

## Extended versus Ensemble Kalman Filtering for Land Data Assimilation

ROLF H. REICHLE

*Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, and  
Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland*

JEFFREY P. WALKER

*The University of Melbourne, Parkville, Victoria, Australia*

RANDAL D. KOSTER AND PAUL R. HOUSER

*Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland*

(Manuscript received 1 October 2001, in final form 5 April 2002)

### ABSTRACT

The performance of the extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF) are assessed for soil moisture estimation. In a twin experiment for the southeastern United States synthetic observations of near-surface soil moisture are assimilated once every 3 days, neglecting horizontal error correlations and treating catchments independently. Both filters provide satisfactory estimates of soil moisture. The average actual estimation error in volumetric moisture content of the soil profile is 2.2% for the EKF and 2.2% (or 2.1%; or 2.0%) for the EnKF with 4 (or 10; or 500) ensemble members. Expected error covariances of both filters generally differ from actual estimation errors. Nevertheless, nonlinearities in soil processes are treated adequately by both filters. In the application presented herein the EKF and the EnKF with four ensemble members are equally accurate at comparable computational cost. Because of its flexibility and its performance in this study, the EnKF is a promising approach for soil moisture initialization problems.

### 1. Introduction

Climate prediction at seasonal-to-interannual time-scales depends on accurate initialization of the slowly varying components of the earth's system, most notably sea surface temperature (SST) and soil moisture. While tropical SST is often the dominant source of predictability, its influence appears to be mostly limited to the Tropics (Koster et al. 2000b). Skill in the prediction of summertime continental precipitation and temperature anomalies in the extratropics may instead depend on the initialization of soil moisture and other land surface states. Since soil moisture controls the partitioning of the latent and sensible heat fluxes to the atmosphere, it can influence precipitation recycling.

The initialization of the land surface states for a seasonal climate forecast can be accomplished by assimilating soil moisture observations into the land model up to the start time of the prediction. With assimilation we attempt to combine the information from the observations and the model in an optimum way. Since for sea-

sonal forecasts we are only interested in the estimates at the start time of the prediction, sequential assimilation methods like Kalman filters are ideally suited to the task. The well-known extended Kalman filter (EKF) can be used for nonlinear applications, but the computational demand resulting from the error covariance integration limits the size of the problem (Gelb 1974). For this reason, the EKF has been used mostly for problems that focus on the estimation of the vertical soil moisture profile (Katul et al. 1993; Entekhabi et al. 1994). More recently, Walker and Houser (2001) have applied the EKF to soil moisture estimation across the North American continent by neglecting all horizontal error correlations and treating surface hydrological units (catchments) independently. This yields an effectively low-dimensional filter.

The ensemble Kalman filter (EnKF) is an alternative to the EKF (Evensen 1994). The EnKF circumvents the expensive integration of the state error covariance matrix by propagating an ensemble of states from which the required covariance information is obtained at the time of the update. Reichle et al. (2002) applied the EnKF to soil moisture estimation and found that it performed well against a variational assimilation method. Since the variational approach generally requires the

---

*Corresponding author address:* Dr. Rolf Reichle, NASA Goddard Space Flight Center, Code 974, Greenbelt Rd., Greenbelt, MD 20771.  
E-mail: reichle@janus.gsfc.nasa.gov

adjoint of the hydrologic model, which is not usually available and is difficult to derive, the obvious choices for advanced land assimilation algorithms are the EKF and the EnKF. There are many variants of the EKF and the EnKF that have been used in meteorology and oceanography, notably reduced-rank square root algorithms (Verlaan and Heemink 1997), particle filters (Pham 2001), methods that use pairs of ensembles (Houtekamer and Mitchell 1998), and hybrid approaches that combine ensembles with reduced-rank approaches (Heemink et al. 2001; Lermusiaux and Robinson 1999) or with variational methods (Hamill and Snyder 2000). In this paper, we focus on the relative merits of using the traditional EKF and EnKF for soil moisture assimilation.

The major differences between the EKF and the EnKF are (i) the approximation of nonlinearities of the hydrologic model and the measurement process (the EKF uses a linearized equation for the error covariance propagation while the EnKF nonlinearly propagates a finite ensemble of model trajectories), (ii) the range of model errors that can be represented (the EnKF can account for a wider range of model errors), (iii) the ease of implementation (the EKF requires derivatives of the nonlinear hydrologic model, evaluated numerically or from a tangent-linear model), (iv) computational efficiency (it must be determined how many ensemble members are needed in the EnKF to match the performance of the EKF), and (v) the treatment of horizontal correlations in the model or measurement errors (the EKF cannot account for horizontal error correlations in large systems for computational reasons). Insights into many important issues can be gained from low-dimensional versions of both filters.

Although approximate nonlinear filters such as the EKF and the EnKF have been found to work well in some applications, their value in a particular nonlinear problem cannot be assessed a priori but must be determined by simulations (Jazwinski 1970). We investigate the above differences in the context of soil moisture initialization for seasonal prediction using synthetic data in a twin experiment. Since all uncertain inputs are known by design, such experiments are well suited for a first assessment of algorithm performance. Tests with actual observations will be conducted in future studies. For retrospective analysis, surface soil moisture can be retrieved from the Scanning Multifrequency Microwave Radiometer (SMMR) for the period 1979–87 (Owe et al. 2001). These retrievals are derived from the 6.6-GHz (C band) and 37-GHz channels. Similar retrievals should soon be available from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E). In the future, passive 1.4-GHz (L band) sensors should also become available (Kerr et al. 2001).

## 2. Kalman filtering

The standard Kalman filter is the optimal sequential data assimilation method for linear dynamics and mea-

surement processes with Gaussian error statistics. The EKF is a variant of the Kalman filter that can be used for nonlinear problems (Gelb 1974). As an alternative, Evensen (1994) described a Monte Carlo approach to the nonlinear filtering problem, the EnKF, which is based on the approximation of the conditional probability densities of interest by a finite number of randomly generated model trajectories. In this section, we briefly review the filter equations and point out the main differences between the EKF and the EnKF.

### a. System model

We can express the nonlinear land surface model in the generic form

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{w}_k \quad (1)$$

by collecting the model prognostic variables of interest (in our case the soil water excess and deficit variables described in section 3) at time  $k$  into the state vector  $\mathbf{x}_k$  of dimension  $N_x$ . The nonlinear operator  $\mathbf{f}_k(\cdot)$  includes all deterministic forcing data (e.g., observed rainfall). Uncertainties related to errors in the model formulation or the forcing data are summarized in the model error term  $\mathbf{w}_k$ .

Suppose we assimilate soil moisture data that are sparse in time and space. If we assemble all observations taken at time  $k$  into the measurement vector  $\mathbf{y}_k$  the measurement process can be written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \quad (2)$$

where we have assumed a linear relationship to keep the notation simple. The operator  $\mathbf{H}_k$  relates the states (in our case the soil water excess and deficit variables) to the measured variables (in our case near-surface soil moisture). Measurement instrument errors and errors of representativeness are reflected in the measurement error term  $\mathbf{v}_k$ .

Adopting a probabilistic interpretation of uncertainty, we assume that  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are zero mean random variables with covariances  $\mathbf{Q}_k$  and  $\mathbf{R}_k$ , respectively. This provides a full statistical description if these random variables are normally distributed. For the discussion in this section we further assume that  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are mutually uncorrelated and white (uncorrelated in time), although these assumptions can be relaxed (Gelb 1974).

