An ensemble-based reanalysis approach to land data assimilation

Susan Dunne and Dara Entekhabi

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[1] Future pathfinder missions such as NASA's Hydrosphere State (Hydros) and ESA's Soil Moisture and Ocean Salinity (SMOS) will provide satellite-based global observations of surface (0–5 cm) soil moisture. In previous work an ensemble Kalman filter was used to estimate soil moisture, related states, and fluxes by merging noisy low-frequency microwave observations with the forecasts from a conventional yet uncertain land surface model. Kalman filter estimates are only conditioned on observations prior to estimation times. Here it is argued that soil moisture estimation is a reanalysis-type problem as observations beyond the estimation time are useful in the estimation. An ensemble smoother is used in which the state vector and measurement vector are distributed in time and updated as a batch. Its performance in a land data assimilation context is compared to that of the ensemble Kalman filter. Results demonstrate that smoothing yields an improved estimate compared to filtering, reflected in the decreased deviation from truth and the reduction in uncertainty associated with the estimate. Precipitation significantly impacts the performance of the smoother, acting as an information barrier between dry-down events. An adaptive hybrid filter/smoother is presented in which brightness temperature is used to break the study interval into a series of dry-down events. The smoother is used on dry-down events, and the filter is used when precipitation is evident between estimation times. An improved estimate is obtained as all observations in a given dry-down period are used to estimate soil moisture in that period, and backward propagation of information from subsequent precipitation events is avoided.

1. Introduction

1.1. Soil Moisture in the Climate System

[2] Surface soil moisture is a key state variable which integrates much of the land surface hydrology and exerts considerable control on several land-atmosphere exchanges. It is the fastest component of the continental water cycle with a residence time of just a few days. Root zone (5–100 cm) soil moisture determines how much water is available to vegetation, thereby influencing the latent heat flux and hence the surface energy balance.

[3] A consistent data set of soil moisture, ground temperature and surface fluxes would enable a detailed study of land-atmosphere interactions and the role that they play in the climatic system. Global or regional in situ measurements at the scales required to study hydrometeorology (10 km) and hydroclimatology (30–50 km) would require networks that are logistically and economically infeasible. Remote sensing, on the other hand, is ideal for obtaining data at these scales and globally.

1.2. Remote Sensing of Soil Moisture

[4] Passive microwave radiometry has long been recognized as having the potential to measure soil moisture on regional and global scales. Low-frequency passive microwave radiation (1–3 GHz or L band) is particularly suitable as there is a sharp contrast in the dielectric constants for water and soil in this region of the spectrum [*Ulaby et al.*, 1986; *Wang and Schmugge*, 1980]. Furthermore, measurements in this frequency range are relatively unaffected by clouds and can penetrate light to moderate vegetation.

[5] As early as the late 1960s and early 1970s small studies were undertaken to determine the feasibility of using microwave brightness temperatures to estimate soil moisture. L band remote sensing of soil moisture can be used to estimate volumetric water content in the top 5 cm of the soil column with a precision of a few percent [*Jackson et al.*, 1995; *Jackson*, 1997; *Jackson et al.*, 1999]. Future pathfinder missions such as NASA's Hydrosphere State (Hydros) and ESA's Soil Moisture and Ocean Salinity (SMOS) will provide global L band observations from which a global soil moisture data set can be obtained [*Entekhabi et al.*, 2004].

1.3. Land Data Assimilation

[6] While remote sensing offers the advantage of global coverage, the temporal resolution of observations is limited by the revisit time. The Hydros satellite will revisit a given location on the Earth's surface just once every 2–3 days. Furthermore, the L band brightness temperature relates to the soil moisture at the surface (top 5 cm) and yields no

Table 1. Advantages and Disadvantages of Ensemble-Based Filters Compared to Those of Variational Techniques

	Ensemble-Based Filters	Variational Techniques
Advantages	Any model can be used. Model does not need to be differentiable. Noise can be placed anywhere, for example, on uncertain parameters and forcing. Noise can be non-Gaussian and nonadditive.	Uses all data in a batch window to estimate the state.
Disadvantages	Estimates are conditioned on past measurements only.	Model must be differentiable to obtain tangent-linear model. Process noise can only be additive and Gaussian. Changes to model require that adjoint be obtained again.

information on the root zone. Forcing a land surface model with meteorological data can produce soil moisture and temperature estimates, along with the associated fluxes at the temporal resolution of the model yielding information on the diurnal cycle. However, such unconstrained simulations are subject to the errors in model structure and forcing uncertainty. Data assimilation offers a means to combine the advantages of modeling with those of remote sensing.

[7] Data assimilation techniques have been used in meteorology and oceanography for decades. A comparison of the various techniques is provided by *Ghil and Manalotte-Rizzoli* [1991]. *Courtier et al.* [1993] compiled a list of significant papers in the application of data assimilation techniques to meteorology problems. Data assimilation techniques can be roughly divided into two categories; variational techniques and those derived from the classic Kalman filter. Both methods have been applied to hydrological research in recent years.

[8] The central concept in variational data assimilation is the adjoint model. This is obtained by linearizing the forward model along a trajectory producing the tangent-linear model, and obtaining the adjoint. Thus variational techniques require that the model be differentiable. *Lorenc* [1986] describes various variational techniques which have been applied in meteorology. Several applications in oceanography and meteorology are discussed by *Ghil and Manalotte-Rizzoli* [1991] and *Wunsch* [1996]. Variational techniques have been successfully applied to hydrological applications in recent years [*Castelli et al.*, 1999; *Boni et al.*, 2001; *Reichle*, 2000; *Reichle et al.*, 2001a, 2001b; *Margulis*, 2002]. 4DVAR, in which observations distributed in space and time are used with knowledge of temporal evolution of the state, is particularly suited to our problem as demonstrated by *Reichle* [2000], but it requires development of the adjoint. While automatic adjoint compilers are available [*Giering and Kaminski*, 1998], they can prove difficult to use with large and intricate numerical models, and typically involve extensive tuning and sensitivity studies to validate the adjoint model generated. A means is sought by which temporally distributed observations may be used in a smoothing approach like 4DVAR without resorting to a simplified land surface model.

[9] The classic Kalman filter as discussed by *Gelb* [1974] provides the optimal state estimate for linear systems. It is therefore of little use in hydrological applications where the

physical model equations are often nonlinear and contain thresholds. In the extended Kalman filter for nonlinear systems [*Gelb*, 1974; *Jazwinski*, 1970], approximate expressions are found for the propagation of the conditional mean and its associated covariance matrix. The structure of the propagation equations is similar to those of the classic Kalman filter for a linear system, as they are linearized about the conditional mean. The extended Kalman filter has been successfully applied to the land data assimilation problem [*Entekhabi et al.*, 1994; *Galantowicz et al.*, 1999; *Walker et al.*, 2001; *Walker and Houser*, 2001; *Crosson et al.*, 2002], but its use in this application would require derivation of a tangent linear model to approximate the land surface model, as well as techniques to treat the instabilities which might arise from such an approximation. *Ljung* [1979] performed a convergence analysis of the extended Kalman filter and demonstrated the potential for divergence or bias in estimates in nonlinear systems. *Nakamura et al.* [1994] encountered such instability in their application of the extended Kalman filter to soil moisture estimation.

[10] An alternative sequential estimation technique for nonlinear problems was proposed by *Evensen* [1994]. In the ensemble Kalman filter (EnKF) an ensemble of model states is integrated forward in time using the nonlinear forward model with replicates of system noise. At update times, the error covariance is calculated from the ensemble. The traditional update equation from the classical Kalman filter is used, with the Kalman gain calculated from the error covariances provided by the ensemble. The EnKF has been successfully implemented by *Evensen and Van Leeuwen* [1996], *Houtekamer and Mitchell* [1998], and *Houtekamer and Mitchell* [2001] and has already been used to merge L band observations with model output to estimate soil moisture [*Reichle et al.*, 2002; *Margulis et al.*, 2002; *Crow*, 2003; *Crow and Wood*, 2003]. Research in ensemble techniques has yielded innovative methods of improving estimates and reducing the computational burden [*Segers et al.*, 2000; *Heemink et al.*, 2001; *Verlaan*, 1998; *Verlaan and Heemink*, 2001]. The advantages and disadvantages of the EnKF are compared to those of variational techniques in Table 1.

[11] In the past, soil moisture observations have typically been gathered during field experiments such as the Southern Great Plains Field Experiments (SGP97 and SGP99) and Soil Moisture Experiments in 2002 (SMEX02) and 2003 (SMEX03). Smoothing is ideal for analyzing historic data or

data which are not available in real time, as is the case with data from field experiments or exploratory missions such as Hydros and SMOS. Smoothing involves using all measurements in an interval $\mathbf{T} = [0, T]$, to estimate the state of the system at some time t where $0 \leq t \leq T$, so that the state estimate at a given time is determined by including information from subsequent observations. It will be argued that an ensemble-based smoothing (or batch estimation) approach is most suited to the soil moisture estimation problem.

[12] Results from the EnKF experiment [*Margulis et al.*, 2002] suggest that the estimate could be improved through the inclusion of subsequent observations. Precipitation events divide the study interval into a series of dry-down events. In estimating soil moisture at a given time, one is estimating a single point value in a series. It is intuitive that the manner in which that series evolves in the future is related to the state at the estimation time. Future observations provide information on the shape of this series in the future and so contain useful information on the current state. Correlation between the states and the observations decreases with depth as the observations relate to the surface conditions. Consequently the impact of the observations is lessened with increasing depth. This means that it takes longer to correct for spurious initial conditions at depth than close to the surface. As the impact of the observations eventually penetrates the deeper layers, the latent heat flux estimate is seen to approach the observed values. Difficulty in estimating the root zone soil moisture results in poor initial estimates of the latent heat flux [*Margulis et al.*, 2002]. If including subsequent observations can improve on the initial conditions at depth, it would result in improved latent heat flux estimates.

[13] In the following section an ensemble-based smoother will be developed as an extension of the conventional EnKF which, by including information on how the state evolves beyond the estimation time, should yield improved estimates of the soil moisture at the surface and at depth.

2. Ensemble Smoother Algorithm

[14] Several ensemble smoothers exist in data assimilation literature, for example, the ensemble smoother (ES) of *Van Leeuwen and Evensen* [1996] and the ensemble Kalman smoother (EnKS) of *Evensen and Van Leeuwen* [2000]. They have been used in various applications such as ocean forecasting [*Brusdal et al.*, 2003; *Van Leeuwen*, 2001; *Van Leeuwen and Evensen*, 1996] and fish stock assessment [*Gronnevik and Evensen*, 2001] and the objective of this paper is to determine their applicability to soil moisture estimation. The simplest smoother is an extension of ensemble Kalman filtering in which the state and measurement vectors are distributed in time, and the augmented state vector is updated using the traditional ensemble Kalman filter equations. Its performance is compared to the EnKF to determine if an improved estimate of soil moisture can be obtained with ensemble smoothing and to identify issues which may be significant in the implementation of an ensemble smoother in a land data assimilation framework.

