Reducing Water Imbalance in Land Data Assimilation: Ensemble Filtering without Perturbed Observations

M. TUGRUL YILMAZ*

George Mason University, Fairfax, Virginia

TIMOTHY DELSOLE

George Mason University, Fairfax, Virginia, and Virginia Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

PAUL R. HOUSER

George Mason University, Fairfax, Virginia

ABSTRACT

It is well known that the ensemble Kalman filter (EnKF) requires updating each ensemble member with perturbed observations in order to produce the proper analysis-error covariances. While increased accuracy in a mean square sense may be preferable in many applications, less accuracy might be preferable in other applications, especially if the variables being assimilated obey certain conservation laws. In land data assimilation, for instance, the update in soil moisture often produces a water balance residual, in the sense that the input water is not equal to output water. This study shows that suppressing the perturbation of observations in the EnKF and in the weakly constrained ensemble Kalman filter significantly improves the water balance residuals without significantly increasing the state errors.

1. Introduction

Shortly after Evensen (1994) proposed an ensemble technique for solving the Kalman filter equations, Burgers et al. (1998) and Houtekamer and Mitchell (1998) showed that the filter proposed by Evensen (1994) underestimates analysis errors. To remedy this shortcoming, Burgers et al. (1998) and Houtekamer and Mitchell (1998) proposed perturbing the observations in a manner consistent with their error statistics. For large ensemble sizes, this procedure produces an analysis ensemble with the correct covariance matrix. The resulting solution of the Kalman filter equations is now called the ensemble Kalman filter (EnKF).

Unfortunately, the addition of random errors on top of preexisting observation errors adds an additional source of sampling errors. Whitaker and Hamill (2002) showed that this additional source of sampling error increases the probability that the analysis-error covariance will be underestimated. Several ensemble-based solutions to the Kalman filter equations that avoid perturbed observations were proposed (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002), which were subsequently shown to belong to a family of solutions known as the square root filter (Tippett et al. 2003).

Interestingly, Burgers et al. (1998) found the state errors of perturbed and nonperturbed EnKF schemes were almost indistinguishable (see Burgers et al. 1998, their Fig. 1). Also, before the underestimation of analysis errors was noted by Burgers et al. (1998), several studies had already implemented ensemble-based Kalman filters without perturbed observations and found no obvious problems (Evensen 1994, 1997; Evensen and van Leeuwen 1996). The fact that the nonperturbed EnKF

---
* Current affiliation: Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland.

Corresponding author address: M. Tugrul Yilmaz, Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, 10300 Baltimore Ave., BARC-WEST, Bldg 007, Room 104, Beltsville, MD 20705.
E-mail: tugrul.yilmaz@ars.usda.gov

can give state estimates almost as good as the perturbed EnKF, even though it is not optimal and has biased analysis-error covariances, does not seem to be appreciated in the literature.

In some contexts, underestimation of analysis errors can have benefits that outweigh the negative effects associated with suboptimal filter performance. For instance, consider the assimilation of conserved quantities (e.g., water or energy). For the kinds of conservation laws that appear in geophysical applications, conserved quantities do not remain conserved after assimilation; moreover, the degree of imbalance invariably increases when the observations are perturbed. The resulting imbalance can be critical for some land data assimilation applications (Pan and Wood 2006; Yilmaz et al. 2011). In such cases, suboptimal schemes that avoid perturbed observations might be preferable to optimal schemes that perturb observations, especially if the suboptimal schemes have smaller imbalances.

Recently, Yilmaz et al. (2011) proposed new filter schemes in which the water budget is weakly constrained and the water imbalance (or, equivalently, the soil moisture state update) is reduced. One of these schemes is a variant of the traditional EnKF, called the weakly constrained ensemble Kalman filter (WCEnKF; Yilmaz et al. 2011). The question arises as to whether the water imbalance produced by the EnKF and the WCEnKF can be reduced by not perturbing observations. This paper investigates the effect of not perturbing observations in the EnKF and WCEnKF on water imbalance and state errors. Results show that not perturbing observations consistently reduces the water imbalance while negligibly changing the state errors.

The standard EnKF and the WCEnKF, as well as their nonperturbed variants, are discussed in the next section. Section 3 briefly reviews the models and performance metrics used to compare filters, section 4 presents the results, and section 5 summarizes our conclusions.