### b. Extended and ensemble Kalman filtering

Both the EKF and the EnKF work sequentially from one measurement time to the next, applying in turn a forecast step and an update step. Figure 1 highlights the key differences between the two filters. During the forecast step, the EKF propagates a single estimate of the state vector (from  $\mathbf{x}_{k-1}^+$  to  $\mathbf{x}_k^-$ ). The EKF also integrates the uncertainty of that estimate (the state error covariance, from  $\mathbf{P}_{k-1}^+$  to  $\mathbf{P}_k^-$ ), which is needed to determine the relative weights of the model forecast and the ob-

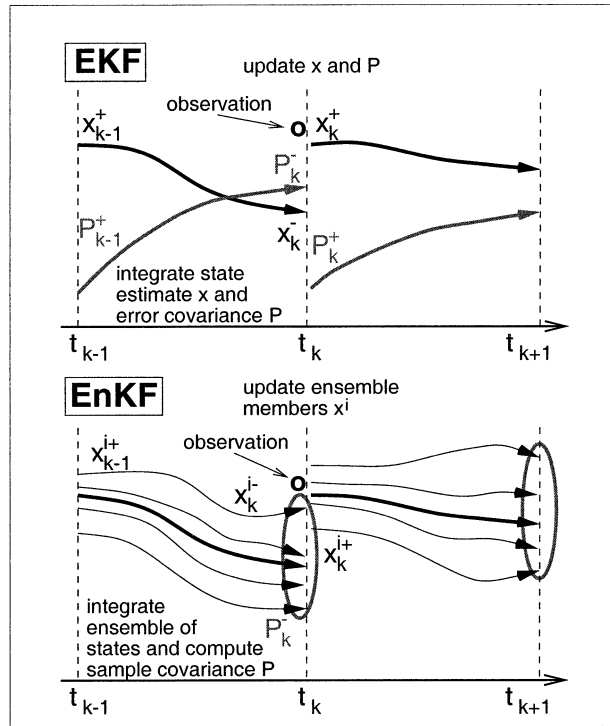


FIG. 1. Schematic of the extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF).

ervation at the update time. The EnKF, on the other hand, propagates an ensemble of state vectors in parallel, each state vector representing a particular realization of the possible model trajectories (e.g., with certain random errors in model parameters and/or a particular set of errors in forcing). The EnKF does not explicitly integrate the state error covariance but computes it instead diagnostically from the distribution of the model states across the ensemble.

During the update step, the EKF revises its estimate of the state vector (from  $\mathbf{x}_k^-$  to  $\mathbf{x}_k^+$ ) using the observation and the prognostic state error covariance  $\mathbf{P}_k^-$ . This reduces the uncertainty in the state estimate, which is reflected in the EKF update of the state error covariance (from  $\mathbf{P}_k^-$  to  $\mathbf{P}_k^+$ ). The EnKF, on the other hand, updates each ensemble member separately, using the observation and the diagnosed state error covariance  $\mathbf{P}_k^-$ . In the EnKF, the reduction of the uncertainty is reflected in the reduction of the ensemble spread. While the EKF state estimate at any time is simply the value of the state vector  $\mathbf{x}_k^-$  or  $\mathbf{x}_k^+$ , the EnKF state estimate is given by the mean of the ensemble members.

We now present a more formal discussion of the two approaches. Our knowledge of the state at the initial time  $k = 0$  is reflected by the mean state  $\mathbf{x}_0^-$  and its covariance  $\mathbf{P}_0^-$ , which are used to initialize the EKF. The EnKF is initialized by generating an ensemble of initial condition fields  $\mathbf{x}_0^{i-}$ ,  $i = 1, \dots, N$ , with mean  $\mathbf{x}_0^-$  and covariance  $\mathbf{P}_0^-$ . We start the assimilation cycle by cal-

culating a matrix of weights  $\mathbf{K}_k$  (the Kalman gain) for the update:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k]^{-1}. \quad (3)$$

If no observations are available at time  $k$  we formally set  $\mathbf{K}_k \equiv 0$ . Next, we update the state estimate (EKF) or each ensemble member (EnKF) using a linear combination of forecast model states and the observations:

$$\begin{aligned} \text{EKF: } \mathbf{x}_k^+ &= \mathbf{x}_k^- + \mathbf{K}_k [\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^-], \\ \mathbf{P}_k^+ &= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^-; \end{aligned} \quad (4a)$$

$$\begin{aligned} \text{EnKF: } \mathbf{x}_k^{i+} &= \mathbf{x}_k^{i-} + \mathbf{K}_k [\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^{i-} + \mathbf{v}_k^i], \\ i &= 1, \dots, N. \end{aligned} \quad (4b)$$

Here, the superscripts  $-$  and  $+$  refer to the state estimates, individual ensemble members, or covariances before and after the update, respectively. They are also known as forecast and analysis, respectively. Note that in the EnKF the data are perturbed by adding a random realization  $\mathbf{v}_k^i$  of the measurement error (Burgers et al. 1998).

In the forecast step, the EKF estimate is propagated forward in time with the nonlinear model, and in the EnKF each ensemble member is integrated using a corresponding ensemble of  $N$  random realizations of model error fields  $\mathbf{w}_k^i$ :

$$\text{EKF: } \mathbf{x}_{k+1}^- = \mathbf{f}_k(\mathbf{x}_k^+); \quad (5a)$$

$$\text{EnKF: } \mathbf{x}_{k+1}^{i-} = \mathbf{f}_k(\mathbf{x}_k^{i+}) + \mathbf{w}_k^i, \quad i = 1, \dots, N. \quad (5b)$$

We also propagate the state error covariance to account for the evolution of the uncertainty in the state estimates:

$$\begin{aligned} \text{EKF: } \mathbf{P}_{k+1}^- &= \mathbf{F}_k \mathbf{P}_k^+ \mathbf{F}_k^T + \mathbf{Q}_k, \\ [\mathbf{F}_k]_{mn} &= \left. \frac{\partial f_m}{\partial x_n} \right|_{\mathbf{x}_k^-}, \quad m, n = 1, \dots, N_x; \end{aligned} \quad (6a)$$

$$\begin{aligned} \text{EnKF: } \mathbf{P}_{k+1}^- &= \frac{1}{N-1} \mathbf{D}_{k+1} \mathbf{D}_{k+1}^T, \\ \mathbf{D}_{k+1} &= [\mathbf{x}_{k+1}^{1-} - \mathbf{x}_{k+1}^-, \dots, \mathbf{x}_{k+1}^{N-} - \mathbf{x}_{k+1}^-], \\ \mathbf{x}_{k+1}^- &= \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{k+1}^{i-}. \end{aligned} \quad (6b)$$

The importance of error covariance propagation is evident from Eq. (3), which describes how the optimal weights for the update depend on the error covariances. In the EKF,  $\mathbf{P}_k^-$  is obtained by propagating the posterior state error covariance from the last update time with a linearized matrix dynamic equation (6a). Integrating this equation for large  $N_x$  is very computationally demanding. This makes the application of the EKF to large-scale environmental assimilation problems impossible unless further approximations are made. In this study we use the EKF implementation of Walker and Houser (2001), in which all correlations between different

catchments are neglected. In this case the error covariance matrix is block-diagonal at all times and the computational burden is reduced considerably.

In the EnKF,  $\mathbf{P}_k^-$  is estimated from the spread of the ensemble prior to the update (6b). In this way,  $\mathbf{P}_k^-$  includes some effects of the nonlinear dynamics that are neglected in the EKF. On the other hand, the accurate estimation of  $\mathbf{P}_k^-$  now depends on the size of the ensemble. Note that in the EnKF the (analysis) error covariance  $\mathbf{P}_k^+$  is never needed, but parts or all of it can be computed at any time from the ensemble. Moreover, the forecast error covariance  $\mathbf{P}_k^-$  need not be constructed explicitly. For details of the implementation see Kepenne (2000).