2.1. Ensemble Kalman Filter Equations

[15] In the EnKF an ensemble of model states, y(t) is integrated forward in time using the full nonlinear model, A[·].

$$y(t) = A[y(\tau), \alpha, u(\tau), w(t), t, \tau]$$
$$y(t_0) = y_0 \tag{1}$$

The state at time t depends on the state at a previous time τ, the time invariant parameters α of the model, the forcing applied to the model $u(\tau)$ and system error $w(t)$. Here y(t) contains the soil moisture in six layers of the soil column and A[·] is the NOAH land surface model (LSM) [*Chen et al.*, 1996]. Time-invariant parameters include descriptors of the soil texture and vegetation cover. The model is initialized with random initial conditions y_0. The observations z are related to the state y through the measurement operator M[·] and have additive Gaussian error $\omega(t)$.

$$z = M[y(t)] + \omega \tag{2}$$

Here the radiative transfer model (RTM) is the measurement operator, relating the volumetric soil moisture values in y(t) to the observed L band brightness temperature (section 3.2). The EnKF is a sequential processor, updating the state through (3) when observations become available. Each ensemble member $y^j(-)$ is updated individually using the Kalman gain, K, which is calculated from the ensemble statistics in (4).

$$y^j(+) = y^j(-) + K\{z + \omega^j - M[y^j(-)]\} \tag{3}$$

$$K = C_{yz}(C_{zz} + C_\nu)^{-1} \tag{4}$$

$$C_{yz} = \frac{1}{N_R}\tilde{\mathbf{y}}(-)\tilde{\mathbf{z}}^T \tag{5}$$

$$C_{zz} = \frac{1}{N_R}\tilde{\mathbf{z}}\tilde{\mathbf{z}}^T \tag{6}$$

K weighs the relative uncertainty in the modeled estimate to that associated with the observation. C_{yz} is the cross covariance between the prior state and its transformed value in observations space, C_{zz} is the covariance of the transformed prior states in observation space and C_ν is the known variance of the observations (here $C_\nu = (3K)^2$ for L band observations). In equations (5) and (6), () denotes the perturbation matrix, N_Y is the number of states(6), N_Z is the number of observations (1) and N_R is the number of ensemble members (2000). For each ensemble member random noise ω^j is added to the observation z to account for the contribution of observation error to the posterior covariance [*Burgers et al.*, 1998]. In the soil moisture estimation problem, the model is highly nonlinear, and uncertainty in parameters and forcing can result in non-Gaussian distributions of the states. By updating each ensemble member individually, this algorithm is particularly
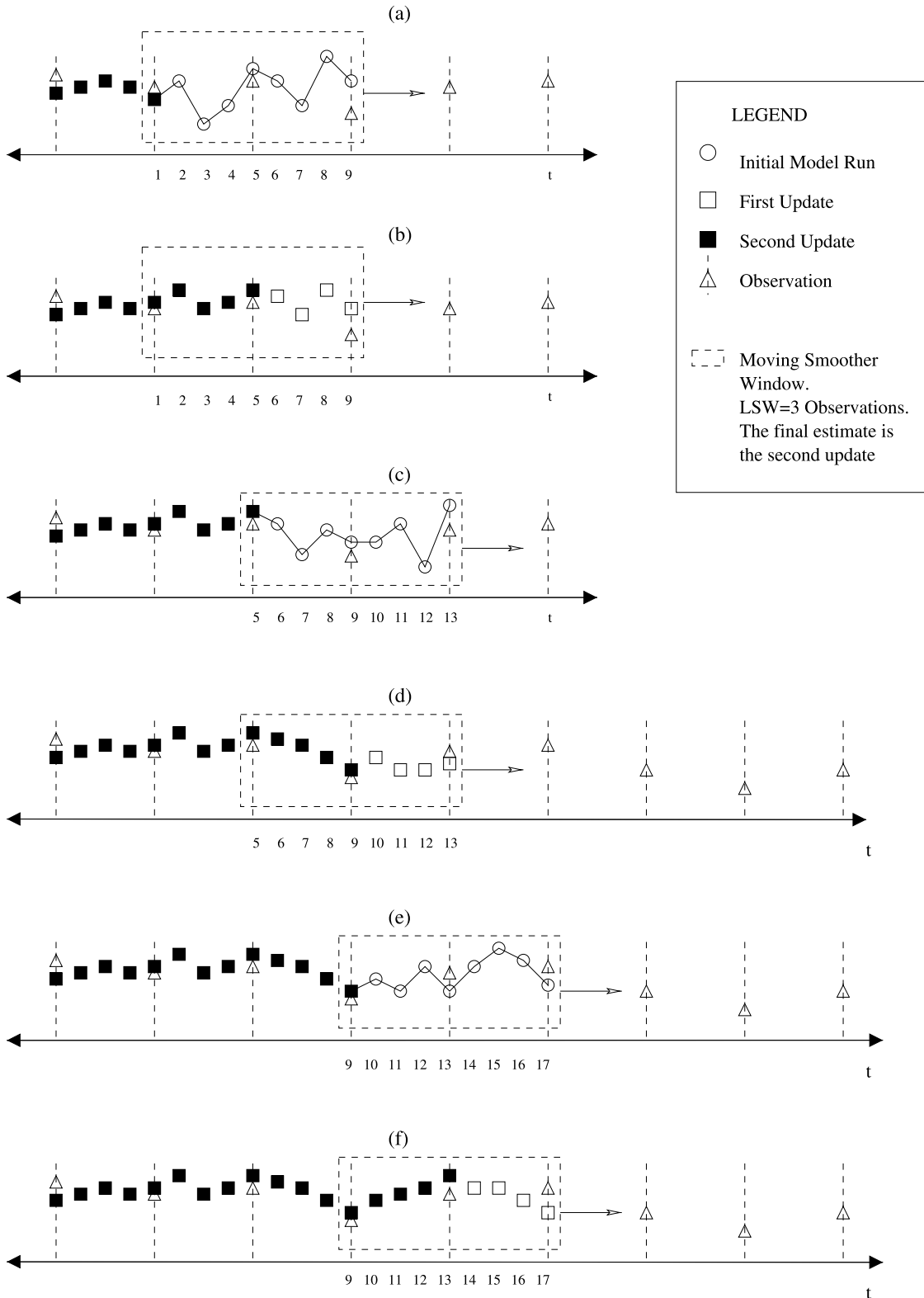
Figure 1. Conceptual diagram of ensemble moving batch smoother algorithm. An estimate of the state is required at every time step, while observations are available at every fourth time step. The length of the smoother window (CSW) in this example is three as the smoother window encompasses three observations.

advantageous as it does not force a Gaussian posterior distribution. A thorough description of the ensemble Kalman filter and its implementation is provided by *Evensen* [2003].

2.2. Ensemble Moving Batch Smoother

[16] The EnKF described above has been used to estimate soil moisture during SGP97 [*Margulis et al.*, 2002]. The smoothing algorithm used here is a simple extension of the

EnKF in which the states are distributed in time and updated in a "batch". The number of observations included determines the length of the observation vector, the state vector and consequently the covariance matrices. Including observations too far into the future would increase the computational burden without adding any useful information. Fortuitously, the memory in soil moisture is limited by the occurrence of precipitation which disrupts the dry-down and effectively reinitializes the problem.

[17] In the conceptual diagram in Figure 1 the batch contains three observations. The smoother window refers to the interval between the first and last observation. The forward model is run through to the end of the smoother window to obtain the prior estimate of the state. An augmented state vector Y contains the states of interest (y) at all time steps of interest, which may include times at which the state is not observed.

$$Y = \begin{bmatrix} y_1 & y_2 & ... & y_9 \end{bmatrix}^T \qquad (7)$$

The augmented measurement vector Z contains all the observations in the smoother window:

$$Z = \begin{bmatrix} z_1 & z_5 & z_9 \end{bmatrix}^T \qquad (8)$$

The EnKF equations in the previous section are applied to these augmented vectors to yield an updated estimate. When the EnKF equations are implemented for a batch of observations, the covariance matrices relate the state at multiple times to all observations in the batch. The Kalman Gain matrix reflects the relevance of future observations to the current state. The smoother window is moved along the study interval one observation at a time, as including a new observation introduces new information.

[18] Here observations will be available every 3 days, while an estimate of soil moisture is desired four times daily, based on the data assimilation product requirement of the Hydros mission [Entekhabi et al., 2004]. The batch includes just two observations, to demonstrate that the inclusion of any information on the future state would yield an improved estimate. Consequently, the state vector will consist of the volumetric soil moisture in six layers at 12 time steps, and the measurement vector will contain two brightness temperatures.

[19] Computational burden is a concern in employing batch smoothing techniques and ensemble techniques. As the length of the augmented vectors grow, larger memory will be required to make estimates conditioned on all measurements in the batch window. A concern is that including spatial correlation would increase the computational burden indefinitely. However, estimation variables can be a combination of model states. The standard Hydros data product is $0-5$ cm and $5-100$ cm soil moisture, so the dimension of the state vector can be significantly reduced even though the land surface model may have more layers for computational stability. There are computationally more efficient ways of implementing the ensemble smoother [Evensen, 2003]. In subsequent studies where the spatial dimension is added to the problem, the ensemble smoother

for land data assimilation will take advantage of the improved implementation.

3. Data Assimilation Framework

[20] Here the ensemble moving batch (EnMB) smoothing algorithm was evaluated using data from the Southern Great Plains Experiment 1997 (SGP97) to facilitate comparison with results from Margulis et al. [2002]. Experiments focused on two points in the SGP97 domain, namely, El Reno and Little Washita.

[21] Data from the Oklahoma Mesonet were used to create an Observing System Simulation Experiment (OSSE). The land surface model (section 3.1) was forced using meteorological data to create a synthetic truth. Synthetic observations were generated from this "truth" using the radiative transfer model (section 3.2). Additive zero-mean Gaussian noise with standard deviation of 3K was added to the synthetic observations to account for observation error. Using synthetic rather than real observations offers the following advantages.

[22] 1. The estimation technique can be evaluated since the synthetic truth is known. Furthermore, this obviates the need to compare the estimate from data assimilation to ground observations which are prone to added sampling error.

[23] 2. The availability of observations is not constrained by adverse weather or instrument troubles.

[24] 3. Observations can be made at any time. Here they were taken at 6am every 3 days to simulate the revisit time of the Hydros mission.

[25] 4. Meteorological data from the Oklahoma Mesonet was used to generate the truth from the land surface model, so the experiment duration could be extended to run from 1st May to 1st September 1997.

3.1. Model

[26] The NOAH Land Surface Model is used to propagate the ensemble of states forward between observations. This 1-D model of the soil column provides estimates of soil moisture and temperature profiles in addition to the mass and energy terms of the surface water and energy balances. It is a widely used and freely available community land surface model which has been extensively validated and is currently used in the NASA land data assimilation system [Lohmann et al., 2004]. It was used by Margulis et al. [2002] to estimate soil moisture using the EnKF.

3.2. Radiative Transfer Model

[27] A Radiative Transfer Model (RTM) is required to transform the states from state-space to observation space. The RTM used here is identical to that used by Margulis et al. [2002]. It is based on the retrieval algorithm used by Jackson et al. [1999] to retrieve soil moisture from ESTAR observations during SGP97 but using the mixing model of Wang and Schmugge [1980]. Surface roughness and vegetation effects are also accounted for [Choudhury et al., 1979; Jackson and Schmugge, 1991].