2. Ensemble filters

We consider two classes of ensemble filters in this study: the ensemble transform Kalman filter (ETKF; Bishop et al. 2001) and the EnKF (Burgers et al. 1998). In both filters, the best estimate of the state, called the analysis, is

$$\boldsymbol{\mu}_a = \boldsymbol{\mu}_f + \mathbf{K}(\mathbf{o} - \mathbf{H}\boldsymbol{\mu}_f), \qquad (1)$$

where $\boldsymbol{\mu}_f$ is the ensemble mean forecast, $\mathbf{o}$ is the observation vector, $\mathbf{H}$ is the interpolation operator that maps the forecast into observation space, and $\mathbf{K}$ is the Kalman gain matrix. The ensemble of anomalies about $\boldsymbol{\mu}_a$, denoted by the matrix $\mathbf{X}_a$, are generated differently in the two filters. In the ETKF, the analysis anomalies are generated by

$$\mathbf{X_a} = \mathbf{X}_f \mathbf{A}, \qquad (2)$$

where $\mathbf{A}$ is a suitably chosen transformation matrix. In the EnKF, analysis anomalies are generated by

$$\mathbf{X}_a = \mathbf{X}_f + \mathbf{K}(\mathbf{o}' - \mathbf{H}\mathbf{X}_f), \qquad (3)$$

where $\mathbf{o}'$ denotes "perturbed observations." Note that the symbol $\mathbf{X}_a$ is used to represent the analysis anomalies regardless of filter.

In standard Kalman filters, the solutions can be found equivalently by minimizing a cost function that is dependent on the differences between the model forecast and the updated analysis and between the observations and the updated analysis. In the constrained filter, the solution is sought by minimizing a cost function that includes a penalty function that also depends on the amount of imbalance created by the state update.

In the strongly constrained filter the water balance is closed perfectly. This perfect balance is accomplished by redistributing water to different components based on the degree of uncertainty in the components. However, if the errors in any of the water balance components are large, then these large errors would be redistributed to the other terms too, resulting in potentially unrealistic simulations. Such an example is given in Yilmaz et al. (2011). In the weakly constrained filter, the penalty term is weak in the sense that the water balance is not closed exactly and a residual is allowed, acknowledging the presence of errors in the water balance closure. The weak constraint may be preferable to the strong constraint when the terms in the balance have large errors, since the strong constraint preserves errors in the water balance closure.

Yilmaz et al. (2011) proposed new versions of the above filters in which a water budget constraint is imposed in the assimilation procedure. These filters differ from the constrained filters proposed by Simon and Chia (2002) and Pan and Wood (2006) in that the water budget is weakly constrained. The ensemble mean water budget equation can be expressed as

$$\beta - \mathbf{c}^{\mathrm{T}}\mathbf{x} = 0, \qquad (4)$$

where β depends both on nonprognostic hydrological variables at the current time step (e.g., precipitation, evaporation, and runoff) and on prognostic variables at a previous time step (soil variables), $\mathbf{c}$ is a unit conversion vector, and $\mathbf{x}$ is the state. Model forecasts already satisfy the water balance (4) in an individual sense and in an ensemble mean sense. The weakly constrained filter is designed to produce analyses that more closely satisfy water balance than unconstrained filters. In the weakly

constrained filters proposed by Yilmaz et al. (2011), the best estimate of the state is

$$\boldsymbol{\mu}_a = \boldsymbol{\mu}_f + \mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\boldsymbol{\mu}_f) + \mathbf{P}_a\mathbf{c}\varphi^{-1}(\mu_\beta - \mathbf{c}^{\mathrm{T}}\boldsymbol{\mu}_f),$$
(5)

where $\mathbf{P}_a$ is the sample covariance matrix of the analysis anomalies $\mathbf{X}_a$ (which differs for the different filters), $\mathbf{R}$ is observation error covariance matrix, μ_β is the ensemble mean constraint, and φ is the error variance of β' (deviations about ensemble mean β). We note that the last additive term in (5) vanishes because of (4), but we retain it for clarity in some terms below. Here $\mathbf{P}_a$ depends on the constraint, hence the mean update also depends on the constraint. The analysis anomalies of the weakly constrained ETKF (WCETKF) and the WCEnKF differ. In the WCETKF, the analysis anomalies are derived from an equation of the form (2), except with a different transformation matrix $\mathbf{A}$ that depends on the constraint. In the WCEnKF, the analysis anomalies are

$$\mathbf{X}_a = \mathbf{X}_f + \mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{o}' - \mathbf{H}\mathbf{X}_f) + \mathbf{P}_a\mathbf{c}\varphi^{-1}(\beta' - \mathbf{c}^{\mathrm{T}}\mathbf{X}_f).$$
(6)

For further details, see Yilmaz et al. (2011).