### 3. Land model and experiment setup

#### a. Land surface model

Koster et al. (2000a) have developed a new land surface model, the Catchment Model, that uses hydrological catchments rather than a regular grid as the computational unit. The viability of their approach has been demonstrated by Ducharme et al. (2000). The Catchment Model has also performed well in the Project for Intercomparison of Land-Surface Parameterization Schemes Phase 2e and the Rhone Aggregation Experiment, which will be documented in forthcoming publications (S. Mahanama 2001, personal communication).

In the Catchment Model, vertical soil water transfer as well as lateral redistribution are modeled. The lateral movement of water is based on equilibrium concepts for the soil moisture profile (Beven and Kirkby 1979). The equilibrium soil moisture profile is determined from the catchment deficit, which is defined as the amount of water that would need to be added to bring the entire catchment to saturation. To allow for nonequilibrium vertical transfer of water, two additional variables are used. The surface excess and the root zone excess describe deviations from the equilibrium profile in the surface and root zone layers. We use a surface layer depth of 2 cm and a root zone layer that extends from the surface down to 1 m. The catchment deficit, root zone excess, and surface excess are model prognostic variables from which we can diagnose soil moisture content in the 2-cm surface layer, the 1-m root zone layer, and the total profile down to the water table (Walker and Houser 2001). We refer to these diagnostic variables as surface, root zone, and profile soil moisture, respectively.

In addition to soil moisture, the Catchment Model also predicts snow, heat transfer in the soil, and moisture and heat transfer in the canopy layer. Diagnostic outputs include the latent and sensible heat fluxes to the atmosphere as well as base flow and runoff. The total number of prognostic variables per catchment is 25 (3 for soil moisture, 3 for surface and canopy temperatures, 6 for subsurface temperature, 9 for snow, 3 for near-

surface humidity, and 1 for canopy interception). For our Kalman filtering applications, we use only the model prognostic variables that are directly related to soil moisture as state variables for the assimilation (catchment deficit, root zone excess, and surface excess). This means that we consider just three states per catchment. We assimilate synthetic observations of the (diagnostic) surface soil moisture (section 3b).

#### b. Twin experiment

Our twin experiment is conducted over a region of the southeastern United States that extends from 95° to 76°W longitude and from 24° to 35°N latitude. The domain contains 208 catchments with an average area of 3600 km<sup>2</sup>. In this region snow processes are relatively unimportant, which is ideal for our focus on soil moisture. On the other hand, parts of the region are covered by dense vegetation during the summer, and accurate remote sensing observations of soil moisture may be difficult to obtain (Jackson and Schmugge 1991). While dense vegetation could lead to a loss of accuracy in the soil moisture estimates when satellite data are assimilated, this does not influence the synthetic observations that we use, and our results about the relative performance of the EKF and the EnKF are not affected. The twin experiment starts with a model integration that serves as the “true” solution and is meant to represent nature. We start from a spinup initial condition on 1 January 1987 and integrate the model until 31 December 1987 using standard model parameters and forcing data from the International Satellite Land Surface Climatology Project (ISLSCP) (Sellers et al. 1996).

Next, we integrate the model again over the same time period but with an intentionally poor initial condition and different forcing data and model parameters. We use a perturbed initial condition generated by adding random noise to the initial surface excess, root zone excess, and catchment deficit with 1-, 10-, and 100-mm standard deviation, respectively. Instead of the ISLSCP data we use the reanalysis data of the European Centre for Medium-Range Weather Forecasts (ECMWF) as forcing inputs (Gibson et al. 1997). Table 1 gives an overview of the differences between the two forcing datasets. Precipitation, which is the most important input so far as soil moisture is concerned, is illustrated in Fig. 2 for a representative catchment. Moreover, we change the timescale parameters for moisture flow between the surface excess, root zone excess, and catchment deficit. Specifically, we use timescale parameters that have been derived for a 5-cm surface layer and a vertical decay factor  $\gamma = 2.17$  for the saturated hydraulic conductivity with depth (rather than for the 2-cm layer and  $\gamma = 3.26$  that we use in the true integration). Collectively, these “wrong” inputs and parameters represent our imperfect knowledge of the true land processes. The resulting fields constitute our best guess prior to assimilating the



TABLE 1. Space–time averages of the meteorological forcing inputs for the true model integration (ISLSCP) and root-mean-square difference between the true forcing and the forcing used in the estimation (ECMWF).

	Precipitation		Air pressure	Radiation		Temperature		Wind speed
	Total	Convective		Longwave	Shortwave	Air	Dewpoint	
	(mm day <sup>-1</sup> )	(mm day <sup>-1</sup> )	(mbar)	(W m <sup>-2</sup> )	(W m <sup>-2</sup> )	(°C)	(°C)	(m s <sup>-1</sup> )
ISLSCP	3.4	2.5	1006.5	353	189	18.3	12.5	3.1
ISLSCP – ECMWF	8.2	6.1	5.1	29	72	2.8	2.8	1.2

remote sensing data and will be referred to as the “prior” (no assimilation) solution.

The synthetic observations used in the assimilation are derived from the true fields by adding random measurement noise. In particular, we generate synthetic observations of the soil moisture content in the 2-cm surface layer (“surface soil moisture”) with an error of 5% (volumetric) once every 3 days for all catchments. These data are subsequently assimilated into the model using the “wrong” forcing and model parameters described above. The resulting fields are referred to as the “estimates.”

### c. Filter calibration

The setup of the twin experiment implies that we do not know the exact statistics of the model errors. In fact, we do not even expect that additive model errors will fully account for the differences between the true and prior fields. In any case, filter performance depends strongly on our choice of model error parameters, so we must choose them very carefully. To ensure a fair comparison of the EKF and the EnKF, we find the parameters that allow each filter to perform the best it can.

An advantage of the EnKF is its flexibility in representing various types of model errors. Besides adding synthetic model error fields, which will be described below, we could use different forcing fields and model parameters for each ensemble member or even use different models altogether, provided that the models describe identical physical variables. In this study, we perturb the ECMWF meteorological data that are used to

force each ensemble member. Standard deviations for these perturbations are 5 K for air and dewpoint temperatures, 1 m s<sup>-1</sup> for wind speeds, 50 W m<sup>-2</sup> (25 W m<sup>-2</sup>) for shortwave (longwave) radiative fluxes, 10 mbar for surface pressure, and 50% of magnitude for precipitation. These numbers are based on simple order-of-magnitude considerations and have not been tuned. They can, however, be compared to the actual differences of the ISLSCP and ECMWF datasets listed in Table 1. Such forcing perturbations represent nonadditive model errors.

In the EKF, only additive model errors can be taken into account by specifying the model error covariances  $\mathbf{Q}_k$ . In the EnKF, we add synthetic model error fields to each ensemble member (in addition to the forcing perturbations that represent nonadditive model errors). These synthetic error fields are generated from a specified covariance matrix assuming a normal probability distribution. In both cases we assume that the standard deviation of each type of model error is identical for all catchments. Furthermore, all model errors are assumed uncorrelated; that is,  $\mathbf{Q}_k$  is diagonal. For the EnKF, we also impose a correlation time of 3 days on the model error time series (autoregressive process of order one), which is inexpensive when the state is not augmented (Reichle et al. 2002). Temporally correlated model errors are not considered in the EKF because this would require state augmentation and significantly increase the computational burden.

With all inputs fixed except the magnitude of the model error variances for the surface excess, root zone excess, and catchment deficit, we calibrate these remaining parameters to achieve the best possible filter performance. Since the twin experiment is designed such that the true solution is known, a convenient measure of estimation performance is the actual error, which is the difference between the true soil moisture and its EKF or EnKF estimate. As an aggregate measure of filter performance we sum up the average actual errors in the surface excess, root zone excess, and catchment deficit, where the average is taken in the root-mean-square sense over all catchments from February to December 1987. The first month of the assimilation is excluded to avoid initialization effects. This aggregate measure gives more relative weight to errors in the root zone excess and the catchment deficit, which are more important for seasonal prediction than errors in the surface excess.

