Improving Land Data Assimilation Performance with a Water Budget Constraint

M. Tugrul Yilmaz *

George Mason University, Fairfax, VA

Timothy DelSole

George Mason University, Fairfax, VA

Center for Ocean-Land-Atmosphere Studies, Calverton, MD

Paul R. Houser

George Mason University, Fairfax, VA

* *Corresponding author address:* Now at: Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, 10300 Baltimore ave., BARC-WEST, bldg 007, room 104, Beltsville, Maryland, 20705, USA

E-mail: tugrul.yilmaz@ars.usda.gov

ABSTRACT

A weak constraint solution was introduced to reduce the water budget imbalance that appears in land data assimilation as a result of state updates. Constrained Kalman Filter results were shown to be identical in single- or two-stages solutions whereas constrained Ensemble Transform Kalman Filter (ETKF) single- and two-stage solutions form two different square root solutions. The Weakly Constrained Ensemble Kalman Filter (WCEnKF) and the Weakly Constrained Ensemble Transform Kalman Filter (WCETKF) were evaluated for 3-hourly and daily update frequencies with soil moisture only, or soil moisture and soil temperature assimilated together. Simulations were performed using the Noah Land Surface Model (LSM) over Oklahoma, USA, using synthetic observations. State errors of constrained and unconstrained solutions were found to be similar; neither type had significantly smaller errors for most experiments. Constrained filters had smaller water balance residuals than unconstrained standard filters for all tested scenarios. The water balance residual of the ETKF and EnKF were similar for both 3-hourly and daily update experiments. The majority of the total column water change for daily updated filters resulted from the assimilation update.

1. Introduction

Data assimilation is a technique for optimally combining observations and model forecasts into a single best estimate of the state, while taking into account the accuracy of the two independent estimates. Data assimilation systems are optimum only in so far as certain underlying assumptions are fulfilled, namely that the forecast model is perfect, observations and forecasts are unbiased, observation errors are independent of the state, and all the distributions are Gaussian. However, available modeling and observing systems do not satisfy all these assumptions. In practice, the model is not perfect, observations and forecasts are biased, and the error covariances that are needed to solve the optimal solution are unknown.

In land surface applications, data assimilation methods have used satellite-, air-, and ground-based observations to improve estimates of soil moisture, soil skin temperature, discharge, snow water equivalent, snow cover, and water storage estimates (Houser et al. 1998; Lakshmi 2000; Pauwels et al. 2002; Reichle et al. 2008; Zaitchik et al. 2008; Crow and Ryu 2009; Kumar et al. 2009; De Lannoy et al. 2010). However, special problems occur when conserved quantities are assimilated. For instance, assimilation of hydrological observations (e.g. soil moisture) may improve estimates of hydrological variables, but generally degrade the water balance because the analysis increments do not conserve water since they are compensating for system biases or errors. Even if the dynamical model conserves water, the state update generally creates a water budget imbalance. If the degree of water imbalance is excessive, then it is reasonable to question whether an alternative data assimilation system should be employed, particularly one that reduces or removes the imbalance of water.

Skillful water estimation is important for hydrologists since it determines the location

of the stored water on land, eg. for streamflow, agricultural, and water management applications (Alsdorf et al. 2007). Accurate water budgets are important for estimating runoff, because runoff is calculated as a residual of other water balance terms. Skillful estimations of the water and energy cycles are also important for developing and validating hydrological models (Wei et al. 2010); in particular in model skill assessment, facilitating model parameterization developments, calibrating model parameterizations, better understanding the hydrological processes, assessing the role of land over climate predictability (DelSole et al. 2009; Dirmeyer 2003), and predicting future changes. In fact obtaining a "closed" water and energy balance estimate on a continental scale has been focus of many scientific experiments, particularly World Climate Research Program (WCRP) Global Energy and Water-Cycle Experiment (GEWEX) Continental-scale International Project (GCIP; Roads et al. 2003). It has been emphasized that the land-atmosphere interaction and the land water storage still remains as the future issues to be addressed (WCRP JSC Report 2010), which are primarily linked with water and energy cycles.