This paper compares the above filters with modified EnKFs in which observations are not perturbed. We also consider suppressing anomalies in the constraint term or, equivalently, using only the ensemble mean constraint term in the update equation. (Just to be clear, we do not perturb the constraint terms in this paper.) The proposed filters are suboptimal in the sense that they underestimate the analysis-error covariances but are expected to have smaller water budget residuals since imbalances caused by random perturbations are removed. Accordingly, we consider the ensemble Kalman filter with no perturbed observations (EnKF-noPO) (Whitaker and Hamill 2002), whose analysis anomalies are

$$\mathbf{X}_a = \mathbf{X}_f + \mathbf{K}\mathbf{H}\mathbf{X}_f.$$
(7)

We also consider the weakly constrained ensemble Kalman filter with no perturbed observations (WCEnKF-noPO), whose analysis anomalies are

$$\mathbf{X}_a = \mathbf{X}_f - \mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}_f) + \mathbf{P}_a\mathbf{c}\varphi^{-1}(\beta' - \mathbf{c}^{\mathrm{T}}\mathbf{X}_f).$$
(8)

In addition, we consider the weakly constrained ensemble Kalman filter with no constraint anomalies (WCEnKF-noCA), whose analysis anomalies are

$$\mathbf{X}_a = \mathbf{X}_f + \mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{X}_f) - \mathbf{P}_a\mathbf{c}\varphi^{-1}(\mathbf{c}^{\mathrm{T}}\mathbf{X}_f).$$
(9)

Finally, we consider the weakly constrained ensemble Kalman filter with no perturbed observations and no

TABLE 1. Summary of filters and their distinctive analysis properties. Here O' denotes presence of perturbed observations and B' denotes the presence of constraint anomalies. A dash means the particular anomaly does not apply for that filter, ✔ means the anomaly exists in the solution, and X means the anomaly is not used in the solution.

	O'	B'
ETKF	—	—
WCETKF	—	—
EnKF	✔	—
EnKF-noPO	X	—
WCEnKF	✔	✔
WCEnKF-noPO	X	✔
WCEnKF-noCA	✔	X
WCEnKF-noPO-noCA	X	X

constraint anomalies (WCEnKF-noPO-noCA), whose analysis anomalies are

$$\mathbf{X}_a = \mathbf{X}_f - \mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}_f) - \mathbf{P}_a\mathbf{c}\varphi^{-1}(\mathbf{c}^{\mathrm{T}}\mathbf{X}_f).$$
(10)

In sum, there are a total of eight filters considered in this paper; four of them [(7)–(10)] based on some form of "no perturbations." These filters are briefly summarized in Table 1. In this paper, we have compared the above-described eight filters under different ensemble size and assimilation frequency selection scenarios. Special emphasis is given to the comparison of the state error and the water balance residual magnitudes under the different scenarios.

The fact that observations are not perturbed should not be construed as an assumption that observations are perfect. In both the perturbed and the nonperturbed cases, observation errors are taken into account in the Kalman gain, which in turn produces the correct mean analysis. The lack of perturbed observations only implies that the analysis spread is underestimated.

3. Simulations and analysis

a. Experiment setup

Separate simulations were performed for each of the above-described eight filters (ETKF, WCETKF, EnKF, EnKF-noPO, WCEnKF, WCEnKF-noPO, WCEnKF-noCA, and WCEnKF-noPO-noCA). The setups of the experiments in this study were identical to the setups described in the study of Yilmaz et al. (2011). Briefly, synthetic experiments were performed using Noah model (Ek et al. 2003) version 2.7. The study area was chosen as Oklahoma (between 32.0°–37.0°N and 96.0°–91.0°W) with 0.125° spatial resolution. The simulations were performed between April and October 2006. North America Land Data Assimilation (NLDAS; Cosgrove et al. 2003) data were used as the atmospheric forcing.