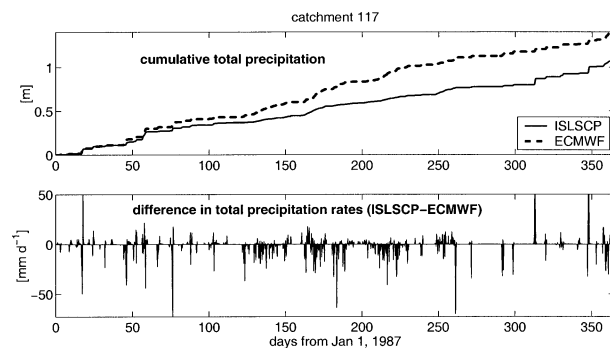


FIG. 2. Comparison between the total precipitation of the ISLSCP and the ECMWF datasets for a representative catchment: (top) the cumulative total precipitation; (bottom) the difference between the total precipitation rates (ISLSCP minus ECMWF).

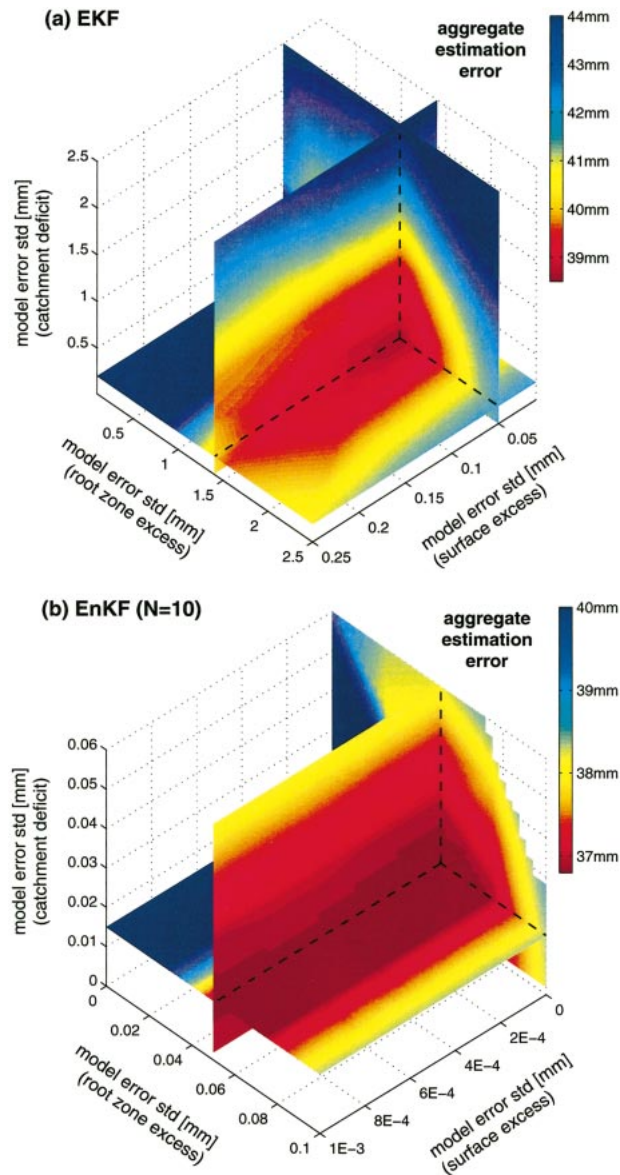


FIG. 3. Aggregate estimation error as a function of the model error standard deviations for the (a) EKF and (b) EnKF with  $N = 10$  ensemble members. The difference in scales in the aggregate estimation error reflects the superior performance of the EnKF. The difference in scales in the model error parameters is due to the difference in model error correlation times (section 3c).

For each filter, we have computed the aggregate estimation errors of about 200 integrations with different model error variances. Figure 3 shows our aggregate performance measures as a function of the model error standard deviations. For both filters we find a single global minimum (at the intersection of the slices in Fig. 3). At the minimum, the model error standard deviation is greater for the root zone excess than for the catchment deficit and the surface excess. Note that for the EnKF the model error variance in the surface excess matters little because we also perturb the forcing inputs.

The true model error statistics are a unique attribute of the model (and associated forcing data) and can be represented only approximately by the assimilation algorithm. In the EKF, for example, there is an implicit assumption of temporally uncorrelated model errors, which explains the larger calibrated model error variances compared to the EnKF. Since soil moisture integrates over the model error, adding temporally correlated model error of a given variance in the EnKF leads to much larger ensemble spread (in soil moisture) than adding temporally uncorrelated error of the same variance. Section 4f will show that the state error variances of the EKF and the EnKF are largely in agreement.

The calibrated parameters are insensitive to our choice of aggregate performance measure. The resulting parameters are almost identical when we calibrate against the average of the errors in the surface, root zone, and profile moisture contents, a criterion which gives much more weight to the surface layer. Our calibration of model error parameters serves mainly to make a fair comparison of the EKF and the EnKF possible. There are many adaptive methods to determine error statistics during filter operation (Dee 1995). Finally, note that the calibration of the EnKF is somewhat incomplete because we did not optimize the size of the forcing perturbations or the correlation time of the model errors, although this might have further improved the EnKF's performance. Table 2 summarizes the calibrated model error parameters.

## 4. Results and discussion

### a. Soil moisture estimates

In this section we discuss the results of the twin experiment described in section 3. Figure 4 shows the time average (root-mean-square) actual errors of the moisture content variables from February to December 1987. Recall that the actual errors are the differences between the true soil moisture (from the control experiment) and its EKF or EnKF estimate. Obviously, the errors are higher for the surface moisture content than for the root zone and profile moisture contents. This is because the surface moisture content varies on timescales of a day or less, while we assimilate observations only once every 3 days. When an observation of surface soil moisture is assimilated, the estimation error of the surface moisture content typically falls well below 5%. But between observation times, errors in the model timescales and in the forcing (notably in precipitation) degrade the surface estimates significantly. Thus, to improve the quality of the surface moisture estimates it would be necessary to assimilate observations more frequently.

The situation is different for the root zone and profile moisture contents. These lower layers exhibit greater memory, and variations in their moisture content occur over longer timescales. Consequently, short-term errors in the forcing do not significantly impact the root zone

TABLE 2. Inputs to the true, prior, and assimilation integrations. Model error standard deviations  $\sigma$  are calibrated (section 3c). Forcing inputs for individual ensemble members are perturbed from ECMWF data (see text). The scalar  $\gamma$  is the exponential decay factor of the saturated hydraulic conductivity with depth.

	True	Prior	EKF	EnKF
Initial condition	Spinup		Perturbed soil water excess/deficit	
Forcing	ISLSCP		ECMWF	Perturbed ECMWF
Parameters	Calibrated for 2-cm surface layer and $\gamma = 3.26$		Timescales for 5-cm surface layer and $\gamma = 2.17$	
Model error				
$\sigma_{\text{surface excess}}$	N/A	Unknown	0.05 mm	0
$\sigma_{\text{root zone excess}}$	N/A	Unknown	1.40 mm	0.050 mm
$\sigma_{\text{catchment deficit}}$	N/A	Unknown	0.20 mm	0.015 mm
Correlation time	N/A	Unknown	0	3 days

and profile estimates. Table 3 lists the time and space average (root-mean-square) actual errors of the moisture content variables and the state variables. We can see that the improvement over the prior estimates from the assimilation is relatively small in the surface and the root zone excess. By comparison, the catchment deficit is much closer to the truth after the assimilation. The difference in the performance of the EKF and the EnKF is small when compared to the prior errors. Nevertheless, the EnKF with  $N = 4$  ensemble members performs as well as the EKF, and it outperforms the EKF for  $N \geq 10$  (section 4b).