3.3. Model Error and Uncertainty

[28] Model error was implicitly added to the data assimilation framework by allowing key parameters to assume different values in an expected range for each ensemble member. Uncertainty was imposed on four key soil and
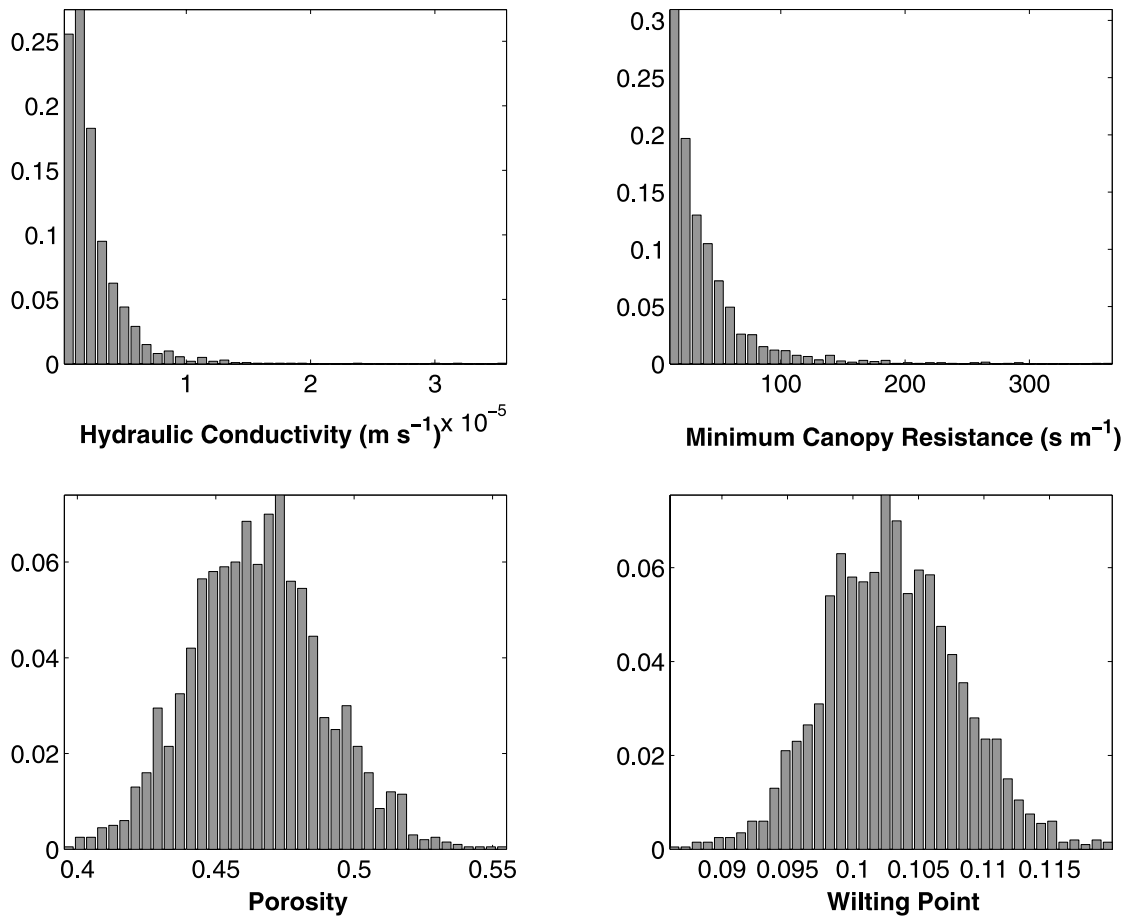
Figure 2. (top left) Relative frequency distribution of saturated hydraulic conductivity, (top right) minimum stomatal resistance, (bottom left) porosity, and (bottom right) wilting point at the El Reno pixel.

vegetation parameters, namely the saturated hydraulic conductivity, the minimum canopy resistance, the porosity and the wilting point. Varying the saturated hydraulic conductivity effectively varies the rate at which water can move through the soil column. Allowing the porosity and wilting point to vary means that each replicate has a distinct possible range of soil moisture values. Each replicate having a different minimum canopy resistance means that the rate of evaporation will be different for each ensemble member.

[29] The values for these parameters afforded by the model based on land class or soil class were used as nominal values. The time-invariant parameter value for each ensemble member consists of the nominal value multiplied by a random variable of mean one and a coefficient of variation of 1.0 for both the saturated hydraulic conductivity and minimum canopy resistance, and 0.05 for the porosity and wilting point. Lognormal multiplicative Gaussian noise was added to yield a large range of values while ensuring that negative values did not occur. The relative frequency distributions of the parameters for the El Reno pixel are shown in Figure 2.

[30] Uncertainty was also included in the initial condition. Nominal relative saturation at the surface was set to 0.5, with the nominal values at depth determined by assuming a hydrostatic profile. Uncertainty was including by adding Gaussian noise of mean 0.0 and standard deviation decaying exponentially with depth from 0.2 at the surface.

[31] Unpublished experiments found that the most effective way to introduce ensemble spread is through uncertainty in precipitation. Further discussion of uncertainty in precipitation is included in sections 4 and 5.

3.4. Algorithm Evaluation

[32] The ensemble open loop (EnOL) provides the model estimate and associated model error in the absence of data assimilation, a valuable benchmark by which to measure the improvement after filtering or smoothing. To evaluate ensemble algorithms, the quantities of interest are the ensemble mean, which will be compared to the "true" state, and the standard deviation across the ensemble. Two summary statistics will also be used to assess the smoother algorithms performance relative to the EnKF and EnOL.

[33] 1. The root-mean-square error (RMSE) provides an average measure of the deviation of the ensemble mean from the true state over all estimation times. Clearly, the data assimilation algorithm is performing well if the ensemble mean is close to the truth.

[34] 2. The estimation error standard deviation (EESD) is the average standard deviation across the ensemble calculated over all estimation times. The ensemble spread is a

Table 2. Rectangular Pulses Model Parameters for Oklahoma City

Month	$E[t_r]$	$E[i_r]$	$E[t_B]$	$E[t_B]'$
May	7	2.616	83	91
June	5	3.220	92	60
July	8	2.264	137	130
August	4	2.374	116	62

measure of the confidence which should be placed in the estimate.

[35] Observations were available every 3 days, and estimates were required four times daily at 6am, 2pm, 6pm and 12am. For the four month experiment duration, this yielded a sample of 493 estimation times with which to calculate the RMSE and EESD.

4. Experiment 1: Precipitation Forcing Derived From Monthly Total Information

[36] In a global land data assimilation application, precipitation data will likely be derived from satellite data such as Global Precipitation Climatology Project (GPCP). Daily, pentad and monthly total precipitation products are available from GPCP. Daily totals provide higher-frequency information than pentad or monthly observations but due to temporal sampling and algorithm uncertainty the monthly total is more reliable. This temporal resolution is too coarse to characterize storm events for the purposes of land surface modeling which requires hourly data or better. The spatial resolution of observations (2.5° × 2.5°) is orders of magnitude greater than that of the estimation pixel (typically kilometers), so information on spatial variability of precipitation is lost. Use of such data requires spatial and temporal disaggregation to the resolution of the model. Consequently,

use of satellite-based data implies uncertainty in the timing, amount, and spatial distribution of precipitation.

4.1. Ensemble Precipitation Using the Rectangular Pulses Model to Disaggregate the Monthly Total

[37] The objective is to generate an ensemble of precipitation forcing which is constrained only by the monthly total precipitation. Using the rectangular pulses model (RPM) of *Rodriguez-Iturbe et al.* [1984], it is assumed for each ensemble member that precipitation occurs as distinct rectangular pulses with random parameters. The expected arrival time, duration and intensity of a storm are exponentially distributed with mean values $E[t_B]$, $E[t_r]$ and $E[i_r]$ respectively.

[38] Using historical meteorological data, *Hawk and Eagleson* [1992] derived these climatological parameters for many stations across the United States. The Hawk and Eagleson parameters for the months of interest are shown in Table 2. The method of *Margulis and Entekhabi* [2001] is used here to derive a modified $E[t_B]$, $E[t_B]'$ which takes into account the observed monthly precipitation. The total monthly precipitation was derived from Oklahoma Mesonet precipitation records at El Reno. Using these "monthly observations," $E[t_B]'$ was calculated for the four months of interest in 1997 (Table 2). Further value can be derived from the monthly measured rainfall, by using it to discriminate between realizations. Here realizations were rejected if they were beyond 25% of the total observed precipitation at the end of the four month period.

4.2. Surface Soil Moisture at El Reno

[39] Figure 3 compares the estimated surface soil moisture from the EnOL, EnKF, and EnMB to the truth. In the absence of information on the timing and magnitude of
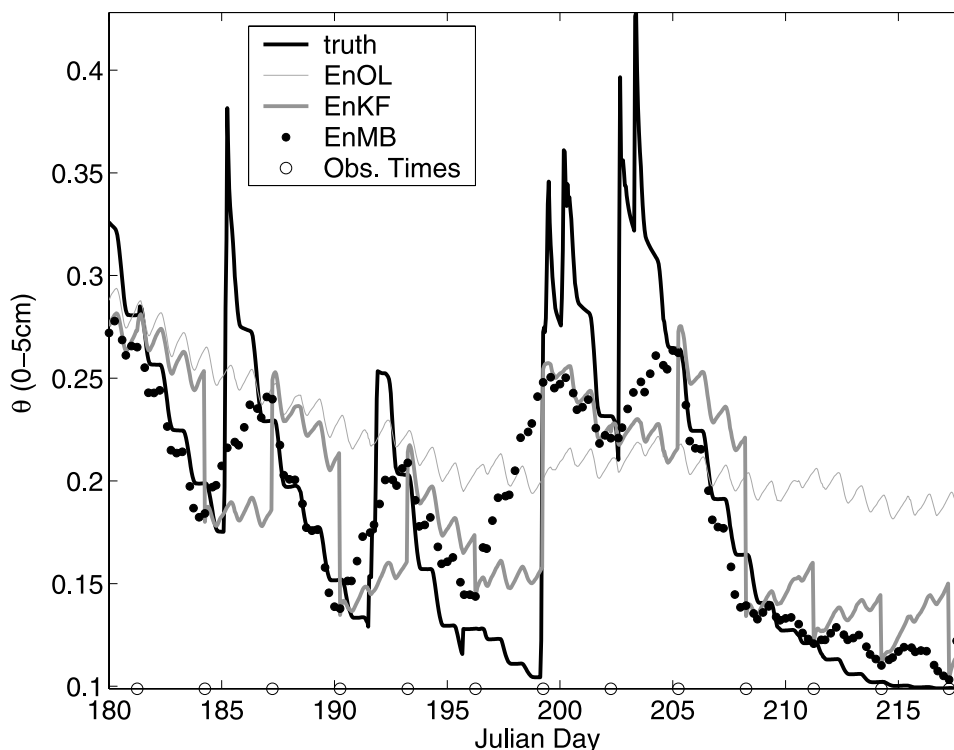