However, obtaining a balanced or closed water budget is not trivial: observations are not temporally and spatially adequate to obtain useful closure information, or to estimate their sampling uncertainties. Models have the potential to completely cover the region of interest temporally and spatially, but, they may suffer from inaccurate parameterizations. Hence, correct closure information may not be obtained from models alone. Data assimilation combines both observations and models by taking into account their error structures; however, as described above, their corrections may lead to water budget imbalance due to the state updates that correct system bias or error.

Pan and Wood (2006) proposed a constraint in land-data assimilation to ensure that

the data assimilation system conserved water. They have derived a two-stage constrained Kalman Filter solution in which the first stage is a traditional Kalman Filter and the second stage imposes a water balance constraint in an optimal manner. They have also included precipitation, evaporation, and runoff in their state vector and thereby used the filter to update these quantities. Pan and Wood (2006) showed that the constrained Kalman Filter gave estimates not far from the unconstrained filter, except that the water imbalance was removed.

In this study, it is shown that the constrained Kalman Filter can lead to very unrealistic state estimates. Specifically, if individual terms in the water budget have large errors, then imposing the budget to balance exactly requires these errors to be distributed among the state variables. If these errors are sufficiently large, then the budget constraint will cause some state variables to deviate beyond their natural range. There are at least two ways for dealing with large errors in the budget terms: include forcing terms in the data assimilation procedure, as showed by Pan and Wood (2006), or to impose a weak constraint in which the water budget derived from observed components is assumed to hold only approximately.

The purpose of this study is to present a weakly constrained data assimilation system in which a water budget constraint is imposed on the conventional data assimilation systems while taking into account the uncertainties of the water balance elements. Weakly constrained solutions were introduced for both the Ensemble Kalman Filter (EnKF) and Ensemble Transform Kalman Filter (ETKF). The weakly constrained Kalman Filter is applied to idealized experiments and its performance was compared to the unconstrained Kalman Filter. It is shown that the weakly constrained solution improves the water budget imbalance without increasing the errors of the hydrological variables.

This paper is organized as follows: section 2 briefly reviews the standard EnKF and ETKF; section 3 introduces the water budget constraint and its implementation in both filters; section 4 outlines the details of the synthetic experiments performed with and without the constraint; section 5 presents results from the experiments; section 6 summarizes the outcomes of the results; and appendix section presents the detailed derivation of the constrained filter.

2. Unconstrained Standard Filters

a. Kalman Filter

Complete derivations of Kalman Filter (KF), EnKF, and ETKF solutions can be found in numerous papers; here, these derivations are reviewed once more to emphasize the differences between the unconstrained and the constrained solutions.

The objective of data assimilation is to "optimally"estimate a set of quantities using all available observations, prior knowledge of the underlying model structure, and associated error statistics. In Kalman Filtering, the goal is to solve for the best state estimate and its uncertainty, where this best estimate and its error covariance information is propagated in time. This optimal estimate can be estimated by minimizing a cost function (Lorenc 1986),

$$\mathbf{J} = (\mathbf{o} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{o} - \mathbf{H}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_f)^T \mathbf{P}_f^{-1} (\mathbf{x} - \mathbf{x}_f), \tag{1}$$

where lower case letters represent vectors, capital letters represent matrices; $\mathbf{o}$ is the observations; $\mathbf{x}$ is the best estimate of the state to be found; $\mathbf{H}$ is a linear observation operator that maps the model state to observation space; superscript T is the transpose operator;