The computational effort of the EnKF is largely de-

termined by the size of the ensemble that is propagated. For the EKF the numerical differentiation scheme implies that the computational cost corresponds roughly to an ensemble of  $m + 1$  members, where  $m$  is the number of state variables per catchment. In our application  $m = 3$  and the computational effort of the EKF corresponds to an ensemble of four members. This means that the EKF and the EnKF are equally expensive for comparable performance (Table 3).

To assess further the performance of the filters, we can compare the actual errors to what the filters “think” they should be. These expected errors are given by the square root of the diagonal elements of the error co-

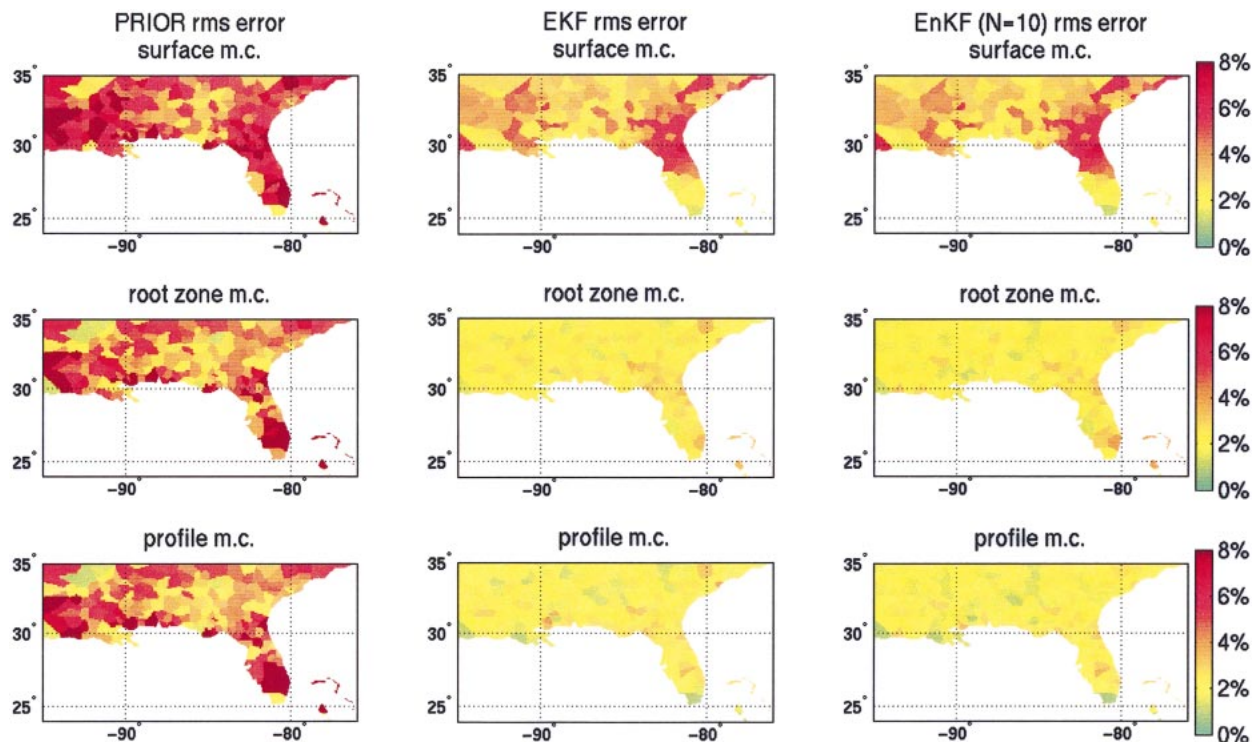


FIG. 4. Time-average error of the moisture content (m.c.) (left-hand column) prior to the assimilation, (middle column) for the EKF, and (right-hand column) for the EnKF with  $N = 10$  ensemble members. Shown are the errors for the (top row) surface, (middle row) root zone, and (bottom row) profile soil moisture content. The average is from Feb to Dec 1987 in the rms sense. Units are volumetric moisture percent.



TABLE 3. Actual errors (root-mean-square average over all catchments from Feb to Dec 1987) of the moisture content (m.c., in volumetric percent) and the state variables. Filter-derived (expected) error standard deviations are shown in parentheses. Moisture content errors are computed from 6-hourly output, excess/deficit errors from daily output.

Ensemble size $N$ (—)	Prior N/A	EKF N/A	EnKF				
			4	10	30	100	500
Surface m.c. (%)	6.1	3.7	3.7	3.6	3.6	3.6	3.6
Root zone m.c. (%)	5.3	2.3	2.3	2.2	2.2	2.2	2.2
Profile m.c. (%)	5.6	2.2	2.2	2.1	2.1	2.0	2.0
Surface excess (mm)	0.86	0.74 (0.43)	0.71 (0.36)	0.69 (0.39)	0.68 (0.41)	0.68 (0.41)	0.68 (0.42)
Root zone excess (mm)	6.3	4.6 (11)	4.5 (5.0)	4.3 (5.1)	4.3 (5.2)	4.3 (5.2)	4.3 (5.2)
Catchment deficit (mm)	120	33 (31)	34 (24)	32 (26)	31 (27)	30 (28)	30 (28)

variance matrix  $\mathbf{P}$  (section 2) and are summarized in Table 3. Both filters clearly underestimate the actual errors in the surface excess, and the EKF overestimates the errors in the root zone excess by more than a factor of 2. While it is possible to tune both filters in such a way that the expected errors match the actual errors more closely, this would imply an increase in the estimation errors, which contradicts the objective of our filter calibration (section 3c).

The discrepancy between expected and actual errors is the result of nonlinearities and our poor knowledge of the true model errors. Recall that the true soil moisture has been derived with forcings and model parameters that are different from the ones used in the estimation (Table 2). In the filtering framework, we try to account for such deficiencies by making statistical assumptions about the model error term  $\mathbf{w}$ . We specify the statistical properties of  $\mathbf{w}$ , most notably its covariance  $\mathbf{Q}$ , which has a direct influence on the weights  $\mathbf{K}$  that are used in the update. The mismatch between expected and actual errors suggests that the differences between the true solution and our best (prior) guess are not fully represented by additive Gaussian errors (or by additional forcing errors in the EnKF). Nevertheless, considering that we only assimilate surface soil moisture once every 3 days, the resulting estimates are quite good.

#### b. Convergence of the EnKF with ensemble size

Obviously, the EnKF's most critical approximation is the finite size of the ensemble, and it is important to understand how many ensemble members are needed to obtain satisfactory estimates. Table 3 shows that for ensemble sizes of four or more the average actual errors of the EnKF are equal to or smaller than the EKF errors. If the problem had been linear and all errors had been Gaussian, the EKF would have been more accurate, and the EnKF errors would have converged to the EKF errors only as the ensemble size tended toward infinity. The superior performance of the EnKF in our application must be due to the nonlinear nature of the prob-

lem and the EnKF's greater flexibility in representing a wide range of model errors (sections 4d and 4f).

The EnKF estimation errors change little with the size of the ensemble, and convergence is achieved quickly. This fast convergence is, of course, related to the effectively very small size of the state vector. Since there are only 3 degrees of freedom in each catchment (the three soil water excess and deficit variables), and since all catchments are treated independently, a small ensemble is sufficient to achieve good results. To suppress statistical noise in small ensembles, we force the sample mean of the synthetic error fields to match the theoretical mean of zero. This idea could be taken further by generating second-order accurate ensembles (Pham 2001) in which the model error trajectories are generated in such a way that their sample covariance is exactly equal to the prescribed theoretical covariance  $\mathbf{Q}$ .