Figure 3. Ensemble mean volumetric soil moisture (θ) in the top 5 cm of the soil column at El Reno compared to the synthetic truth. Results are shown for the period between Julian days 180 and 218.
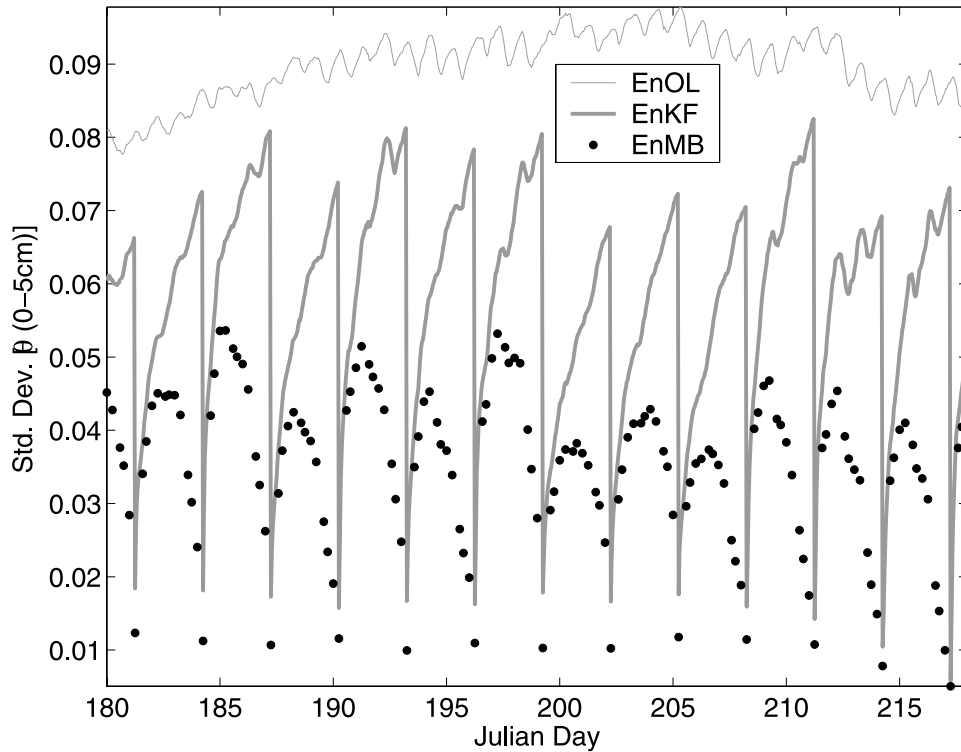
Figure 4. Estimation error standard deviation (EESD) in the estimate of surface (0–5 cm) volumetric soil moisture (θ) at El Reno for the period between Julian days 180 and 218.

precipitation events, the EnOL soil moisture is distributed across the dynamic range. Both the EnKF and EnMB are drawn toward to the truth at observation times. While the EnKF drifts uncorrected toward the EnOL between observations, the backward propagation of subsequent observations yields a smooth transition between observations in the EnMB. This is particularly advantageous during dry-down periods (e.g., Julian days 205 to 219).
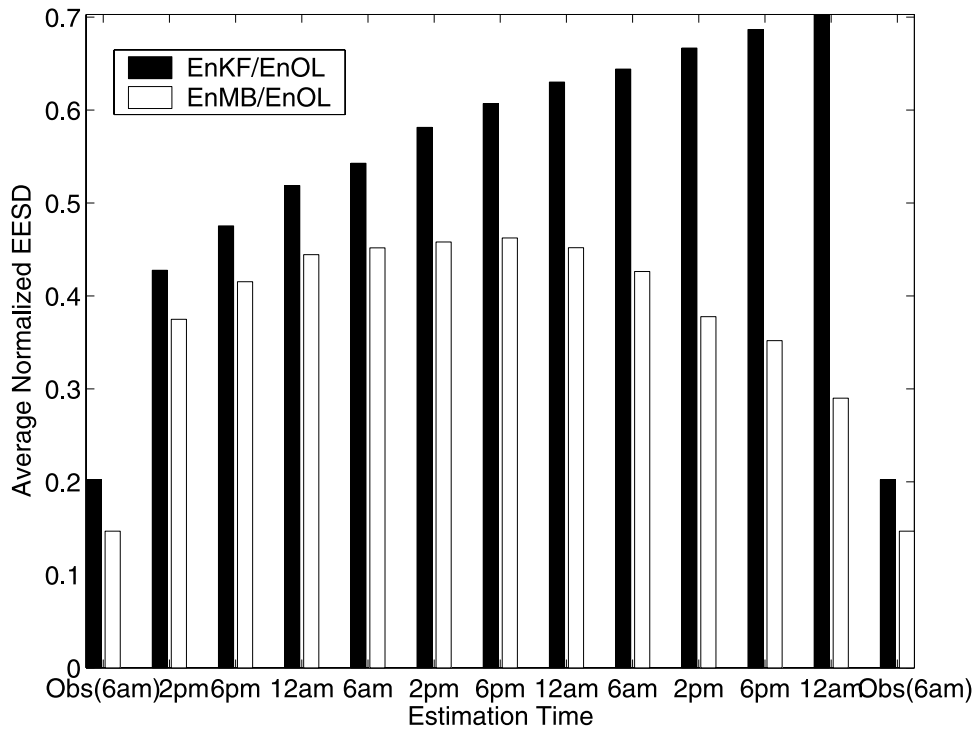
Figure 5. Average normalized EESD in surface volumetric soil moisture as a function of timing within the interobservation period at El Reno.
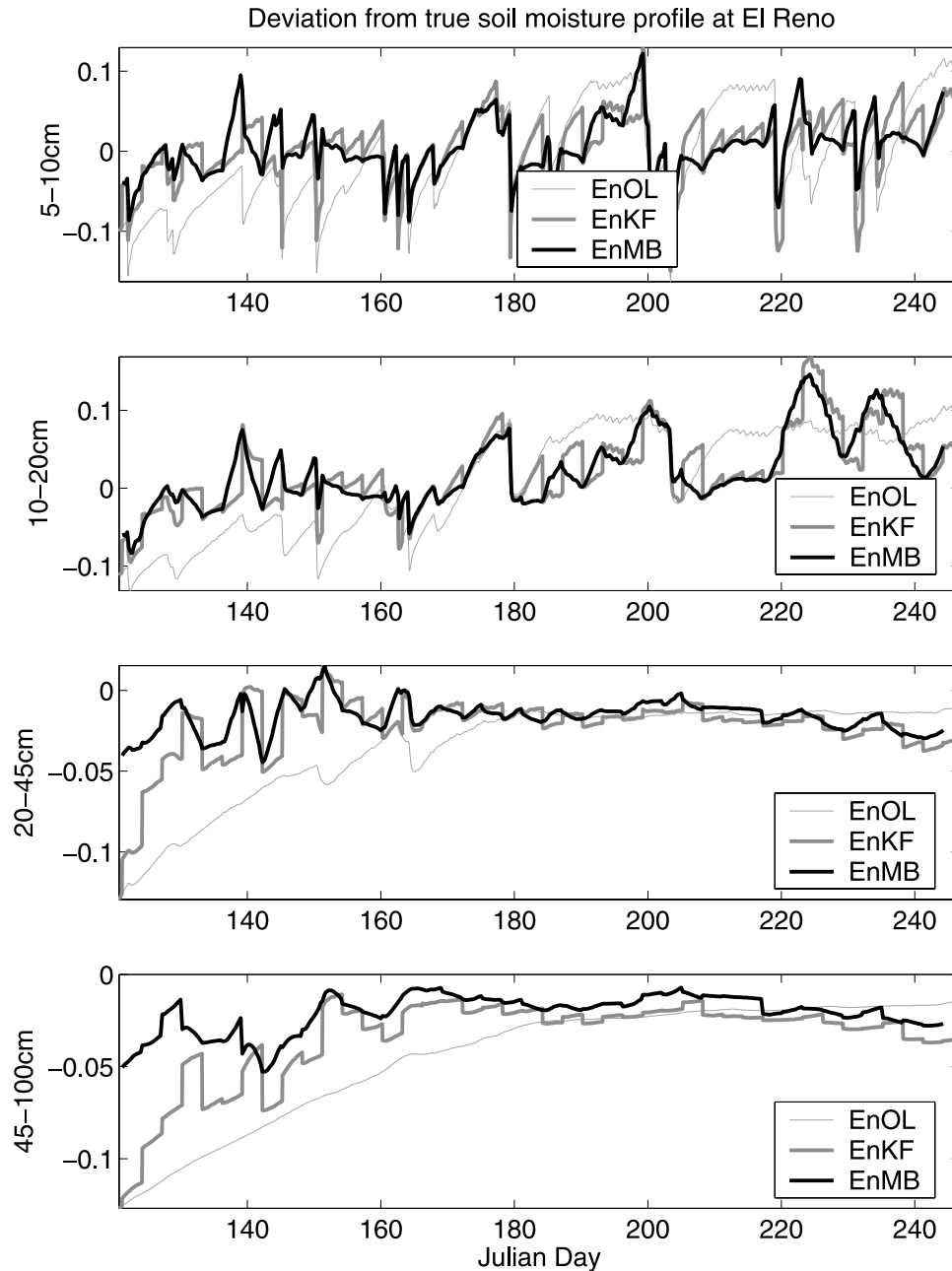
Figure 6. Deviation from "true" soil moisture at El Reno is shown at various depths. The smoothed estimate (EnMB) is compared to the filtered estimate (EnKF) and the ensemble open loop (EnOL).

[40] The relative timing of observations and precipitation significantly impacts the performance of the smoother. Backward propagation of the increased soil moisture following a storm results in spuriously moist estimates in the EnMB. The effect is most detrimental if the observation immediately precedes an observation (days 196–199), and less harmful if the precipitation is early in the interval (days 184–187 and 202–205). As smoothing is most effective on dry-down curves, it would be useful if we could identify dry-down curves over which to smooth. This issue is discussed further in section 6.

[41] Figure 4 shows the reduced EESD obtained from the EnMB compared to the EnKF and EnOL. The EESD in the EnOL is relatively constant at 0.09, about 25% of the dynamic range of soil moisture. In the filter case, the

ensemble spread exhibits a characteristic sawtooth shape, growing rapidly between observations. The symmetry in the EnMB standard deviation indicates that the backward propagation of information through the covariance matrix is improving the estimate. The reduced standard deviation indicates that we should have increased confidence in the smoothed result compared to the filter.

[42] Figure 5 shows the reduction in ensemble spread after filtering/smoothing as a function of timing within the 3-day interobservation period. At each estimation time the EESD for the smoother and filter were normalized by that of the open loop. The results were then averaged for each point in the interobservation period. The EESD in the ensemble filter grew to 0.7 times that of the open loop case as observations were available every 3 days. Short-
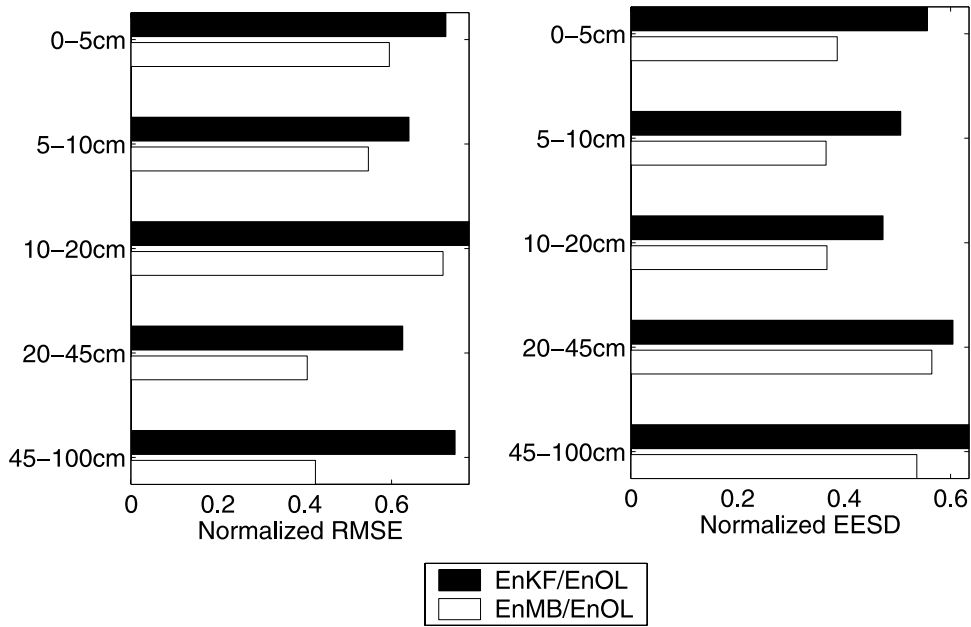
Figure 7. Normalized RMSE and average normalized EESD of volumetric soil moisture at depth (θ) at El Reno.