Model grid spatial resolutions were selected consistent with the forcing data so that no averaging or downscaling was needed. Initial states were obtained after spinning up the model for 10 years. Ensembles were created by perturbing initial states of temperature and soil moisture and by perturbing forcing data based on precipitation, air temperature, and shortwave and longwave radiation, as described by Yilmaz et al. (2011).

The "truth" is identified with a single solution of the model with unperturbed initial conditions and forcing. The "open loop" is defined as the ensemble of simulations with the same perturbed forcing as the assimilation simulations, but without assimilation of observations (Yilmaz et al. 2011). "Observations" are generated by adding random variables with covariance matrix **R** to the truth states. Perturbed observations are generated by perturbing the observations in the same way.

Observations of four layers of soil moisture and four layers of soil temperature were assimilated. For initial assessments, observations were assimilated once a day using 50 ensemble members. Later, for sensitivity analysis, observations were assimilated at assimilation frequencies changing hourly to daily using ensemble sizes changing from 10 to 150.

b. Performance metrics

Because of the time interval selection (April–October; no snow), snow-related variables were effectively not updated, hence snow-related results were not investigated and will not be presented in this study. Root-mean-square error (RMSE) of ensemble means for each soil moisture (SM) and soil temperature (ST) layer and for each of nine experiments were calculated as

$$\text{RMSE.SM}_i = \sqrt{\sum_{\text{sm}}^{4}\sum_{\text{lat}}^{39}\sum_{\text{lon}}^{39}\sum_{t} (\text{SM}_{s\,i\,\text{lon lat}\,t} - \widehat{\text{SM}}_{s\,i\,\text{lon lat}\,t})^2/(4500 \times 4 \times 39 \times 39)} \quad \text{and} \quad (11)$$

$$\text{RMSE.ST}_i = \sqrt{\sum_{\text{sm}}^{4}\sum_{\text{lat}}^{39}\sum_{\text{lon}}^{39}\sum_{t} (\text{ST}_{s\,i\,\text{lon lat}\,t} - \widehat{\text{SM}}_{s\,i\,\text{lon lat}\,t})^2/(4500 \times 4 \times 39 \times 39)}, \quad (12)$$

where $\widehat{\text{SM}}$ and $\widehat{\text{ST}}$ are the true SM and ST states, s is each soil layer (total 4), i identifies the experiments (open loop and eight filters defined above, total 9), lon is longitude pixel number (total 39), lat is latitude pixel number (total 39), and t is each time step (total 4501). The water balance residuals

$$r = \beta - \mathbf{c}^{\text{T}}\mathbf{x}_a \quad (13)$$

were calculated for each ensemble member, at each assimilation time step, at each pixel in the study area, and for each set of experiments separately. In this study, the ensemble mean, the time mean, and the time variance of the residuals are calculated as

$$r_{i\,\text{lon lat}\,.\,t} = \sum_{n} r_{i\,\text{lon lat}\,n\,t}/N, \quad (14)$$

$$r_{i\,\text{lon lat}\,..} = \sum_{t} r_{i\,\text{lon lat}\,.\,t}/\text{ats}, \quad \text{and} \quad (15)$$

$$\sigma^2 r_{i\,\text{lon lat}} = \sum_{t} (r_{i\,\text{lon lat}\,.\,t} - r_{i\,\text{lon lat}\,..})^2/(\text{ats} - 1), \quad (16)$$

where the "dot" subscript denotes an index that is averaged out, $\sigma^2 r$ is the residual variance, n denotes ensemble member, and ats is the total number of assimilation time steps (4501 and 187 for hourly and daily update scenarios, respectively), where only the residuals due to assimilation of observations were included in the residual statistics. Then $\sigma^2 r_{i\,\text{lon lat}}$ values were averaged over the study area into a single number $(\sigma^2 r_{i..})$ for each experiment separately.

Residuals $(r_{i\,\text{lon lat}\,.\,t})$ can be thought as the ensemble mean soil storage update by the filter or equally thought as the ensemble mean of the water budget imbalance from the beginning to the end of the assimilation window. In practice, the truth is never known for the purpose of calculation of the "true" error; however, in this context it is known that the truth has 0 residual (perfect water balance closure).