$\mathbf{R}$ is the observation error covariance matrix; $\mathbf{x}_f$ is the prior estimate of the model state, usually obtained from a model forecast; and $\mathbf{P}_f$ is the model background error covariance matrix. The first term on the right side of (1) measures the distance between the state and the observations, and the second term measures the distance between the state and the background. Both distances are measured using a norm based on the appropriate error covariance matrix. The vector $\mathbf{x}$ that minimizes (1) gives the best estimate according to maximum likelihood or Bayesian derivation methods (Maybeck 1982). The minimization of (1) can be obtained by setting the derivative of $\mathbf{J}$ w.r.t. $\mathbf{x}$ equal to 0 and solving

$$\frac{\partial \mathbf{J}}{\partial \mathbf{x}} = 2(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{P_f}^{-1})\mathbf{x} - 2(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + \mathbf{P_f}^{-1}\mathbf{x}_f) = 0.$$

The solution can be shown to be

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) \tag{2}$$

$$\mathbf{K} = \mathbf{P}_f\mathbf{H}^T(\mathbf{H}\mathbf{P}_f\mathbf{H}^T + \mathbf{R})^{-1} \tag{3}$$

where $\mathbf{x}_a$ is the updated state vector and $\mathbf{K}$ is the Kalman gain matrix. The analysis error covariance is given by

$$\mathbf{P}_a = \mathbf{P}_f - \mathbf{K}\mathbf{H}\mathbf{P}_f. \tag{4}$$

b. Ensemble Kalman Filter

In typical geophysical data assimilation, the KF is prohibitively expensive. Moreover, the background error covariance $\mathbf{P}_f$ is often unavailable due to its large dimension and/or the underlying model is nonlinear. To circumvent these problems, Evensen (1994) introduced the EnKF, whereby ensembles of realizations are created by Monte Carlo methods and carry

the error covariance information. Evensen (1994) proposed updating the individual ensemble members using the equation

$$\mathbf{x}_{ai} = \mathbf{x}_{fi} + \mathbf{K}(\mathbf{o} - \mathbf{H}\mathbf{x}_{fi})$$

where an i-index is included to identify the ensemble member. It proves convenient to collect the ensemble members into a single matrix as

$$\mathbf{X}_f = \frac{1}{\sqrt{(N-1)}}[\mathbf{x}_{f1} - \boldsymbol{\mu}_f, \mathbf{x}_{f2} - \boldsymbol{\mu}_f, \ \ldots, \mathbf{x}_{fN} - \boldsymbol{\mu}_f].$$

where $\boldsymbol{\mu}_f$ denotes the ensemble mean state vector, N is the ensemble size, and similarly for the update $\mathbf{X}_a$. In this notation, the analysis anomaly and the best estimate of the state update equations become

$$\mathbf{X}_a = \mathbf{X}_f + \mathbf{K}(\mathbf{O} - \mathbf{H}\mathbf{X}_f) \tag{5}$$

$$\boldsymbol{\mu}_a = \boldsymbol{\mu}_f + \mathbf{K}(\mathbf{o} - \mathbf{H}\boldsymbol{\mu}_f) \tag{6}$$

where $\mathbf{P}_f = \mathbf{X}_f\mathbf{X}_f^T$ is substituted in (3) and (4); and $\mathbf{O}$ is a matrix of perturbed observations in which each column is of the form $\mathbf{o} + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i$ is drawn from a normal distribution with 0 mean and covariance $\mathbf{R}$ (Burgers et al. 1998).

c. Ensemble Transform Kalman Filter

Bishop et al. (2001), Anderson (2001), and Whitaker and Hamill (2002) proposed alternative ensemble filtering method that avoided perturbed observations. These filters were shown to belong to a single family of filters called square root filters (Tippett et al. 2003). Just as a square root is not unique due to an ambiguity in sign, square root filters are

not unique due an ambiguity in a unitary transformation. Bishop et al. (2001) showed the analysis error covariance matrix (4) can be written as

$$\mathbf{P}_a = \mathbf{X}_f \mathbf{D} \mathbf{X}_f^T \tag{7}$$

where

$$\mathbf{D} = (\mathbf{I} + \mathbf{X}_f^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}_f)^{-1} \tag{8}$$

and (7) is also consistent with $\mathbf{P}_a = \mathbf{X}_a \mathbf{X}_a^T$.