The actual errors of the state estimates are only one of many possible performance criteria. Covariance estimates, for instance, are rather noisy with a small ensemble of 10 or fewer members. To illustrate this point, Fig. 5 shows the analysis error standard deviations for a typical catchment. Here, a larger ensemble is clearly superior. The error standard deviations of the 20-member ensemble are very close to the 500-member ensemble and therefore are not shown in Fig. 5. Similar results are found for the correlation coefficients (section 4f). Additional experiments using synthetic observations of near-surface soil moisture with 2% measurement error (as opposed to the 5% measurement error used throughout the paper) show that the relative advantage of the 500-member EnKF over the EKF is larger when the observations are more accurate. Note finally that the requirements on the size of the ensemble are bound to increase once horizontal correlations are taken into account.

#### c. Numerical considerations and the EKF

In the EKF we need the state transition matrix  $\mathbf{F}$  of the linearized dynamical system for the propagation of the state error covariance (6a). Since the Catchment



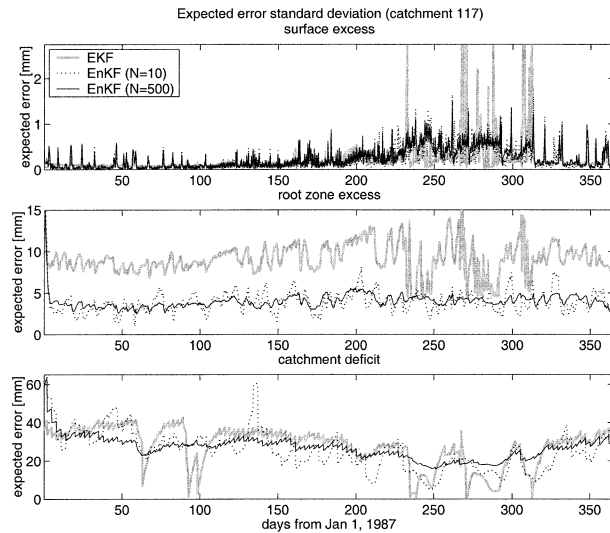


FIG. 5. Filter-derived (expected) error standard deviations of the state variables for a representative catchment: (top) surface excess, (middle) root zone excess, and (bottom) catchment deficit.

Model includes many switches, analytic derivatives are difficult to obtain. Walker and Houser (2001) therefore evaluate  $\mathbf{F}$  numerically and approximate the derivative via  $F = df/dx \approx [f(x+h) - f(x)]/h$ . Although conceptually straightforward, this numerical differentiation scheme is not without problems. As shown in Fig. 5, the error standard deviation of the surface excess in our representative catchment becomes very large around days 232, 267, 287, and 306. This is attributed to numerical problems in the calculation of the state transition matrix  $\mathbf{F}$ . The problem is that small perturbations  $h$  typically lead to numerical problems, while large perturbations result in a loss of accuracy in the derivative and are also more likely to hit nonlinear thresholds. When implemented without additional constraints, the numerical differentiation scheme fails frequently, which has a negative effect on the soil moisture updates.

In practice, Walker and Houser (2001) found it necessary to implement various checks on the EKF covariance propagation (6a). Due to numerical problems with the linearization, the state error covariance matrix  $\mathbf{P}$  is not always positive definite. In such cases, the covariance is reset according to a set of prespecified rules. Likewise, the covariance is confined within prespecified bounds and reset if these bounds are exceeded. Note that every time the error covariance is reset, information from earlier updates is partially lost. We have measured the influence of this by excluding from the error average calculation the first 3 days after each covariance reset. These modified average estimation errors are shown in Table 4. Although the errors generally decrease when the problematic times are excluded from the average, the relative performance of the EKF and the EnKF remains the same. This means that the interruption of the EKF covariance propagation is not a ma-

TABLE 4. Average actual errors of the moisture content (m.c., in volumetric percent) with the first 3 days excluded from the average calculation after each EKF covariance reset (section 4c).

Ensemble size $N$ (—)	Prior N/A	EKF N/A	EnKF		
			4	10	100
Surface m.c. (%)	6.0	3.4	3.5	3.4	3.3
Root zone m.c. (%)	5.3	2.3	2.3	2.2	2.1
Profile m.c. (%)	5.7	2.2	2.2	2.1	2.0

ajor source of error, and the numerical instabilities experienced by the EKF do not affect the comparison with the EnKF.

#### d. Measuring nonlinearity

There are generally two kinds of nonlinearities that appear in a hydrological model: differentiable functions and nondifferentiable switches and steps. The first kind, differentiable functions, can be treated with a standard Taylor series expansion of the model trajectory around the most recent estimate, as is done in the EKF for the error covariance forecast. For nonlinearities of the second kind, which are inherently nondifferentiable, we cannot expect that the linearization approach of the EKF will produce accurate estimates.

Verlaan and Heemink (2001) described a nondimensional number  $\mathcal{V}^2 \equiv \mathbf{b}^T \mathbf{P}^{-1} \mathbf{b}$  to measure nonlinearities, where  $\mathbf{b}$  is the bias in the estimate and  $\mathbf{P}$  is the state error covariance. The bias is related to nonlinearities in the model (see the appendix). By construction,  $\mathcal{V}$  can only measure nonlinearities that are differentiable. Such nonlinearities are significant if  $\mathcal{V} \gg \sqrt{N_x}$ , where  $N_x$  is the dimension of the state vector. We have computed  $\mathcal{V}$  for our application and found that differentiable nonlinearities are largely insignificant. Since the catchments are completely uncorrelated in this study, we compute  $\mathcal{V}$  for each catchment separately and compare it to the effective state dimension  $N'_x = 3$ . It turns out that  $\mathcal{V}$  exceeds  $\sqrt{3}$  at only 1.5% of all computational nodes and time steps between 1 April and 31 December 1987. (The first three months of 1987 are neglected to eliminate initial condition effects of the bias integration.) We have found no evidence that errors in the moisture content grow with  $\mathcal{V}$ . Moreover, the differences between the estimates from the standard EKF and a bias-corrected assimilation are negligible. We conclude that the first-order local linearization of the EKF adequately accounts for differentiable nonlinearities in the Catchment Model and that finite higher-order corrections add little information.

It is important to reiterate that  $\mathcal{V}$  cannot yield information about the impact of step functions and switches. The Catchment Model, like any other land surface model, contains many such nondifferentiable nonlinearities. Another measure for the impact of nonlinearities in the model can be obtained from the EnKF. Nonlinearities, differentiable or not, are likely to induce asymmetries

in the sample distribution of the ensemble members. We define the skewness coefficient as  $s = \mathcal{E}\{[x - \mathcal{E}(x)]^3\} / \sigma_x^3$ , where  $\mathcal{E}(\cdot)$  is the expectation operator and  $\sigma_x$  is the standard deviation of  $x$ . For the surface excess,  $s > 5$  for 8% of all times and catchments (EnKF with 500 ensemble members). The primary reason for high skewness is that soil moisture and the corresponding Catchment Model states have upper and lower bounds. A positive skewness coefficient indicates that the distribution has a large, positive tail and is concentrated at the lower end. We find that such positive skewness occurs when the soil dries out completely at the surface for lack of rain.