Soil water content change due to model integration alone, for eight filters before the analysis is updated with observations, was calculated as

$$\text{mwc}_{i\,\text{lon lat}\,.\,t} = \sum_{n} \text{mwc}_{i\,\text{lon lat}\,n\,t}/N, \quad (17)$$

$$\text{mwc}_{i\,\text{lon lat}\,..} = \sum_{t} \text{mwc}_{i\,\text{lon lat}\,.\,t}/\text{ats}, \quad \text{and} \quad (18)$$

$$\sigma^2\text{mwc}_{i\,\text{lon lat}} = \sum_{t} (\text{mwc}_{i\,\text{lon lat}\,.\,t} - \text{mwc}_{i\,\text{lon lat}\,..})^2/(\text{ats} - 1), \quad (19)$$

where mwc is the model water content change at each time step, and σ^2mwc is its variance. Similar to the residual variance metric, a single averaged mwc variance value (σ^2mwc$_{i..}$) was calculated for each experiment separately.

c. Significance tests

The degree of performance change between filters with perturbed and nonperturbed observations was assessed by a series of F tests using daily residuals (total 187 for each filter and pixel). Residual variance ratios of filters were calculated per pixel as

$$\text{VarRatio1}_{\text{lon lat}} = \sigma^2 r_{\text{EnKF}_{\text{lon lat}}} / \sigma^2 r_{\text{EnKF-noPO}_{\text{lon lat}}}, \tag{20}$$

$$\text{VarRatio2}_{\text{lon lat}} = \sigma^2 r_{\text{wcEnKF}_{\text{lon lat}}} / \sigma^2 r_{\text{wcEnKF-noPO}_{\text{lon lat}}}, \quad \text{and} \tag{21}$$

$$\text{VarRatio3}_{\text{lon lat}} = \sigma^2 r_{\text{wcEnKF-noCA}_{\text{lon lat}}} / \sigma^2 r_{\text{wcEnKF-noPO-noCA}_{\text{lon lat}}}. \tag{22}$$

The residual improvement is counted as significant if Var Ratio at any given pixel is higher than the critical F value. The one-tailed F-test critical value at the 5% significance level for 186 degrees of freedom (dof) is 1.27.

The above F-test configuration assumes that the residuals are independent in time. However, any significant autocorrelation in the daily residuals may decrease the d.o.f. and thus requires a new F test. Separate d.o.f. were calculated to reflect this temporal relation. Autocorrelations of the residuals were calculated for each pixel and filter separately. Decorrelation time scale (Shukla and Gutzler 1983), where no significant autocorrelation is assumed for longer time lags, was calculated as

$$\text{dts} = 1 + 2\sum_{k=1}^{20} \rho_k, \tag{23}$$

where dts is the decorrelation time scale, k is the time lag, and ρ_k is the correlation at k time lag. Using this time scale, new adjusted degrees of freedom was calculated for each pixel and filter separately as

$$\text{dof}_e = \frac{187}{\text{dts}} - 1, \tag{24}$$

where dof_e is the number of effective degrees of freedom and dof is the nonadjusted degrees of freedom (187). Two sets of F tests were performed using dof_e and dof, respectively, for each of these three filter pairs and pixels. Then, for each of these sets, the percentage of pixels that has significant residual improvement was calculated.

4. Results

The RMSE of soil temperature and soil moisture for all eight assimilation schemes, as well as for the observations and open loop, are shown in Fig. 1. We see immediately that the errors of the different assimilation schemes are comparable to each other, and less than those of observations or the open loop. As anticipated, the errors for the nonperturbed EnKF schemes (EnKF-noPO, WCEnKF-noPO, WCEnKF-noCA, or WCEnKF-noPO-noCA) tend to be larger than those of their perturbed counter parts (EnKF or WCEnKF). Nevertheless, for most practical purposes, the error differences between assimilation schemes are indistinguishable.