The square root of $\mathbf{D}$ can be derived from the eigenvectors of $\mathbf{X}_f^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}_f$. Specifically, if the eigenvector decomposition of this matrix is expressed as

$$\mathbf{X}_f^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{X}_f = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$$

where $\mathbf{U}$ is unitary and $\mathbf{\Lambda}$ is a real positive diagonal matrix, then the most general square root of $\mathbf{D}$ is

$$\mathbf{A} = \mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{V}^T \tag{9}$$

where $\mathbf{V}$ is <u>any</u> unitary matrix and $\mathbf{A}\mathbf{A}^T = \mathbf{D}$. This expression allows us to write the updated analysis anomaly matrix as

$$\mathbf{X}_a = \mathbf{X}_f \mathbf{A}. \tag{10}$$

Finally, the Kalman Gain can be written as

$$\mathbf{K} = \mathbf{X}_f \mathbf{D} \mathbf{X}_f^T \mathbf{H}^T \mathbf{R}^{-1}.$$

It should be recognized that the square root matrix $\mathbf{A}$ depends on the choice of $\mathbf{V}^T$. In contrast, the matrices $\mathbf{P}_a$, $\mathbf{D}$, and $\mathbf{K}$ are independent of $\mathbf{V}^T$ and hence unique. Choosing

$\mathbf{V}^T = \mathbf{U}^T$ makes the square root matrix $\mathbf{A}$ symmetric. Ott et al. (2004) show that the quadratic form $(\mathbf{X}_a - \mathbf{X}_f)^T {\mathbf{P}_a}^{-1}(\mathbf{X}_a - \mathbf{X}_f)$, which is a measure for the magnitude of the analysis update, is also minimized if $\mathbf{A}$ is selected as the symmetric square root of $\mathbf{D}$ (which is unique). Accordingly, in the present study $\mathbf{V}^T$ is chosen to be $\mathbf{U}^T$.

Although both EnKF and ETKF have the same solution for $\mathbf{P}_a$, $\mathbf{D}$, $\mathbf{K}$, and $\boldsymbol{\mu}_a$ when starting with the same ensemble, they produce different ensemble anomalies − the EnKF produces the anomalies $\mathbf{X}_a$ as defined in (5), while the ETKF produces anomalies given in (10). The EnKF requires inverting the matrix $(\mathbf{H}\mathbf{P}_f\mathbf{H}^T + \mathbf{R})$, which is expensive for meteorological data assimilation applications, but relatively cheaper for land data assimilation applications when the simulations at different pixels are assumed uncoupled. In contrast, the ETKF requires calculating the eigenvector decomposition $(\mathbf{X}_f^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_f)$ and inverting the matrix $\mathbf{R}$, both of which are feasible for moderate ensemble sizes and diagonal $\mathbf{R}$.

3. Constrained Filter

a. Water Budget Constraint

In land data assimilation, assimilation of soil moisture (SM) results in an analysis update that does not conserve water. In this section, a water budget constraint is introduced to reduce the water imbalance. The water balance residual at time step t is

$$r_t = c_{sm}^T(SM_{at-1} - SM_{at}) + c_{cmc}(CMC_{at-1} - CMC_{at}) + c_{swe}(SWE_{at-1} - SWE_{at}) + c_p Pr_t - c_e Ev_t - c_r Rn_t \quad (11)$$

where SM is a 4-dimensional vector specifying the soil moisture in each of the 4 layers (assuming there are 4 soil layers); the scalar CMC specifies canopy moisture content; the scalar SWE specifies the snow water equivalent; the scalars Pr, Ev, and Rn specify the integrated precipitation, evapotranspiration, and runoff respectively, during the data assimilation window; prefactors c_{cmc}, c_{swe}, c_p, c_e, and c_r are constants for unit conversion; and subscript a denotes the analysis. Note that SM, CMC, and SWE are prognostic variables; Pr is a forcing variable; Ev, and Rn are diagnostic variables. It is of interest to write the residual equation as combination of state and non-state variables. For a given time step all terms in (11), except for the analysis states, are known. Hence, these water balance terms can be condensed into the form