ening/lengthening the observation interval would reduce/increase this value. The maximum ratio in EESD between the smoother and the open loop is around 0.45, two thirds of the maximum from the filter. The greatest improvement due to smoothing is immediately prior to the later obser-

vation. The correlation between states and the future observation is highest immediately prior to the observation and is diminished as the difference between the estimation time and the future observation increases. This is counterbalanced by the fact that EESD is at a minimum at the
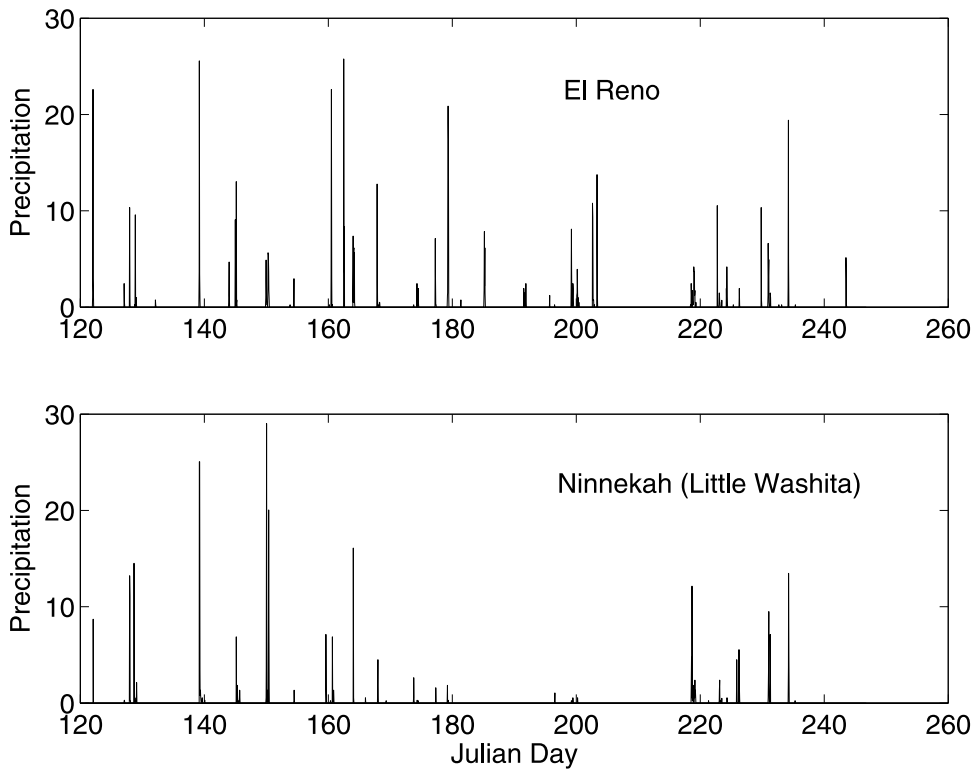

Figure 8. Observed precipitation (in mm h^{-1}) time series for El Reno and Ninnekah (Little Washita). This illustrates the difference in the timing and quantity of precipitation at the two stations.
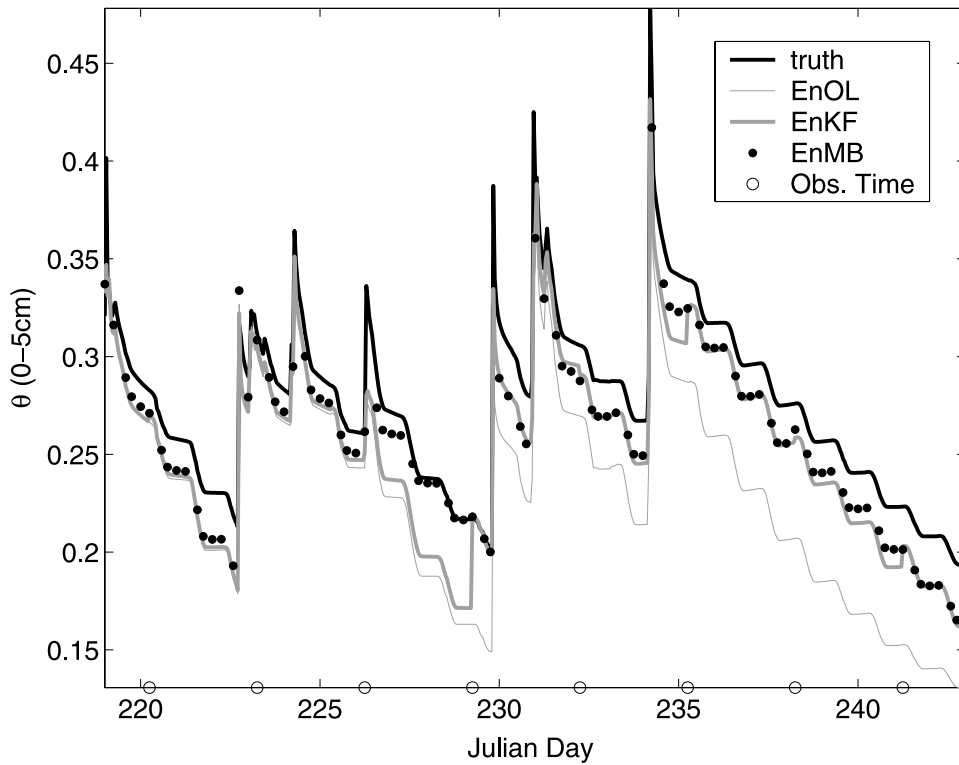
Figure 9. Ensemble mean volumetric soil moisture (θ) in the top 5 cm of the soil column at El Reno compared to the synthetic truth. Results are shown for the period between Julian days 219 and 243.

observation time and grows with time. The combination of the two effects is a symmetric rather than sawtooth evolution of the EESD between observations in the smoothed case.

4.3. Subsurface Soil Moisture Estimation at El Reno

[43] Figure 6 shows the deviation between the estimated and true soil moisture at depth. With increasing depth the length of time required by the EnOL to recover from
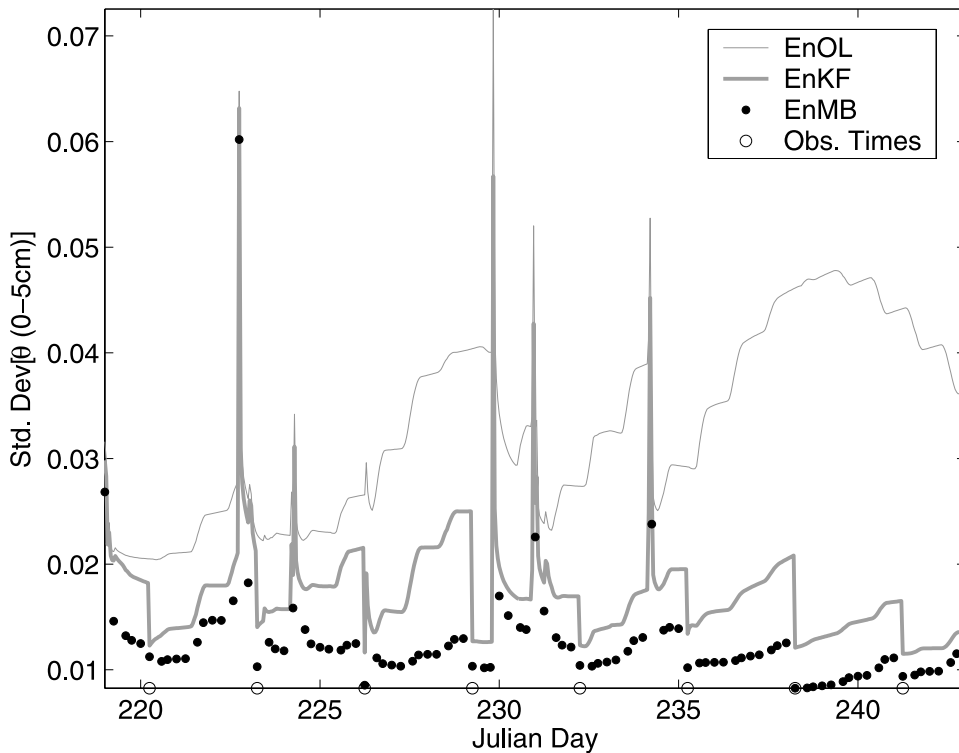

Figure 10. Estimation error standard deviation in the estimate of surface (0−5 cm) volumetric soil moisture (θ) at El Reno. Results are shown for the period between Julian days 219 and 243.
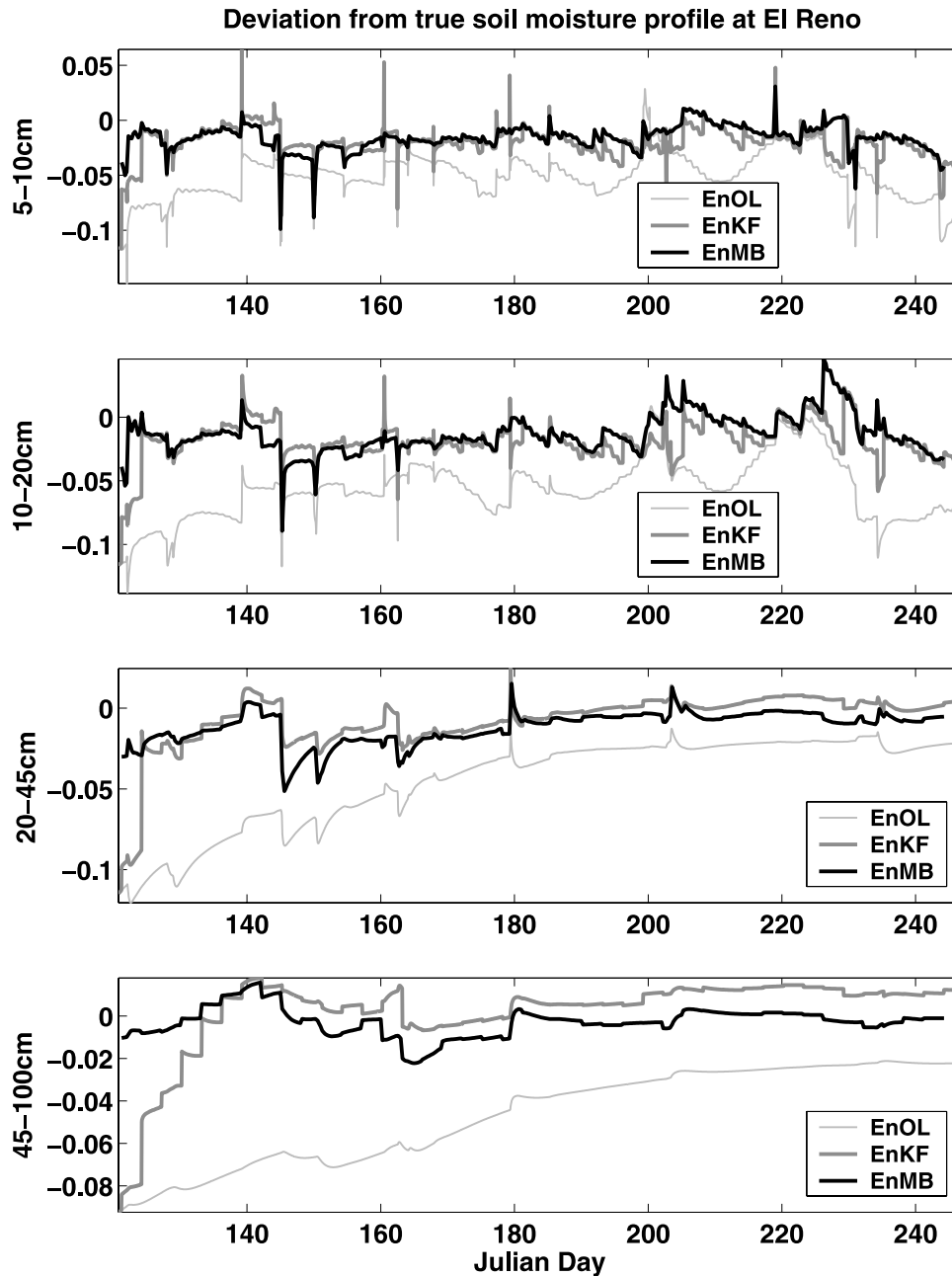
Figure 11. Deviation from "true" soil moisture at El Reno shown for layers 2–5. The results from the moving batch smoother (EnMB) are compared to that of the EnKF and the ensemble open loop (EnOL).

spurious initial conditions increases. While the EnKF improves on the open loop, the EnMB reduces the deviation by over 50% as soon as the first observations become available. In the deepest layer, the EnKF takes 20–30 days to catch up with the smoother.