The skewness information that we gain in the EnKF is very informative, but it is not fully used in the EnKF update. Recall that for the update we derive only the sample covariance from the ensemble. Higher-order moments, although present and fully propagated in the ensemble, are not used in the computation of the gain matrix (3). Fortunately, high skewness does not imply large estimation errors. In fact, when the distribution of the surface excess is very positively skewed with a narrow peak close to the lower bound, it is likely that the soil is in fact very dry. An update in such a case will most likely produce a soil moisture value close to the lower bound, regardless of the sophistication of the update scheme. In summary, the bias and skewness results demonstrate that nonlinearities are in fact present but are not a dominant source of estimation error in the EKF and the EnKF.

#### e. Innovations

Examination of the innovations sequence is a standard tool to evaluate filter performance. This tool is particularly important because it can also be applied in an operational setting when the true soil moisture is unknown and actual errors cannot be derived. The innovations sequence  $\mathbf{v}_k \equiv \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^-$  describes the difference between the actual observations and the forecast. If the problem is linear and the filter operates in accordance with its underlying statistical assumptions,  $\mathbf{v}_k$  is a Gaussian and white (temporally uncorrelated) process with covariance  $\mathcal{E}(\mathbf{v}_k \mathbf{v}_k^T) = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k$ . This term is easily output from both filters, which allows us to normalize the innovations. Figure 6 shows a relative histogram of the normalized innovations of all catchments and all updates. Also shown is the standard normal distribution  $\mathcal{N}(0, 1)$ . We can see that the normalized innovations are not fully consistent with a standard normal distribution. The histogram is broader, which reflects the underestimation of the actual covariance of the innovations. This is to be expected, given that the filter-derived error variances typically underestimate the actual errors (Table 3).

We can test for the whiteness of the innovations sequence by computing its sample autocorrelation function (Jenkins and Watts 1968). Out of the 208 catch-

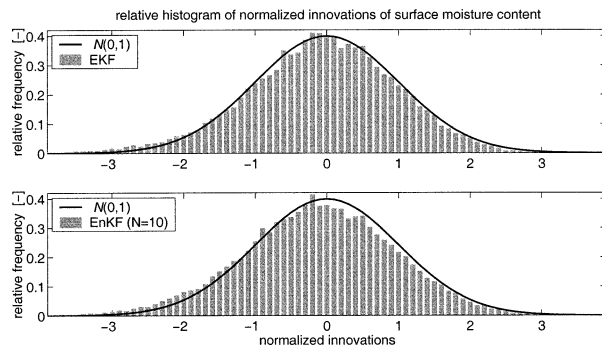


FIG. 6. Relative histogram of the innovations for all catchments and all update times. For comparison, the probability density of the standard normal distribution  $\mathcal{N}(0, 1)$  is also shown.

ments, the EKF innovations sequence of 24 catchments is not white at the 5% significance level (its lag-one autocorrelation coefficient does not contain zero in a 95% confidence interval). Similarly, the EnKF innovations for 27 (or 24; or 22) catchments using  $N = 4$  (or  $N = 10$ ; or  $N = 500$ ) are not white at the 5% significance level. Moreover, for some catchments (and both filters) the sample autocorrelation function exhibits oscillatory behavior, which also suggests that the innovations sequence is not perfectly white. In summary, both filters produce innovations that indicate slightly suboptimal performance, which stems from the imperfect representation of the model errors and from the presence of nonlinearities.

#### f. Error covariance modeling

The EKF and the EnKF differ mostly in how they approximate the error covariance propagation (6a,b). This has implications for how model error covariances can be represented in each filter. Note the difference in the analysis error standard deviation of the surface excess for the typical catchment shown in Fig. 5. While the error standard deviation of the surface excess varies rapidly in the EnKF, the EKF produces much smoother error standard deviations at the beginning of the year. This difference is entirely dependent upon the experiment setup. Here, we choose to add errors to the forcings of each ensemble member according to the actual forcing conditions. For example, we added larger errors when the forcing indicated that precipitation was falling. This leads to the very nonstationary behavior of the EnKF error standard deviation. In the EKF, on the other hand, a constant model error covariance  $\mathbf{Q}$  was added at each time step to the forecast error covariance (6a).

In addition to the error standard deviations, the filters also produce error correlations for the states and the measured variable, in our case surface soil moisture. These correlations can be derived easily from the off-diagonal elements of the state error covariance matrix  $\mathbf{P}$  and the measurement operator  $\mathbf{H}$  (EKF), or directly from the ensemble (EnKF). Figure 7 shows time series

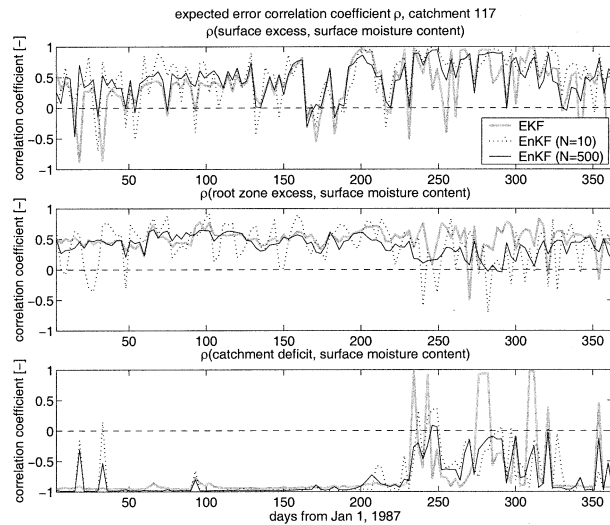


FIG. 7. Filter-derived error correlation coefficients for a representative catchment.

of the correlation coefficients for a representative catchment. As expected, the error correlation between the surface excess (or root zone excess) and the surface moisture content is mostly positive, with the correlation being more erratic in the case of the surface excess. Likewise, we find the expected anticorrelation between errors in the catchment deficit and the surface moisture content. This strong coupling between the surface soil moisture and the profile variables is particular to the Catchment Model. The catchment deficit describes the equilibrium profile for a given amount of water within the catchment and thereby determines the surface soil moisture to first order. The surface and root zone excess terms are only corrections to the equilibrium profile. Provided that we succeed in a satisfactory model calibration, the Catchment Model approach offers great advantages for estimating deep soil moisture from observations of the surface moisture content.

Figure 7 also illustrates that the correlations change with the general hydrologic conditions. There are strong (anti-) correlations between the root zone excess (or the catchment deficit) and the surface moisture content in the first half of the year, when the catchment is relatively wet. In the second half of the year, the catchment is much drier and the root zone excess and catchment deficit decouple from the surface soil moisture, while the surface excess is more strongly correlated to surface moisture content. Generally, the EKF and the EnKF produce correlations that are consistent.

## 5. Summary and conclusions

In this paper, we compare two promising data assimilation methods for soil moisture initialization in seasonal climate prediction. The extended Kalman filter (EKF) and the ensemble Kalman filter (EnKF) were

used to assimilate synthetic surface soil moisture observations into the Catchment Model, with model error parameters calibrated against actual estimation errors. The best results are obtained for both filters when the model error in the root zone excess is large compared to the model errors in the surface excess and the catchment deficit. Using the calibrated filter parameters we find that the EKF and the EnKF produce satisfactory estimates of soil moisture.

The EKF and the EnKF (with four ensemble members) show comparable performance for comparable computational effort. For 10 or more ensemble members, the EnKF outperforms the EKF. This is ascribed to the EnKF's flexibility in representing nonadditive model errors. The actual estimation errors of the EnKF converge quickly with increasing ensemble size, even though the filter-derived (expected) error covariances are noisy for small ensembles. The numerical differentiation scheme used in the EKF requires frequent checks in order to avoid divergent error covariances or loss of positive definiteness. Although these checks interrupt the integration of the error covariances, and information from earlier updates is partially lost, they are not a major source of error.