$$r_t = \beta_t - \mathbf{c}_{\mathbf{x}}^T \mathbf{x}_t \tag{12}$$

where

$$\beta_t = c_r Pr_t - c_e Ev_t - c_r Rn_t + c_{sm}^T SM_{at-1} + c_{cmc} CMC_{t-1} + c_{swe} SWE_{t-1} \tag{13}$$

$$\mathbf{x}_t = [SM1_{at}, SM2_{at}, SM3_{at}, SM4_{at}, ST1_{at}, ST2_{at}, ST3_{at}, ST4_{at},$$
$$SkT_{at}, CMC_{at}, SWE_{at}]^T \tag{14}$$

where r_t is residual, β_t is a known constant that holds the residual terms involving non-prognostic variables; where $SM1_{at}$, $SM2_{at}$, $SM3_{at}$, and $SM4_{at}$ are the soil moistures in the four layers, $ST1_{at}$, $ST2_{at}$, $ST3_{at}$, and $ST4_{at}$ are the soil temperatures in four soil layers, and SkT_{at} is the skin temperature; and $\mathbf{c}_{\mathbf{x}}^T$ is the unit conversion vector, where temperature terms that are not part of water balance are weighted as zero (eg. assuming the units of SM, CMC, and SWE are same, then $\mathbf{c}_{\mathbf{x}}^T = [1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1]$) in order (12) to be consistent with (11).

Applying a strong constraint (i.e. forcing $r_t = 0$) would preserve the total amount of water in the water storage terms (soil moisture at different soil layers, canopy moisture content, and snow water content). In a system where precipitation, runoff, and evapotranspiration are not updated, the strongly constrained solution would redistribute the water between the storages and would preserve the total amount of water in the storage terms. However, the problem with enforcing a strong constraint is that the individual terms (including the non-storage terms) in the water budget have error, and the errors themselves are not conserved. Thus, it is inappropriate to force an imperfectly observed budget to be held exactly. One approach is to correct the forcing terms, as described by Pan and Wood (2006). Here a weak constraint is imposed, which accounts for uncertainty in the water budget itself.

One way to impose a residual constraint is to add another term to the cost function (1) of the form $\lambda * f(r)$, where λ is a Lagrange multiplier and $f(r)$ is a positive definite function of r. For a strongly constrained solution (as in Simon and Chia 2002) the weighting factor λ can be determined by setting the derivation of the chosen cost function w.r.t. to λ to 0 and solving. However, for a weakly constrained system, it is not clear how this λ should be selected. Here the penalizing function $f(r)$ is set to be $(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x})^2$, and the Lagrange multiplier λ is chosen as φ^{-1}, where φ is the error variance of β [how φ is calculated is given below in (16)]. Note that the Lagrange multiplier is objectively estimated (is not assumed a predetermined value or is not estimated through tuning a parameter). Hence the imposed constraint is of the form $(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x})^T\varphi^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x})$, and the cost function to be minimized is of the form

$$\mathbf{J}_c = (\mathbf{o} - \mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_f)^T{\mathbf{P}_\mathbf{f}}^{-1}(\mathbf{x} - \mathbf{x}_f) + (\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x})^T\varphi^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}) \quad (15)$$

where the constraint is conceived as a third penalization function which measures the degree of water imbalance ($r = 0$).