4.4. Summary Statistics at El Reno

[44] Figure 7 demonstrates that smoothing improves over filtering at all depths in terms of both RMSE and EESD. The filtered estimate is quickly drawn toward the EnOL between observations, limiting the reduction in RMSE to 25% at the surface. The smoother leads to a further 20% improvement over the filter. At depth, smoothing alleviates the impact of initial conditions much faster than the filter. In layers 4 and 5 (20–45 cm and

45–100 cm, respectively), the RMSE is close to half that of the filter.

[45] There is almost a 50% reduction in average EESD due to the filter compared to the EnOL. There is a further 33% reduction at the surface when the EnMB is employed. Although ensemble growth is slower at depth due to the dampened response to incident precipitation, there is a persistent reduction in EESD due to smoothing.

5. Experiment 2: Precipitation Forcing From Rain Gauge Data

5.1. Derivation of Precipitation Forcing Data

[46] The objective of this experiment is to evaluate the performance of the EnMB in a data assimilation framework
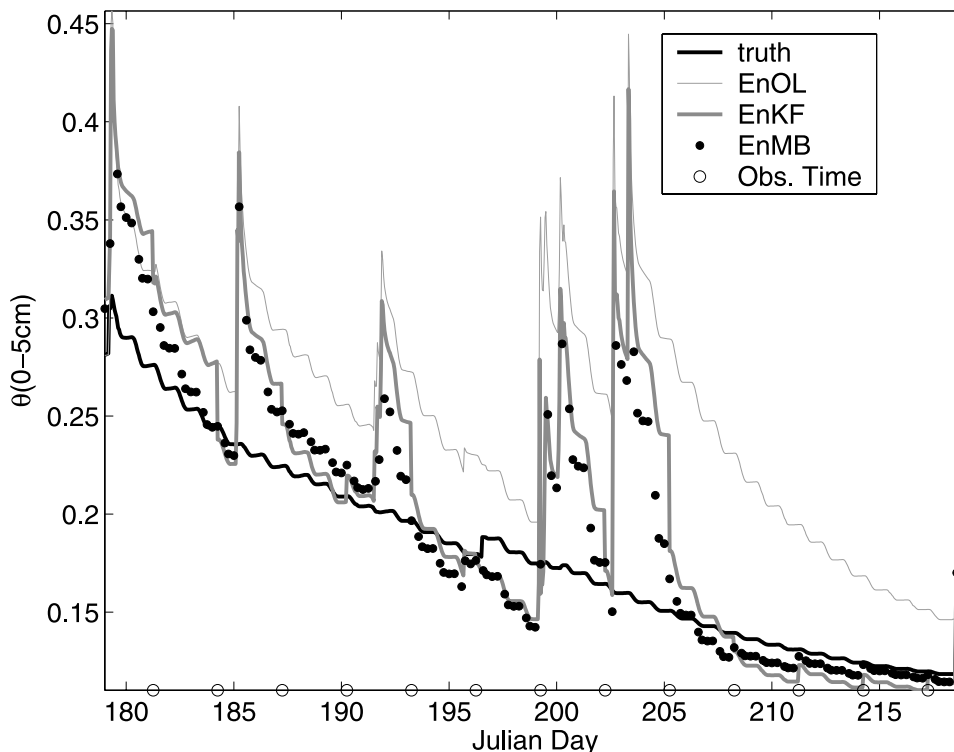
Figure 12. Ensemble mean volumetric soil moisture θ in the top 5 cm of the soil column at Little Washita is compared to the synthetic truth for the interval from Julian day 179 to 219.

where precipitation data are from rain gauges. While gauge data are a useful indicator of when precipitation occurs, the amount is uncertain as the measurement is at a point and is prone to errors due to spatial variability and underreach. An ensemble of precipitation forcing was generated to reflect this uncertainty. Nominal precipitation was multiplied by a lognormally distributed random factor of mean 1.0 and standard deviation set equal to 50% of the nominal precipitation. The performance of the EnMB was evaluated at two locations.

[47] 1. At El Reno (the gauge location) the timing of precipitation is known. A single realization of the precipitation forcing was used to generate "truth."

[48] 2. At Little Washita it was assumed that the best available data is that recorded at El Reno. Gauge density in the SGP97 region is considerably higher than the rest of the world. This experiment evaluates the performance of the EnMB under the incorrect assumption that storm timing is perfectly known. Figure 8 shows that the amount and timing of precipitation are considerably different at Little Washita and El Reno.

5.2. Estimating Surface Soil Moisture at El Reno With Precipitation Forcing From Gauge Data at El Reno

[49] Figure 9 compares the true soil moisture to that estimated using the EnOL, EnKF, and EnMB. The benefit of smoothing is particularly noticeable between Julian days 226 and 229. Elsewhere, the improvement over filtering is relatively modest. This may be due to the limited growth of uncertainty in this experiment due to the assumption that the timing of precipitation is perfectly known. The growth of uncertainty between observations

is limited to the uncertainty associated with the unknown parameters.

[50] From Figure 10 the reduction in standard deviation due to smoothing exceeds that achieved by filtering. The uncertainty introduced in this experiment is very small; the standard deviation across the filtered ensemble is on the order of 0.02, about 5% of the dynamic range of volumetric soil moisture. The limited improvement due to data assimilation between days 219 and 225 suggests that the uncertainty in the modeled estimate is comparable with the observation error.

5.3. Estimating Soil Moisture at Depth at El Reno Using El Reno Forcing Data

[51] Figure 11 shows the deviation between the true soil moisture at depth and the the estimate from the ensemble algorithms. Spurious dry initial conditions persist longer at depth, as illustrated by the EnOL estimate in layers 4 and 5. In the filter and smoother, the states at depth are updated through their correlation with the surface state and the observations. The filter improves more slowly than the smoother as it processes the observations sequentially. The smoother updates using observations in a batch, therby tying the estimate closer to the truth between observations.

5.4. Estimating Surface Soil Moisture at Little Washita With Precipitation Forcing From Gauge Data at El Reno

[52] Figure 12 shows the estimated surface soil moisture at Little Washita. The EnMB improves over the EnKF and EnOL, but is unable to correct entirely for the fact that storms occurred at El Reno while Little Washita was dry. When precipitation occurs at El Reno all ensemble members
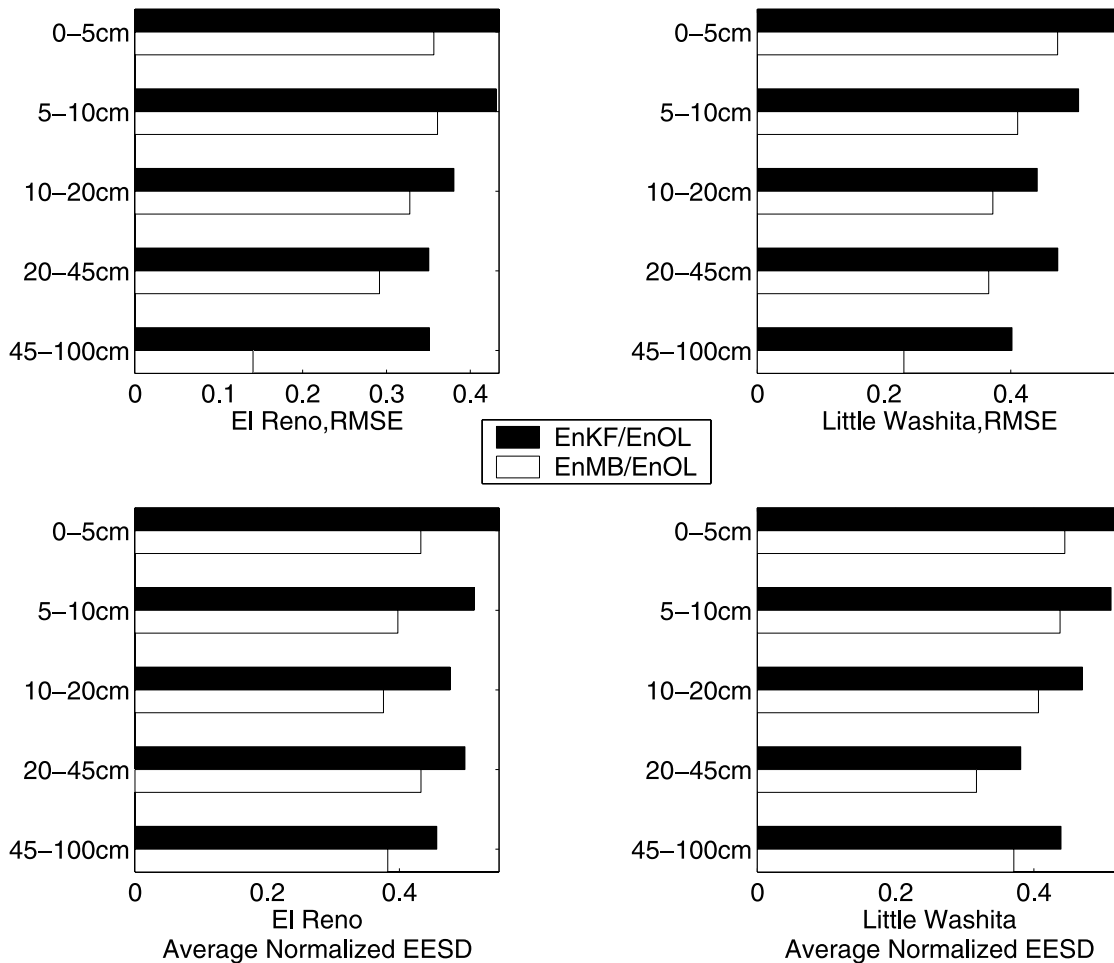
Figure 13. Normalized RMSE and average normalized EESD in the smoothed estimate of volumetric soil moisture θ compared to that of the filter and open loop.

receive precipitation, thereby reducing ensemble spread. Because of this apparent certainty that the soil at Little Washita is wet, the filter and smoother fail to update the ensemble mean toward the true value. This demonstrates the importance of correctly characterizing the sources of error and uncertainty in land data assimilation.

