c)** and WIPD **(d–f)** scenarios. **(a1)**, **(a2)**, **(d1)**, **(d2)** are for non-parameter-estimation benchmark experiments, **(b1)**, **(b2)**, **(e1)**, **(e2)** are for non-constrained three parameter estimation experiments and **(c1)**, **(c2)**, **(f1)**, **(f2)** are for constrained three parameter estimation experiments. The solid line represents the "true" values, the long dash line represents the ensemble mean values, and the shaded area represents the ensemble spread of soil moisture forecasting.

certain period (e.g., one month), making it difficult to overcome the accumulation of errors in soil moisture state in the process of model integration. Therefore, when the parameter uncertainties reflected by ensemble were insufficient, proper assimilation frequency (about once 10 or 20 days here) displayed the greatest advantages of this constrained estimation method in soil moisture assimilation using the EnKF. Whereas in WIPD scenario, the 1-day assimilation interval

experiment displayed the largest RRE for both near surface layer (66.7 %) and root zone layer (81.3 %). When parameter uncertainties were large enough with a wide prior ensemble sampling distribution, the estimation procedures extracted more useful information from soil moisture observations. In this case, therefore, frequent corrections to model soil moisture state obtained the biggest advantage of the constrained estimation method.
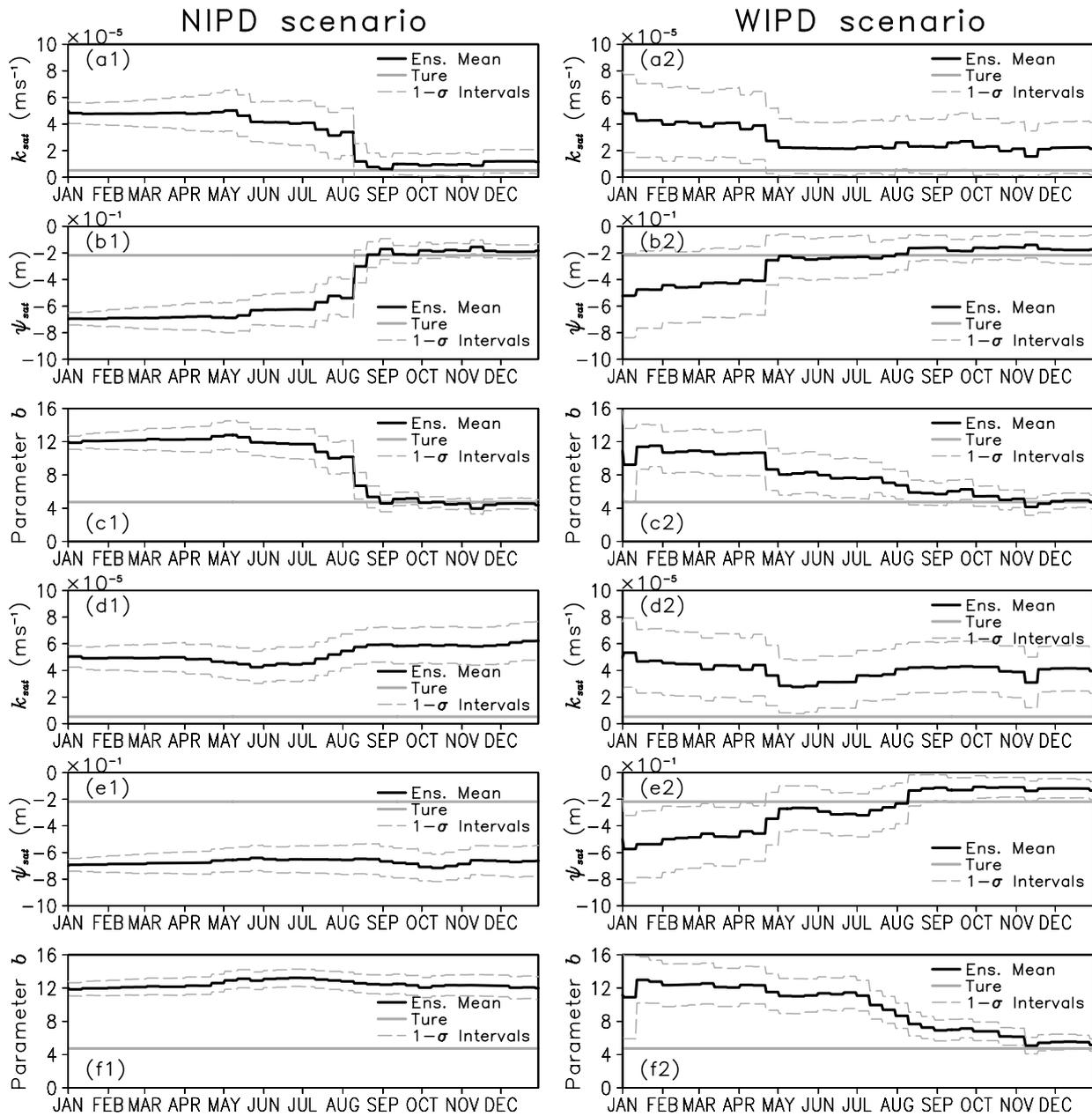
Fig. 10. Same as in Fig. 2 but for the constrained three-parameter estimation experiments **(a1)**, **(a2)**, **(b1)**, **(b2)**, **(c1)**, **(c2)** and non-constrained three-parameter estimation experiments **(d1)**, **(d2)**, **(e1)**, **(e2)**, **(f1)**, **(f2)** in both NIPD and WIPD scenarios with 10-day assimilation interval.