In the standard cost function (1), uncertainty of the observations and the forecast are represented with error covariance matrices of $\mathbf{R}$ and $\mathbf{P}_f$ respectively, that can be obtained from the ensemble of their anomalies for an ensemble filtering framework. Similarly the error variance (φ) of $\boldsymbol{\beta}$ in (15) can be obtained from the ensemble of realizations in the form

$$\varphi = \boldsymbol{\beta}^{'}\boldsymbol{\beta}^{'^{T}}/(N-1) \tag{16}$$

where $\boldsymbol{\beta}^{'}$ is a vector with dimension (1,N) that holds the ensemble anomaly of $\boldsymbol{\beta}$ (13), and it is trivially calculated from the ensemble of variables that are known.

b. Constrained Kalman Filter

The vector $\mathbf{x}$ that minimizes (15) can be found by setting the derivative of $\mathbf{J}_c$ with respect to $\mathbf{x}$ equal to 0 and solving. It is shown in the appendix (A9) that the constrained KF solution is

$$\mathbf{x}_{aa} = \mathbf{x}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) + \mathbf{P}_{aa}\mathbf{c}_\mathbf{x}\varphi^{-1}(\beta - \mathbf{c}_\mathbf{x}{}^T\mathbf{x}_f) \tag{17}$$

where $\mathbf{P}_{aa}$ is the analysis error covariance of the constrained filter which is given in Appendix (A13). Similar to EnKF solution, the Weakly Constrained Ensemble Kalman Filter (WCEnKF) solution is obtained by updating the ith ensemble member using (17) where perturbed observations ($\mathbf{o}^{'}$) are used instead of the observations ($\mathbf{o}$) to update the ith ensemble

member. The best estimate and the anomaly of the state for WCEnKF is found as

$$\boldsymbol{\mu}_{aa} = \boldsymbol{\mu}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\boldsymbol{\mu}_f) + \mathbf{P}_{aa}\mathbf{c}_\mathbf{x}\varphi^{-1}(\beta - \mathbf{c}_\mathbf{x}{}^T\boldsymbol{\mu}_f) \tag{18}$$

$$\mathbf{X}_{aa} = \mathbf{X}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{O}' - \mathbf{H}\mathbf{X}_f) + \mathbf{P}_{aa}\mathbf{c}_\mathbf{x}\varphi^{-1}(\mathbf{B}' - \mathbf{c}_\mathbf{x}{}^T\mathbf{X}_f). \tag{19}$$

Also, it is shown in the appendix (A17) that the anomaly of the state for the Weakly Constrained Ensemble Transform Kalman Filter (WCETKF) is of the form $\mathbf{X}_{aa} = \mathbf{X}_f\mathbf{A}_{aa}$ where $\mathbf{A}_{aa}$ is the symmetric square root of

$$\mathbf{D} = (\mathbf{I} + \mathbf{X}_f^T(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)\mathbf{X}_f)^{-1}.$$

The square root can be obtained from the eigenvector decomposition of $\mathbf{X}_f^T(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)\mathbf{X}_f$.

The above constrained KF solution can also be shown to approach the unconstrained standard KF solution as $\varphi \to \infty$ [see appendix, (A21)]. Moreover, the residual of the constrained filter is shown to be smaller than the residual of the standard filter [see appendix, (A23)]. It is also shown in the appendix (A21) that the constrained KF solution can be solved equivalently in two recursive filters:

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c}_\mathbf{x}(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a)$$

where $\mathbf{x}_a = \mathbf{x}_f + \mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) = \mathbf{x}_f + \mathbf{K}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)$ is the solution of the standard KF without the constraint. This solution implies that the constrained solution can be obtained by first calculating the solution ($\mathbf{x}_a$) for the standard KF, and then adjusting this solution to take into account the constraint by adding $[\mathbf{P}_a\mathbf{c}_\mathbf{x}(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a)]$. The single-stage and the two-stage solutions yield identical WCEnKF updates, but generally different WCETKF analysis anomaly updates due to the fact that the single- and two-stage

WCETKF equations are solved using two different matrix square roots for the same analysis error covariances.

A strongly constrained KF solution (A22) can be estimated by taking the limit $\varphi \to 0$.