At any given assimilation time step (when observations are available), total soil water content change is equal to summation of the change due to the model integration alone and the change due to the state update as a result of assimilation of observations only [namely residuals; Eq. (14)]. The variance of the model-only water change (19) and the residual (16) for all eight assimilation schemes, as well as the truth and open loop, are shown in Fig. 2. Since the truth and open loop conserve water exactly at each time step, the residual vanishes and hence has zero variance. In this figure, residual variances remained much higher than the variance of the model water change for all eight filters [black bars (residuals) are much higher than the gray bars (water change due to model only)]; residuals dominate the total soil water content change (black bars + gray bars) when observations are assimilated daily. The figure also shows that residual variances of the standard ETKF and EnKF are larger than that of either the weakly constrained or nonperturbed EnKF schemes. Among these residual improvements, EnKF improvements were significant 66% of the time when dof is used and 38% of the time when dof_e is used; WCEnKF and WCEnKF-noCA improvements were significant 1%–3% of the time when dof or dof_e is used. Furthermore, the residual variance of the nonperturbed filters is consistently less than the
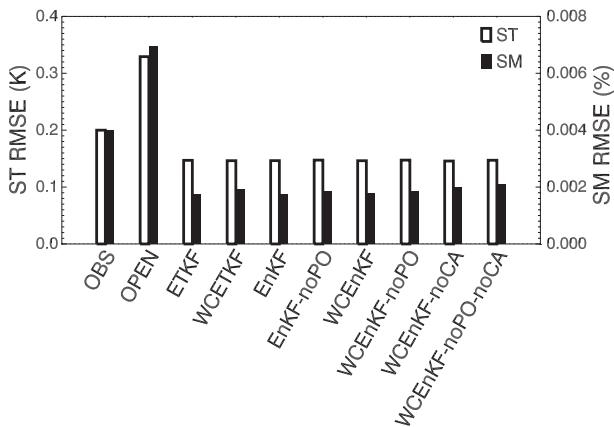
FIG. 1. SM and ST RMSE of observations, open loop, and ETKF, WCETKF, EnKF, EnKF-noPO, WCEnKF, WCEnKF-noPO, WCEnKF-noCA, and WCEnKF-noPO-noCA filters (assimilating daily SM and ST observations with 50 ensemble members). ST errors are shown in gray color on left y axis and SM errors are shown in black color on the right y axis.
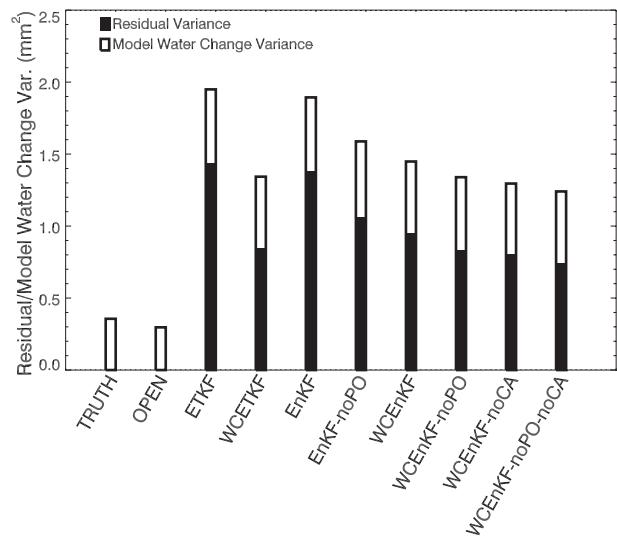

FIG. 2. Residual variances (16) for various filters (assimilating daily SM and ST observations with 50 ensemble members, shown in black bars), and model water change variances (19) for various filters, truth, and open loop simulations (shown in gray bars).

residual variance of their perturbed counterparts regardless of its significance; residual variances were improved 100%, 100%, and 98% of the time for EnKF, WCEnKF, and WCEnKF-noCA when observations are not perturbed. The fact that the improvements for the constrained filters are less significant than the unconstrained filters is not surprising since the constrained filters explicitly constrain the water balance residual to be small. On the other hand, 6-month cumulative residuals of nonperturbed filters (EnKF-noPO, WCEnKF-noPO, and WCEnKF-noPO-noCA) were only marginally (1%–2%) improved when compared to their perturbed pairs (EnKF, WCEnKF, and WCEnKF-noCA). The difference between the cumulative residuals of constrained and unconstrained cases are small because the analysis increments are unbiased with zero mean and the residuals average out for both scenarios over long period of time. In short, suppressing perturbations in the EnKF consistently improves the water budget residual without significantly increasing the state errors.