5 Conclusions

This study explores the applicability of the EnKF-based state-parameter estimation in soil moisture data assimilation using a physical process land surface model by a series of identical twin experiments. Uncertainties in model parameters, soil moisture initial condition and atmospheric forcing data were considered primary sources of model errors. By the means of state augmentation, model-based pseudo near surface soil moisture observations estimated model param-

eter and soil moisture state simultaneously. Three key hydraulic parameters: the saturated hydraulic conductivity k_{sat}, the saturated soil moisture suction ψ_{sat} and a soil texture empirical parameter b, were subjected to estimation attempts in various experiments. Because an accurate estimate of prior uncertainty of hydraulic parameters in nature differed greatly, two kinds of initial parameter distribution (IPD) scenarios were considered in this study to explore fully the parameter estimation capability of the EnKF. These included the narrow IPD (NIPD) scenario representing our insufficient

NIPD scenario

WIPD scenario
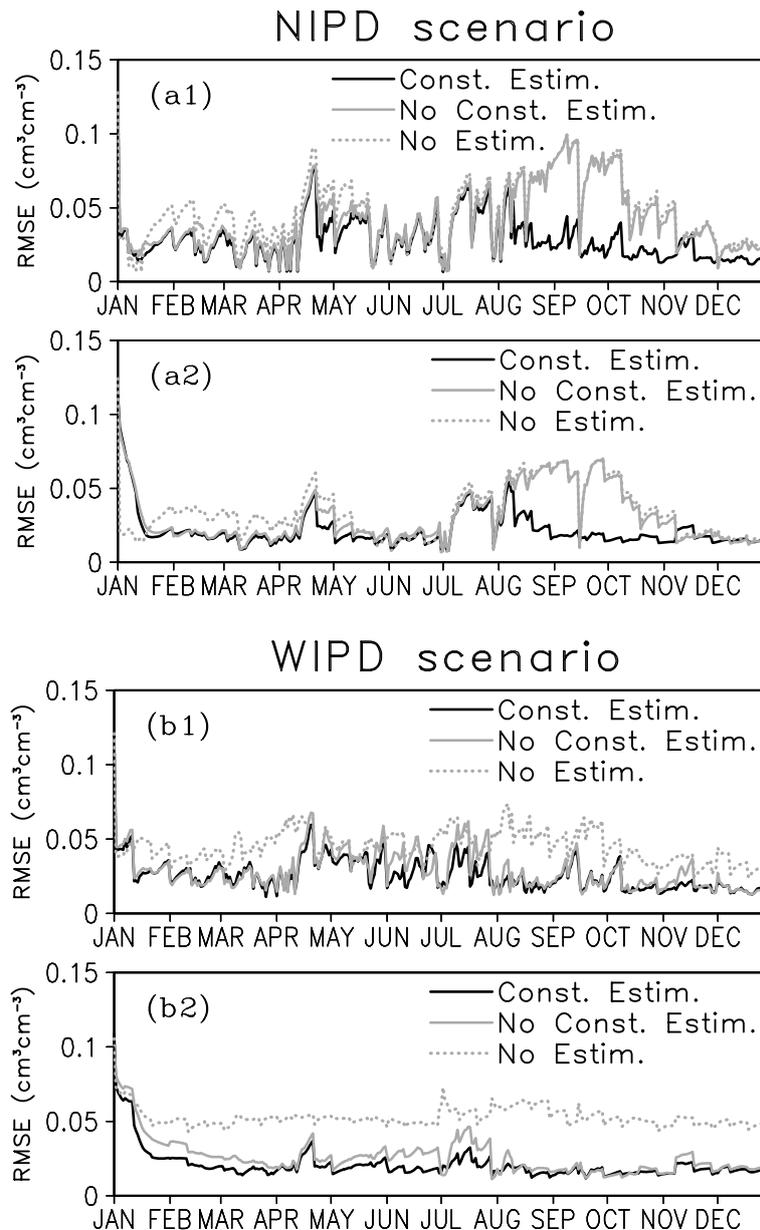
Fig. 11. Same as in Fig. 3 but for the time evolution of RMSE of near surface layer soil moisture **(a1)**, **(b1)** and root zone layer soil moisture **(a2)**, **(b2)** from the constrained (solid black lines) and non-constrained (solid gray lines) three-parameter estimation experiments in both NIPD and WIPD scenarios with 10-day assimilation interval corresponding to Fig. 8. Short dashed gray lines represent RMSE from the non-parameter-estimation benchmark experiment.

knowledge about the first guess error of these hydraulic parameters and the wide IPD (WIPD) scenario indicating we know enough about the first guess error of all parameters.

The estimation of single imperfect parameter was successful for all three estimated parameters in both NIPD and WIPD scenarios. The ensemble mean value of each estimated parameters converge to its true value successfully. Increasing the number of imperfect parameters led to a decline in the performance of parameter estimation, especially

for the NIPD scenario. A wide initial distribution of estimated parameters made more appropriate sampling information included in parameter ensembles, improving simultaneous multi-parameter estimation performances compared to that of the NIPD scenario. However, when the number of estimated parameters increased to three, experiments in both NIPD and WIPD scenarios could not lead all estimated parameters to converge to their true values validly, although the overall performance of the EnKF in terms of soil moisture

Table 5. Summary of the time average root mean squared error (RMSE) and relative root mean squared error (RRE) of soil moisture in near surface layer (SM1) and root zone layer (SM2) in three-parameter constrained estimation (CE) experiments, non-constrained estimation (NCE) experiments, and non-parameter-estimation (NPE) experiments in both NIPD and WIPD scenarios with assimilation intervals of 1 day, 10 days, 20 days, 30 days and 40 days, respectively.