The normalized innovations are found to be inconsistent with a standard normal distribution. This is because our representation of model errors cannot fully account for the effects of uncertainties in the forcing and imperfectly known model parameters that we use in our twin experiment. Nonlinearities in the land model generate skewness in the distribution of ensemble states. But this skewness information is only very approximately used in the EnKF update and is not available in the EKF. Fortunately, the nonlinearities are not a dominant source of error, because the local linearization strategy of the EKF is for the most part successful and because the nature of the soil moisture bounds limits the actual estimation errors.

Catchment-to-catchment error correlations could arise from large-scale errors in the forcing or from unmodeled lateral fluxes such as river or groundwater flow. Moreover, satellite data are likely to exhibit horizontal error correlations. The present paper compares the EKF and EnKF under the assumption that horizontal error correlations can be neglected. The importance of such correlations is a topic of active research. If horizontal error correlations turn out to be important, information can be spread laterally, in particular from observed to unobserved catchments. When horizontal error correlations are taken into account in the EnKF, small error correlations associated with observations that are far apart must be filtered out (Mitchell and Houtekamer 2000). For computational reasons, the EKF must be approximated using a rank-reduction technique such as the reduced-rank square root method (Verlaan and Heemink 1997).

Before soil moisture assimilation can become a routine tool for seasonal climate prediction, many more

questions will need to be addressed. Important areas of research include the investigation of multivariate assimilation using more Catchment Model prognostic variables as states, the direct assimilation of radiances as opposed to soil moisture retrievals, and the assimilation of other types of remote sensing data such as soil temperatures or vegetation parameters. Finally, soil moisture estimates from the assimilation must then be shown to improve the accuracy of seasonal climate forecasts. In summary we can say that the EnKF is more robust and offers more flexibility in covariance modeling (including horizontal error correlations). This leads to its slightly superior performance in our study and makes the EnKF a promising approach for soil moisture initialization of seasonal climate forecasts.

**Acknowledgments.** This research was sponsored by the NASA Seasonal-to-Interannual Prediction Project. We would like to thank Kenneth Mitchell, Wade Crow, and two anonymous reviewers for insightful reviews, Michele Rienecker and Max Suarez for many discussions, and Sarith Mahanama, Aaron Berg, and Sally Holl for their support with the data.

## APPENDIX

### Bias and Nonlinearity

The differentiable part of the nonlinearities can be examined by integrating an estimate of the bias along with the state estimate. This yields a nondimensional number that describes the importance of nonlinearities. We follow the method described by Verlaan and Heemink (2001) and refer to their paper for details. If the forward operator  $\mathbf{f}$  is nonlinear, the EKF forecast equation (5a) becomes biased. Using a Taylor series expansion we get

$$\mathcal{E}[\mathbf{f}(\mathbf{x})] = \mathbf{f}[\mathcal{E}(\mathbf{x})] + \frac{1}{2}(\mathbf{P}\partial^2\mathbf{f}) + \dots, \quad (\text{A1})$$

where  $\mathcal{E}(\cdot)$  is the expectation operator and  $[\mathbf{P}\partial^2\mathbf{f}]_i = \sum_{m,n=1}^{N_x} \mathbf{P}_{mn}(\partial^2 f_i / \partial x_m \partial x_n)$  ( $i = 1 \dots N_x$ ) is a second-order correction term. Higher-order terms are neglected. Let us now define the bias as the expected error of the estimate with respect to the true state  $\mathbf{x}_k$ , that is,  $\mathbf{b}_k^+ \equiv \mathcal{E}[\mathbf{x}_k - \mathbf{x}_k^+]$  and  $\mathbf{b}_k^- \equiv \mathcal{E}[\mathbf{x}_k - \mathbf{x}_k^-]$ . The bias is integrated according to (Verlaan and Heemink 2001)

$$\mathbf{b}_k^- = \mathbf{F}_{k-1}\mathbf{b}_{k-1}^+ + [\mathbf{P}\partial^2\mathbf{f}]_{k-1} \quad (\text{A2})$$

$$\mathbf{b}_k^+ = \mathbf{b}_k^- - \mathbf{K}_k \mathbf{H}_k \mathbf{b}_k^-, \quad (\text{A3})$$

starting from the initial condition  $\mathbf{b}_0^+ \equiv 0$ . Obviously, for linear  $\mathbf{f}(\cdot)$  we have  $\mathbf{b}_k \equiv 0$ . The bias is forced by the product of system nonlinearities ( $\partial^2\mathbf{f}$ ) and the uncertainty ( $\mathbf{P}$ ). This means that system nonlinearities are less important when the uncertainty is small.

When the estimate is biased, the expected magnitude of the state estimation errors is approximately given by

$\mathbf{P} + \mathbf{b}\mathbf{b}^T$  (Dee and da Silva 1998). The relative importance of the bias can then be measured by  $\mathcal{V}^2 = \mathbf{b}^T \mathbf{P}^{-1} \mathbf{b}$  (Verlaan and Heemink 2001). Since for unbiased estimates the expected value of the log likelihood function is equal to  $N_x$ , the bias is significant for  $\mathcal{V} \gg \sqrt{N_x}$  and insignificant for  $\mathcal{V} \ll \sqrt{N_x}$ . Note again that  $\mathcal{V}$  can only represent nonlinearities in the model that are differentiable.

## REFERENCES

- Beven, K. J., and M. J. Kirkby, 1979: A physically-based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.*, **24**, 43–69.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- , and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295.
- Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model. 2: Parameter estimation and model demonstration. *J. Geophys. Res.*, **105**, 24 823–24 838.
- Entekhabi, D., H. Nakamura, and E. G. Njoku, 1994: Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Trans. Geosci. Remote Sens.*, **32**, 438–448.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- Gelb, A., Ed., 1974: *Applied Optimal Estimation*. The MIT Press, 374 pp.
- Gibson, J., P. Kallberg, S. Uppala, A. Hernandez, A. Nomura, and E. Serrano, 1997: ERA description. ECMWF Re-Analysis Project Report Series, No. 1, European Centre for Medium-Range Weather Forecasts, 84 pp.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- Heemink, A. W., M. Verlaan, and A. J. Seegers, 2001: Variance reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718–1728.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Jackson, T. J., and T. J. Schmugge, 1991: Vegetation effects on the microwave emission of soils. *Remote Sens. Environ.*, **36**, 203–212.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Jenkins, G. M., and D. G. Watts, 1968: *Spectral Analysis and Its Applications*. Holden-Day, 525 pp.
- Katul, G. G., O. Wendroth, M. B. Parlange, C. E. Puente, M. V. Folegatti, and D. R. Nielsen, 1993: Estimation of in situ hydraulic conductivity function from nonlinear filtering theory. *Water Resour. Res.*, **29**, 1063–1070.
- Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971–1981.
- Kerr, Y. H., P. Waldeufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger, 2001: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1729–1735.
- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000a: A catchment-based approach to modeling land surface



- processes in a general circulation model. 1: Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822.
- , —, and M. Heiser, 2000b: Variance and predictability of precipitation at seasonal to interannual timescales. *J. Hydrometeor.*, **1**, 26–46.
- Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, **127**, 1385–1407.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–432.
- Owe, M., R. de Jeu, and J. Walker, 2001: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1643–1654.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.
- Reichle, R., D. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114.
- Sellers, P. J., and Coauthors, 1996: The ISLSCP Initiative I global datasets: Surface boundary conditions and atmospheric forcings for land–atmosphere studies. *Bull. Amer. Meteor. Soc.*, **77**, 1987–2005.
- Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrol. Hydraul.*, **11**, 349–368.
- , and —, 2001: Nonlinearity in data assimilation applications: A practical method for analysis. *Mon. Wea. Rev.*, **129**, 1578–1589.
- Walker, J. P., and P. R. Houser, 2001: A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *J. Geophys. Res.*, **106**, 11 761–11 774.