To test the sensitivity of the above results, we repeated the above experiments for a range of ensemble sizes and observation frequency. The results, presented in Fig. 3, consistently support the conclusion that filters without perturbed observations (or constraint anomalies) produce smaller budget residuals than their perturbed counterparts (in graphical terms, the symbols for the nonperturbed filters are shifted to the left of those for the perturbed filters). It is interesting to note that ensemble size has relatively little effect on the residual variance. Although the soil moisture errors tend to decrease with increasing ensemble size, this is not always

the case. In general, more frequent observations resulted in smaller residuals and soil moisture errors.

One anomalous result is worth noting: the residuals of the EnKF for hourly observations increased dramatically with ensemble size (red stars in top-left panel of Fig. 3). This sensitivity is substantially reduced when the filter is weakly constrained, and disappears altogether when observations are not perturbed. This behavior, which was not seen in square root–based filters (ETKF and WCETKF), was investigated extensively, including attempts to recreate the behavior with simple autoregressive models, but no satisfactory explanation could be found.

5. Conclusions

This paper investigated the use of "nonperturbed" ensemble Kalman filters in land data assimilation. Although these filters are suboptimal compared to the standard ensemble Kalman filters, in the sense that they underestimate the analysis errors, they produce smaller budget errors of conserved quantities because of the absence of artificial random perturbations. A total of eight data assimilation schemes were investigated (see Table 1), four of them being nonperturbed variants of previously proposed schemes. The major finding of this study is that suppressing perturbations in the EnKF significantly improves the water budget residual without significantly increasing the state errors. This finding was shown to be independent of ensemble size or observation frequency.

This improvement in the residual is expected for the following reason. Suppressing the perturbations in
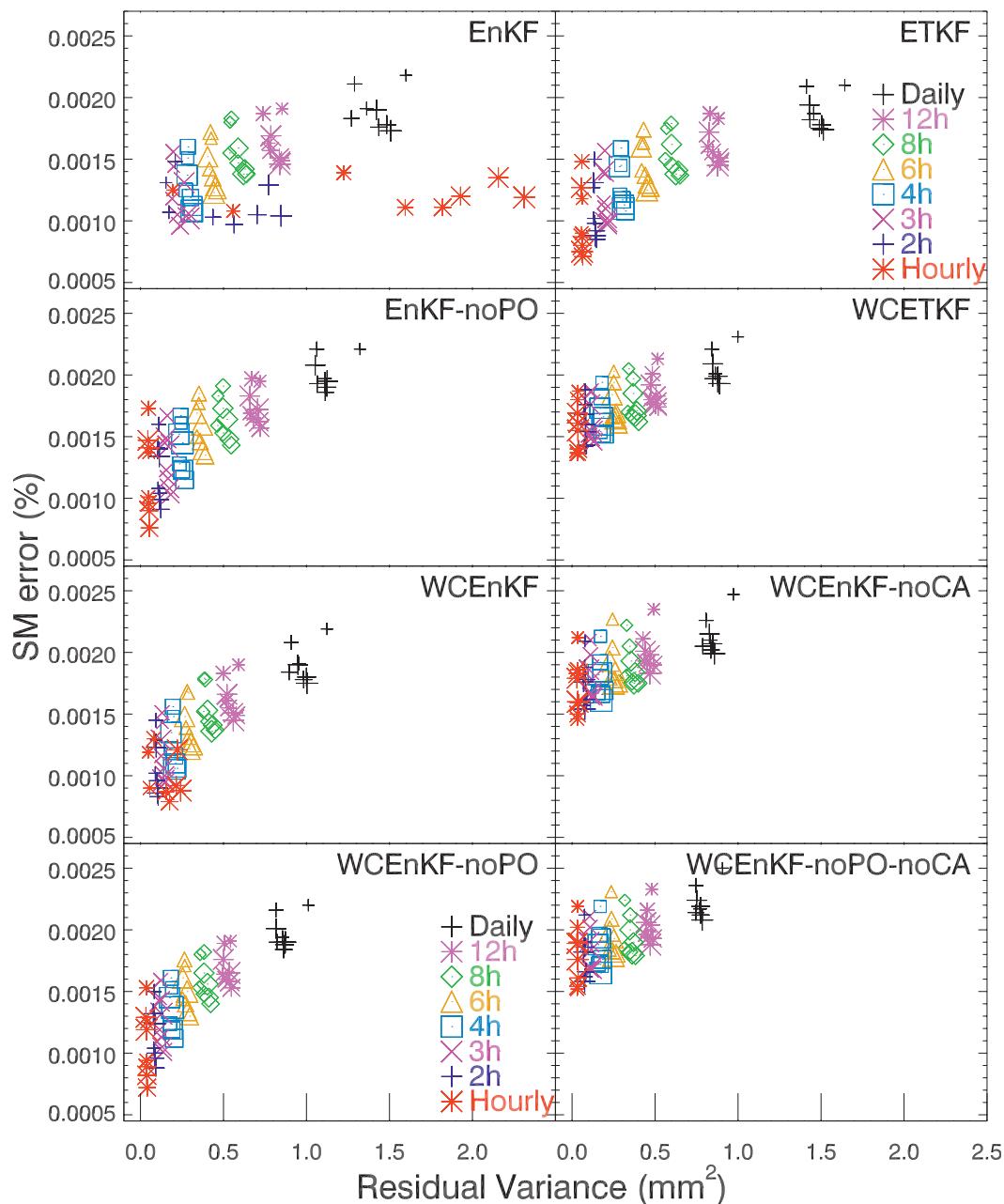
FIG. 3. Sensitivities of the eight filter residuals and the SM errors to the ensemble size and the assimilation frequency. Each panel represent a different filter, and each color in each panel represents a different assimilation frequency varying from hourly to daily; points with the same color and symbol represent a simulation with a different ensemble size (10, 15, 20, 30, 50, 80, 120, and 150) where the ensemble size increases with the increasing symbol size. All simulations are performed with both SM and ST observations assimilated.