Scenarios	Assimilation intervals	CE RMSE ($cm^3 cm^{-3}$)		NCE RMSE ($cm^3 cm^{-3}$)		NPE RMSE ($cm^3 cm^{-3}$)		CE RRE (%)		NCE RRE (%)	
		SM1	SM2	SM1	SM2	SM1	SM2	SM1	SM2	SM1	SM2
NIPD	1 day	0.015	0.009	0.021	0.013	0.022	0.014	31.8	35.7	4.5	7.1
	10 days	0.028	0.020	0.042	0.031	0.046	0.034	39.1	41.2	8.7	8.8
	20 days	0.034	0.025	0.053	0.041	0.054	0.043	37.0	41.9	1.9	4.7
	30 days	0.039	0.031	0.056	0.045	0.059	0.047	33.9	34.0	5.1	4.3
	40 days	0.043	0.035	0.060	0.049	0.063	0.052	31.7	32.7	4.8	5.8
WIPD	1 day	0.012	0.011	0.014	0.014	0.036	0.059	66.7	81.3	61.1	76.3
	10 days	0.026	0.021	0.029	0.026	0.045	0.053	42.2	60.4	35.6	50.9
	20 days	0.032	0.027	0.035	0.029	0.055	0.059	41.8	54.2	36.4	50.8
	30 days	0.035	0.029	0.036	0.031	0.057	0.058	38.6	50.0	36.9	46.6
	40 days	0.036	0.031	0.037	0.031	0.059	0.059	39.0	47.5	38.3	47.5

RMSE was still superior to the non-parameter-estimation benchmark experiments with parameter error reduced in the estimation processes. The failure of estimation was ascribed to independent perturbations on different estimated parameters, causing the increase of degree of freedom of assimilation system and the breakage of the statistical relationships between these parameters in assimilation processes.

A strategy of considering constraints between estimated hydraulic parameters in the filter was introduced to improve the performance of estimating all three imperfect parameters simultaneously. The constraints used were obtained from Cosby et al. (1984), and are the approximation to actual statistical relationships between these parameters. The performance of the constrained three-parameter estimation was successful, even if observations were available with temporally sparse intervals such as 10 days or much longer. From the viewpoint of adding information, imposing the equality constraints between estimated parameters also adds new information to the EnKF assimilation system and the available observations, namely, the statistical balance correlation between these parameters. For this reason, the constrained estimation method effectively overcame the negative impacts of the non-closure problem in the multi-parameter estimation, especially for the NIPD scenario which features an insufficient guess for the prior error of estimated parameters.

Compared to non-parameter-estimation and non-constrained estimation cases, the constrained three-parameter estimation case can reduce RMSE in soil moisture forecasting much more. The benefits of this constrained estimation method in estimating all imperfect parameters simultaneously are fully demonstrated when the corresponding non-constrained estimation case has a relatively poor performance for parameter estimation as in the NIPD

scenario. Given that true values of these hydraulic parameters (and also most of other parameters in LSM) cannot be obtained in nature, the constrained estimation method may provide an effective way for reducing parameter related uncertainties in actual soil moisture assimilation with its poor parameter settings. Moreover, the greatest advantages of the constrained estimation method in improving soil moisture forecasting were displayed with different assimilation intervals for the NIPD and WIPD scenarios. When the first guess of these hydraulic parameters are insufficient, 10- to 20- days may be a proper assimilation frequency. Once the first guess error of parameters is given correctly in the EnKF assimilation processes, however, frequent corrections using soil moisture observation are able to utilize the greatest advantage of this constrained estimation method. Although obtained from identical twin experiments, these results still provide an instructive analysis of the advantages of this constrained estimation method in soil moisture assimilation.

Because all constraints considered in this study were always in a statistical sense and obtained from literature or standard parameter tables of the model, this constrained parameter estimation method is applicable to other land surface models with even more imperfect parameters for estimation. Due to its superiorities to the non-constrained estimation method and the traditional non-parameter-estimation method in the performance of soil moisture forecasting, even if available observations are rare in time, this method is suitable well for assimilating actual temporally sparse in situ soil moisture observations in China. The application of this constrained multi-parameter estimation method in assimilating temporally sparse in situ soil moisture observations in China will be studied in the future.

Edited by: Y. Liu

References

Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W.: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model, Mon. Weather Rev., 134, 2951–2970, 2006.

Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, Mon. Weather Rev., 129, 2884–2903, 2001.

Annan, J. D., Hargreaves, J. C., Edwards, N. R., and Marsh, R.: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter, Ocean Modell., 8, 135–154, 2005.

Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev., 126, 1719–1724, 1998.

Chen, M., Liu, S., Tieszen, L. L., and Hollinger, D. Y.: An improved state-parameter analysis of ecosystem models using data assimilation, Ecol. Model., 219, 317–326, 2008.

Clapp, R. B. and Hornberger, G.. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, 1978.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils, Water Resour. Res., 20(6), 682–690, 1984.

Cornelis, W., Ronsyn, M. J., van Meirvenne, M., and Hartmann, R.: Evaluation of pedotransfer functions for predicting the soil moisture retention curve, Soil Sci. Soc. Am. J., 65, 638–648, 2001.

Coudert, B., Ottlé, C., Boudevillain, B., Demarty, J., and Guillevic, P.: Contribution of thermal infrared remote sensing data in multiobjective calibration of a dual-source SVAT model, J. Hydrometeorol., 7, 404–420, 2006.

Crow, W. T. and van den Berg, M. J.: An improved approach for estimating observation and model error parameters in soil moisture data assimilation, Water Resour. Res., 46, W12519, doi:10.1029/2010WR009402, 2010.

Crow, W. T. and Van Loon, E.: Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, J. Hydrometeorol., 7, 421–432, 2006.

Dechant C. and Moradkhani H.: Radiance data assimilation for operational snow and streamflow forecasting, Adv. Water Resour., 34, 351–364, 2010.

Dickinson, R., Henderson-Sellers A., and Kennedy P.: Biosphere-Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model, NCAR Tech. Note TN-387+STR, Natl. Cent. for Atmos. Res., Boulder, Colo, 1993.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143–10162, 1994.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dyn., 53, 343–367, 2003.

Franssen H. J. H. and Kinzelbach W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resour. Res., 44, W09408, doi:10.1029/2007WR006505, 2008.

Fujiyoshi, Y., Ding, Y. H., and Zhang, Y.: Outline of GAME/HUBEX, in: Final Report of GAME/HUBEX, Edited by: Fujiyoshi, Y., and Ding, Y. H., GAME/HUBEX Project Office, Beijing, 1–6, 610 pp., 2006.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, Water Resour. Res., 34, 3405–3420, 1998.

Ji, J. J.: A climate-vegetation interaction model: Simulating physical and biological processes at the surface, J. Biogeogr., 22, 445–451, 1995.

Ji, J. J. and Hu, Y.: A simple land surface process model for use in climate studies, Acta. Meteorologic. Siniea, 3, 342–351, 1989.

Jung, Y., Xue, M., and Zhang, G.: Simultaneous estimation of microphysical parameters and the atmospheric state using simulated polarimetric radar data and an ensemble Kalman filter in the presence of an observation operator error, Mon. Weather Rev., 138, 539–562, doi:10.1175/2009MWR2748.1, 2010.

Kalman, R. and Bucy, R.: New results in linear prediction and filtering theory, Trans. AMSE J. Basic Eng., 83D, 95–108, 1961.

Leisenring M. and Moradkhani, H.: Snow water equivalent estimation using Bayesian data assimilation methods, Stoch. Environ. Res. Risk Assess., 25, 253–270, doi:10.1007/s00477-010-0445-5, 2010.

Margulis, S. A., McLaughlin, D., Entekhabi, D., and Dunne, S.: Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, Water Resour. Res., 38, 1299, doi:10.1029/2001WR001114, 2002.

Montaldo, N., Albertson, J. D., and Mancini, M.: Dynamic calibration with an ensemble Kalman filter based data assimilation approach for root-zone moisture predictions, J. Hydrometeorol., 8, 910–921, 2007.

Montzka, C., Moradkhani, H., Weihermuller, L., Canty, M., Hendricks Franssen, H. J., and Vereecken, H.: Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, J. Hydrol., 399, 410–421, 2011.

Moradkhani H., Hsu, K., Gupta, H. V., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41, W05012, doi:10.1029/2004WR003604, 2005a.