5.5. Summary Statistics at El Reno and Little Washita

[53] From Figure 13, filtering yields about a 50% reduction in RMSE compared to the EnOL at the surface. Ensemble smoothing yields a further 20% reduction on average. At depth, the greater improvement in smoothing over filtering is largely due to smoothing's ability to correct for erroneous initial conditions. With longer experiments, this effect would be reduced. The EnKF yields a 50% reduction in EESD over the EnOL, but the EnMB yields a further 20% improvement over the sequential filter. The improvement is apparent at both El Reno and Little Washita.

6. Hybrid Filter/Smoother Approach

[54] Recall from section 4 that while the EnMB yielded improved results compared to filtering, the backward propagation of information pertaining to the soil's response to

subsequent precipitation led to spuriously moist estimates. From Figure 3 it is evident that smoothing is most advantageous where the objective is to measure a particular dry-down series. Conversely, the smoother is least beneficial when the smoothing interval is disrupted by intermittent precipitation.

[55] Here a method is proposed to objectively divide the study interval into a series of dry-down events over which to smooth. It would be undesirable to use precipitation data for this purpose, as the objective is to estimate soil moisture with uncertain precipitation data and satellite data only. Fortuitously, the L band brightness temperature observations can be used to make a first-order assessment of when in the study interval wetting has occurred.

[56] Figure 14 shows the precipitation recorded at El Reno over the 4-month-long synthetic experiment (top). The resultant modeled soil moisture at the surface at El Reno is shown in the middle panel. L band brightness temperature T_B observations were simulated from this soil moisture using the Radiative Transfer Model (bottom). T_B is a function of soil moisture, soil temperature, and many soil and vegetation parameters. In general, however, a decrease in T_B indicates the interim occurrence of precipitation so smoothing would offer no improvement over filtering. Because the observations have an error of 3K (1σ),
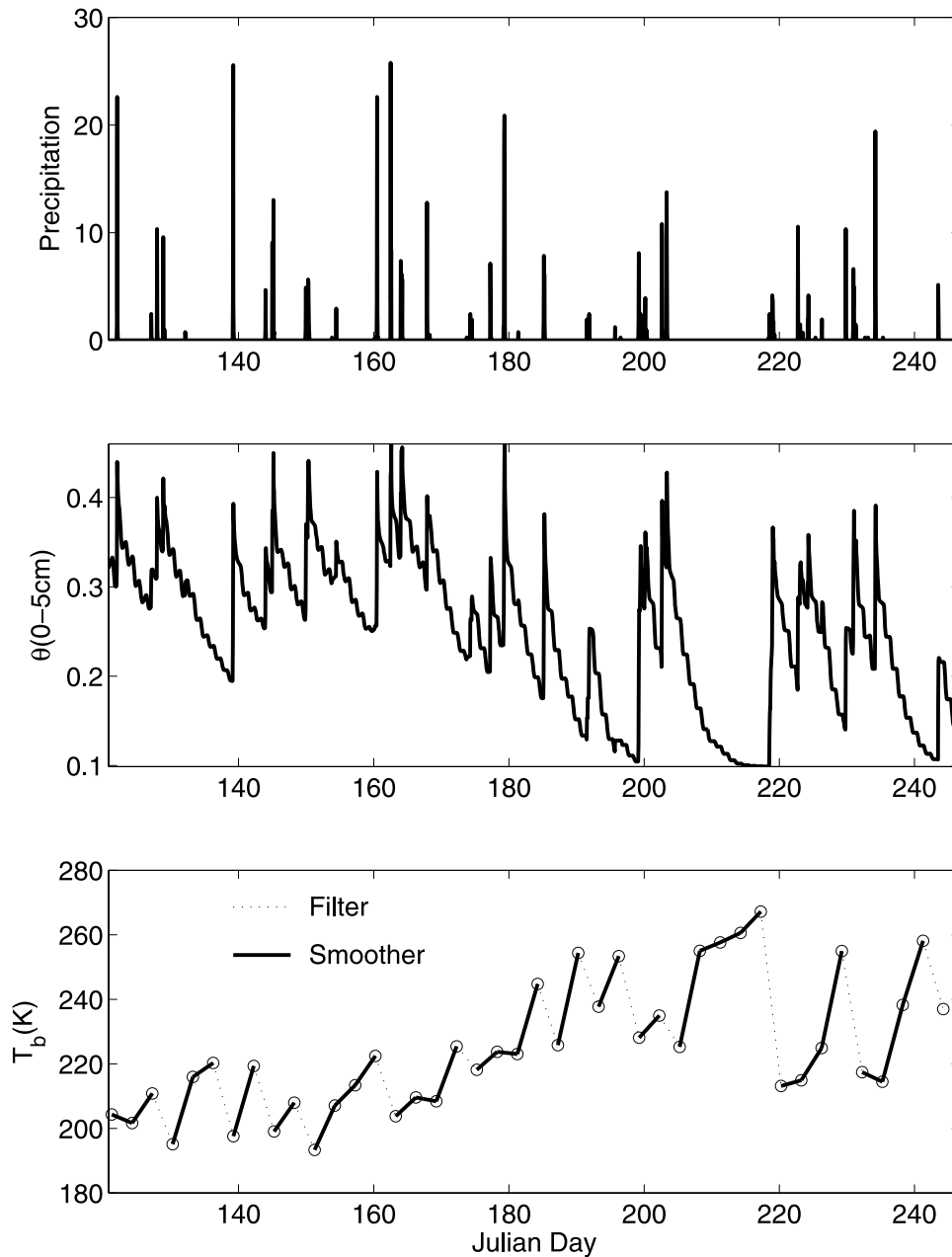
Figure 14. (top) Incident precipitation (in mm h^{-1}) at El Reno. (middle) Resultant modeled volumetric soil moisture θ at El Reno. (bottom) Simulated brightness temperatures T_B associated with these soil moisture values. The solid lines indicated smoothing intervals which are separated by filtered intervals.

decreases of less than 6K are disregarded. Provided the brightness temperature is increasing the soil is drying down, and smoothing should improve over filtering.

[57] Instead of prescribing the fixed length of a moving smoother window, in this approach the length of the smoother window is dynamic such that the augmented state vector consists of all estimation times on a given dry-down curve. Soil moisture estimation using this technique should yield improved estimates through two mechanisms: (1) Preventing backward propagation of information from a subsequent dry-down and (2) lengthening the smoother window to encompass all observations on the dry-down curve of interest

guarantees that the state is estimated using all relevant observations.

6.1. Results

[58] In Figure 15 the "hybrid smoother/filter" performance is compared to that of the EnKF and the EnMB. The key benefit of this hybrid algorithm is seen, for example, on the dry-down beginning on day 205. As the brightness temperatures are increasing for 12 days, the smoother window encompasses five observations. Using all of these observations in a single batch to estimate soil moisture at all estimation times in that interval yields an improvement over
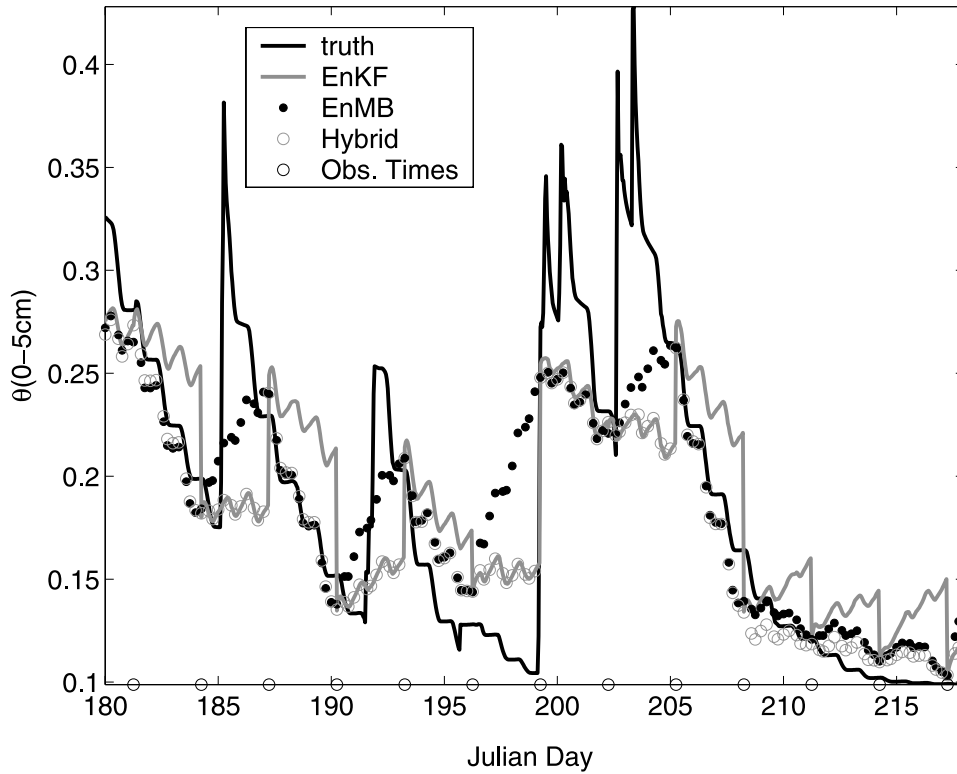
Figure 15. Ensemble mean volumetric soil moisture θ in the top 5 cm of the soil column at El Reno compared to the synthetic truth for the period between Julian days 180 and 218.
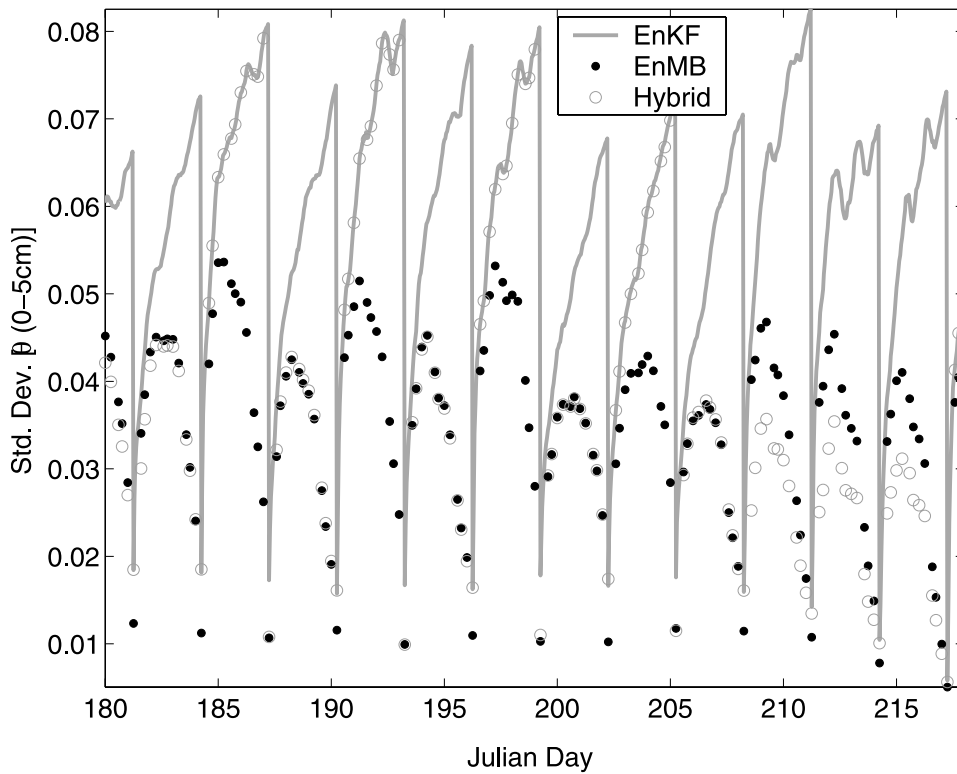

Figure 16. Estimation error standard deviation in the estimate of surface (0–5 cm) volumetric soil moisture at El Reno. The EESD from the hybrid smoother/filter approach is compared to that obtained using the EnKF and EnMB alone. Results are shown for the period between Julian days 180 and 218.

using the EnMB (LSW = 2). The hybrid algorithm also improves the estimate during wetting periods where the filter is used instead of the EnMB. Precipitation occurs immediately prior to the observation on day 199. The EnMB propagates information on the wet condition back in time yielding a moister estimate between days 196 and 199. In the hybrid algorithm the filter is used in this interval, preserving the drier soil moisture condition.