observations leads to generally smaller analysis updates. A smaller increment implies that the analysis is closer to the forecast, which in turn has zero residual because the forecast model maintains water balance exactly. Note that suppressing perturbations on the observations has no effect on the mean increment at the current time step, but it reduces the analysis spread at the current time step

and subsequently reduces the mean analysis increment at the next time step.

The issue of not perturbing the observations, particularly the fact that earlier studies (Evensen 1994, 1997; Evensen and van Leeuwen 1996) had found no obvious problems in using nonperturbed suboptimal filters, have been discussed previously. Both Burgers et al. (1998)

and Whitaker and Hamill (2002) speculated that in oceanography the observation errors are much smaller than forecast errors and therefore there is relatively little difference between the perturbed and unperturbed cases. In the case of land data assimilation, however, we believe there is another relevant fact: land models are strongly dependent on the forcing (e.g., precipitation). As a result, forecast spread tends to be a stronger function of forcing spread than on initial condition spread. This strong dependence on forcing perhaps makes the perturbed observations less critical in land data assimilation than in atmospheric or oceanographic data assimilation. This dependence also may explain why inflation of the forecast covariance described by Anderson and Anderson (1999) usually is not needed in land data assimilation studies as much as in the atmospheric and the oceanographic data assimilation applications.

REFERENCES

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

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.,* **127,** 2741–2758.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420–436.

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

Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the north american land data assimilation system (NLDAS) project. *J. Geophys. Res.,* **108,** 8842, doi:10.1029/2002JD003118.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayand, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers For Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

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), 10 143–10 162.

——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.

——, 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. Wea. Rev.,* **124,** 85–96.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.

Pan, M., and E. F. Wood, 2006: Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *J. Hydrometeor.,* **7,** 534–547.

Shukla, J., and D. S. Gutzler, 1983: Interannual variability and predictability of 500 mb geopotential heights over the Northern Hemisphere. *Mon. Wea. Rev.,* **111,** 1273–1279.

Simon, D., and T. L. Chia, 2002: Kalman filtering with state equality constraints. *IEEE Trans. Aerosp. Electron. Syst.,* **38,** 128–136.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

Whitaker, J., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1924.

Yilmaz, M. T., T. DelSole, and P. R. Houser, 2011: Improving land data assimilation performance with a water budget constraint. *J. Hydrometeor.,* **12,** 1040–1055.