Moradkhani, H., Sorooshian S., Gupta, H. V., and Houser, P.: Dual State-Parameter Estimation of Hydrological Models using Ensemble Kalman Filter, Adv. Water Resour., 28, 135–147, 2005b.

Nie, S., Luo, Y., and Zhu, J.: Trends and scales of observed soil moisture variations in China, Adv. Atmos. Sci., 25(1), 43–58, 2008.

Ni-Meister, W., Houser, P. R., and Walker, J. P.: Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data

into a land surface model, J. Geophys. Res., 111, D20102, doi:10.1029/2006JD007190, 2006.

Niyogi, D. S., Xue, Y., and Raman, S.: Hydrological land surface response in a tropical regime and a midlatitudinal regime, J. Hydrometeorol., 3, 39–56, 2002.

Qin, J., Liang, S., Yang, K., Kaihotsu, I., Liu, R., and Koike, T.: Simultaneous estimation of both soil moisture and model parameters using particle filtering method through the assimilation of microwave signal, J. Geophys. Res., 114, D15103, doi:10.1029/2008JD011358, 2009.

Rawls, W. J., Brakensiek, D. L., and Saxton, K. E.: Estimation of soil water properties, Trans. Amer. Soc. Agric. Eng., 25, 1316–1320, 1982.

Reichle R. H. and Koster, R. D.: Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, Geophys. Res. Lett., 32, L02404, doi:10.1029/2004GL021700, 2005.

Reichle, R. H., Crow, W. T., and Keppenne, C. L.: An adaptive ensemble Kalman filter for soil moisture data assimilation, Water Resour. Res., 44, W03423, doi:10.1029/2007WR006357, 2008.

Reichle, R. H., Entekhabi, D., and McLaughlin, D. B.: Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach, Wate Resour. Res., 37, 2353–2364, 2001a.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Variational data assimilation of microwave radiobrightness observations for land surface hydrologic applications, IEEE Trans. Geosci. Remote Sens., 39, 1708–1718, 2001b.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, Mon. Weather Rev., 130, 103–114, 2002a.

Reichle, R. H., Walker, J. P., Koster, R. D., and Houser, P. R.: Extended versus ensemble Kalman filtering for land data assimilation, J. Hydrometeor., 3, 728–740, 2002b.

Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S., and Namkhai, A.: The global soil moisture data bank, B. Am. Meteorol. Soc., 81, 1281–1299, 2000.

Schaap, M. G. and Leij, F. J.: Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model, Soil Sci. Soc. Am. J., 64, 843–851, 2000.

van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, Soil Sci. Soc. Am. J., 44, 892–898, 1980.

Wang, D., Chen, Y., and Cai, X.: State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter, Water Resour. Res., 45, W11416, doi:10.1029/2008WR007401, 2009.

Walker, J. P. and Houser, P. R.: A methodology for initializing soil moisture in a global climate model: Assimilation of near surface soil moisture observations, J. Geophys. Res., 106, 11761–11774, 2001.

Wen, L., Gallichand, J., Viau, A. A., Delage, Y., and Benoit, R.: Calibration of the CLASS model and its improvement under agricultural conditions, T. ASAE, 41, 1345–1351, 1998.

Wosten, J., Pachepsky Y., and Rawls W.: Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, J. Hydrol., 251, 123–150, 2001.

Xia, Y., Pitman, A. J., Gupta, H. V., Leplastrier, M., Henderson-Sellers, A., and Bastidas, L. A.: Calibrating a land surface model of varying complexity using multicriteria methods and the Cabauw dataset, J. Hydrometeorol., 3, 181–194, 2002.

Zhang, S., Zeng, X., Zhang, W., and Barlage, M.: Revising the ensemble-based Kalman filter covariance for the retrieval of deep-layer soil moisture, J. Hydrometeorol., 11, 219–227, 2010.

Zhuang, J., Nakayama, K., Yu, G. R., and Miyazaki, T.: Predicting unsaturated hydraulic conductivity of soil based on some basic soil properties, Soil Till. Res., 59, 143–154, 2001.