[59] There are cases where the hybrid algorithm can result in a poorer estimate than the EnMB (days 184–187 and 202–205). Here the precipitation occurs just after an observation filtering underestimates the soil moisture, and the hybrid is therefore too dry in this interval. When the precipitation occurs halfway between observations (Julian days 190–193), the hybrid algorithm has no net effect. These results demonstrate the difficulties of estimating soil moisture in intermittent precipitation using temporally sparse observations.

[60] The impact of using the hybrid algorithm is also apparent in the reduction in ensemble standard deviation (Figure 16) compared to the EnKF and the EnMB alone. When there is intermittent precipitation, the algorithm switches constantly between filtering and smoothing. When the hybrid selects the filter, ensemble spread grows like that in the filter, unconstrained by subsequent observations. Similarly, when the smoother is used for an interval of length 2 (i.e., two observations), the standard deviation across the ensemble is comparable to that of the EnMB. However, when the hybrid recognizes a lengthy dry-down and estimates the soil moisture over the entire interval as a long batch, the impact of additional future observations reduces the ensemble standard deviation below that of the moving batch smoother. The issue of the relative timing of precipitation and observations merits further attention. Nonetheless, this approach makes tentative steps to address the apparent pitfalls in using ensemble smoothing techniques in soil moisture estimation.

7. Conclusion and Discussion

[61] It is argued soil moisture estimation is a reanalysis-type problem rather than a control-type or forecast problem and consequently a smoothing approach is more appropriate than filtering. An ensemble-based smoothing algorithm was presented in which all observations within a prescribed window are used in a batch estimator to determine soil moisture at the surface and at depth. The algorithm was compared to the ensemble Kalman filter in two experiments with different precipitation data, and smoothing improved the estimated soil moisture at the surface and at depth. Smoothing was particularly effective in correcting for erroneous initial conditions at depth. This improvement is significant as it may lead to improved surface flux estimation through the dependence of the latent heat flux on root zone soil moisture. The smoother incorporates more observations than the filter to obtain the estimate, and thus is characterized by significantly reduced estimation errors and increased confidence in the estimate.

[62] The use of smoothing in land data assimilation is complicated by the occurrence of precipitation. A hybrid smoother/filter approach was presented to address this by breaking the study interval into a series of smoothing windows. The smoother window length is dynamic rather

than prescribed, including all observations in a single dry-down period. The soil moisture for the whole dry-down is determined in one batch. This method improves the estimate by preventing the backward propagation of information from precipitation events after an observation at the end of a smoothing window. Here the hybrid assumes precipitation has occurred if the decrease in brightness temperature is greater than 2σ, i.e., twice the standard deviation in the observation. While this is simplistic, it demonstrates the feasibility of using brightness temperature to break the interval into dry-down events. Further experiments will address the issue of spatial variability in brightness temperature. The impact of the relative timing of precipitation and observations merits further attention as the performance of the hybrid depends on when in the interobservation period the precipitation occurred.

[63] So far, the performance of the smoother has been evaluated on independent uncorrelated pixels. In future work, the smoother will be used to estimate soil moisture over a grid of spatially correlated pixels to estimate soil moisture from combined active and passive (multiscale) microwave-based observations like those expected from Hydros.

References

Boni, G., F. Castelli, and D. Entekhabi (2001), Sampling strategies and assimilation of ground temperature for the estimation of surface energy balance components, *IEEE Trans. Geosci. Remote Sens.*, *39*, 165–172.

Brusdal, K., J. M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. J. Van Leeuwen, E. Dombrowsky, and J. Verron (2003), A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems, *J. Mar. Syst.*, *40–41*, 253–289.

Burgers, G., P. J. Van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*(6), 1719–1724.

Castelli, F., D. Entekhabi, and E. Caporali (1999), Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance, *Water Resour. Res.*, *35*, 3115–3125.

Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q. Duan, M. Ek, and A. Betts (1996), Modeling of land surface evaporation by four schemes and comparison with FIFE observations, *J. Geophys. Res.*, *101*(D3), 7251–7268.

Choudhury, B. J., T. Schmugge, A. Chang, and R. W. Newton (1979), Effect of surface roughness on the microwave emission from soils, *J. Geophys. Res.*, *84*(C9), 5699–5706.

Courtier, P., J. Derber, R. Errico, J.-F. Louis, and T. Vukicevic (1993), Important iterature on the use of adjoint, variational methods and the Kalman filter in meteorology, *Tellus, Ser. A*, *45*, 342–357.

Crosson, W. L., C. Laymon, R. Inguva, and M. Schamschula (2002), Assimilating remote sensing data in a surface flux-soil moisture model, *Hydrol. Processes*, *16*(8), 1645–1662.

Crow, W. T. (2003), Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble kalman filter and surface brightness temperatures, *J. Hydrometeorol.*, *4*(5), 960–973.

Crow, W. T., and E. F. Wood (2003), The assimilation of remotely-sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on estar measurements during SGP97, *Adv. Water Resour.*, *26*(2), 137–149.

Entekhabi, D., H. Nakamura, and E. 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.

Entekhabi, D., et al. (2004), The Hydrosphere State (Hydros) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *IEEE Trans. Geosci. Remote Sens.*, 42, 2184–2195.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99(C5), 10,143–10,163.

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

Evensen, G., and P. J. Van Leeuwen (1996), Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model, *Mon. Weather Rev.*, 124(1), 85–96.

Evensen, G., and P. J. Van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, 128(6), 1852–1867.

Galantowicz, J., D. Entekhabi, and E. Njoku (1999), Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness, *IEEE Trans. Geosci. Remote Sens.*, 37, 1860–1870.

Gelb, A. (1974), *Applied Optimal Estimation*, MIT Press, Cambridge, Mass.

Ghil, M., and P. Manalotte-Rizzoli (1991), Data assimilation in meteorology and oceanography, *Adv. Geophys.*, 33, 141–266.

Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *Trans. Math. Software*, 24(4), 437–474.

Gronnevik, R., and G. Evensen (2001), Application of ensemble-based techniques in fish stock assessment, *Sarsia*, 86, 517–526.

Hawk, K., and P. Eagleson (1992), Climatology of station storm rainfall in the continental United States: Parameters of the Bartlett-Lewis and Poisson rectangular pulses models, *Tech. Rep. 336*, Mass. Inst. of Technol., Cambridge.

Heemink, A., M. Verlaan, and A. Segers (2001), Variance reduced ensemble Kalman filtering, *Mon. Weather Rev.*, 129(7), 1718–1728.

Houtekamer, P., and H. Mitchell (1998), Data assimilation using an Ensemble Kalman filter technique, *Mon. Weather Rev.*, 126(3), 796–811.

Houtekamer, P., and H. Mitchell (2001), A sequential Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, 129(1), 123–137.

Jackson, T. (1997), Soil moisture estimation using special satellite microwave/imager (SSM/I) satellite data over a grassland region, *Water Resour. Res.*, 33, 1475–1484.

Jackson, T., and T. Schmugge (1991), Vegetation effects on the microwave emission of soils, *Remote Sens. Environ.*, 36(3), 203–212.

Jackson, T., D. L. Vine, C. Swift, T. Schmugge, and F. Schiebe (1995), Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita'92, *Remote Sens. Environ.*, 54(1), 27–37.

Jackson, T., D. LeVine, A. Hsu, A. Oldak, P. Starks, C. Swift, J. Isham, and M. Haken (1999), Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains Hydrology Experiment, *IEEE Trans. Geosci. Remote. Sens.*, 37, 2136–2151.

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Elsevier, New York.

Ljung, L. (1979), Asymptotic behaviour of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Trans. Autom. Control*, 24, 36–50.

Lohmann, D., et al. (2004), Streamflow and water balance intercomparisons of four land surface models in the North America land data assimilation system project, *J. Geophys. Res.*, 109, D07S91, doi:10.1029/2003JD003517.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, 112(474), 1177–1194.

Margulis, S. A. (2002), Variational sensitivity analysis and data assimilation studies of the coupled land surface-atmospheric boundary layer system, Ph.D. thesis, Mass. Inst. of Technol., Cambridge.

Margulis, S. A., and D. Entekhabi (2001), Temporal disaggregation of satellite-derived monthly precipitation estimates and the resulting propagation of error in partitioning of water at the land surface, *Hydrol. Earth Syst. Sci.*, 5(1), 27–38.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment, *Water Resour. Res.*, 38(12), 1299, doi:10.1029/2001WR001114.

Nakamura, H., D. Entekhabi, and E. Njoku (1994), Soil moisture profile retrieval by assimilation by multispectral remote sensing and surface observations, paper presented at Geosciences and Remote Sensing Annual Conference, Inst. of Electr. and Electron. Eng., Pasadena, Calif.

Reichle, R. (2000), Variational assimilation of remote sensing data for land surface hydrologic applications, Ph.D. thesis, Massachusetts Institute of Technology.

Reichle, R., D. Entekhabi, and D. McLaughlin (2001a), Downscaling of radiobrightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach, *Water Resour. Res.*, 37, 2353–2364.

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

Reichle, R., D. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, 130(1), 103–114.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire (1984), Scale considerations in the modeling of temporal rainfall, *Water Resour. Res.*, 20, 1611–1619.

Segers, A., A. Heemink, M. Verlaan, and M. VanLoon (2000), A modified RRSQRT-filter for assimilating data in atmospheric chemistry models, *Environ. Modell. Software*, 15, 663–671.

Ulaby, F., R. Moore, and A. Fung (1986), *Microwave Remote Sensing*, Artech House, Norwood, Mass.

Van Leeuwen, P. J. (2001), An ensemble smoother with error estimates, *Mon. Weather Rev.*, 129(4), 709–728.

Van Leeuwen, P. J., and G. Evensen (1996), Data assimilation and inverse methods in terms of a probabilistic formulation, *Mon. Weather Rev.*, 124(12), 2898–2913.

Verlaan, M. (1998), Efficient Kalman filtering algorithms for hydrodynamic models, Ph.D. thesis, Technische Univ. Delft, Delft, Netherlands.

Verlaan, M., and A. Heemink (2001), Nonlinearity in data assimilation applications: A practical method for analysis, *Mon. Weather Rev.*, 129(6), 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(D11), 11,761–11,774.

Walker, J. P., G. R. Willgoose, and J. D. Kalma (2001), One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: A comparison of retrieval algorithms, *Adv. Water Resour.*, 24, 631–650.

Wang, J., and T. Schmugge (1980), An empirical model for the complex dielectric permittivity of soils as a function of water content, *IEEE Trans. Geosci. Remote Sens.*, 18, 288–295.

Wunsch, C. (1996), *The Ocean Circulation Inverse Problem*, Cambridge Univ. Press, New York.

—————————————

S. Dunne and D. Entekhabi, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. (susan@mit.edu; darae@mit.edu)