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c}_\mathbf{x}(\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a)$$

where this solution is identical to the strongly constrained solution of (Simon and Chia 2002, eq. 25). Similar to the WCEnKF solution, a strongly constrained Ensemble Kalman Filter (SCEnKF) can be estimated by updating the each ensemble using the above equation with perturbed observations. Similarly, the analysis anomaly of a strongly constrained Ensemble Transform Kalman Filter (SCETKF) can be obtained by taking the limit $\varphi \to 0$ in (A26) as

$$\mathbf{X}_{aa} = \mathbf{X}_a - \mathbf{P}_a\mathbf{c}_\mathbf{x}\mathbf{c}_\mathbf{x}^T\mathbf{X}_a/\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}$$

which implies the adjustment term for the constraint in the second state is $\mathbf{P}_a\mathbf{c}_\mathbf{x}\mathbf{c}_\mathbf{x}^T\mathbf{X}_a/\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}$.

In the two-stage constrained filter, the **H**-operator does not even appear in the second stage, so the nonlinearity in **H** can be handled entirely in the first stage, which is identical to the traditional KF. In other words, nonlinearity in the observation error can be handled the same way it is handled in modern filters.

4. Synthetic Simulations

a. Experiment Setup

To illustrate the weakly constrained filters, synthetic experiments were performed using the Noah land surface model (Ek et al. 2003) version 2.7. The study area was chosen to

be Red Arkansas River Basin, US (between 32.0°N - 37.0°N and 96.0°W - 91.0°W) with 0.125°spatial resolution. There are total of 1521 pixels (39*39). The pixels are assumed to have uncorrelated errors. Simulations were performed between April - October 2006 (total 4500 hourly time steps) using hourly North America Land Data Assimilation (NLDAS; Cosgrove et al. 2003) forcing data (precipitation, pressure, relative humidity, wind speed, short wave and long wave radiation, and air temperature) which have 0.125°spatial resolutions. Model grid spatial resolutions were selected consistent with the NLDAS data native resolution, so that no averaging or downscaling was needed. The initial states were generated by running the land model for 10 years, but with repeating 2006 NLDAS forcing data in each of the 10 years where the state obtained after each year of simulation is used as an initial condition for the following year. The state obtained at the end of the 10th year were selected as the initial states for all simulations. Assimilation of observations are performed in warm climate, where the ensemble of model realizations are simulated starting from January to provide a smooth transition before the assimilation of observations. All initial states and the forcing data (air temperature, short and long wave radiations, and precipitation) were perturbed (as described below) to create the ensembles for all simulations. The "truth" run is identified as a single run of the model with unperturbed initial condition and forcing.

The experiments were based on a "perfect model"assumption in which the same model that generated the "truth"was used to generate the prior distribution. The observation operator $\mathbf{H}$ equals to the identity matrix. Initial states were perturbed using additive Gaussian noise [selected from normal distribution with (μ=0,σ=1K°) and (μ=0,σ=0.02%) for ST and SM respectively]. Forcing perturbation standard deviations were selected similar to the ones described in Reichle et al. (2008). Precipitation forcing was perturbed using multi-

plicative noise with a log-normal distribution (μ=1,σ=0.7); short-wave radiative forcing was perturbed using multiplicative noise with normal distribution [N(μ=1, σ=0.25)]; air temperature forcing and long-wave forcing data were perturbed using additive noises with normal distribution [N(μ=0, σ=2.5 K°) and N(μ=0, σ=10 $W.m^{-2}$) respectively]. The above perturbations are independent. The precipitation perturbation multiplication factor was limited between 0 and 4 where the actual precipitation value was further prevented to exceed the true precipitation value with ±5mm/hour in ensemble generation. The short-wave perturbation multiplication factor was limited between 0.2 and 1.8. Temperature and long-wave radiation perturbations were limited to ±4 times their respective standard deviations.

All forecasts were performed for an ensemble size of 50. Ensembles of Open loop simulations (through an ensemble of model simulations without the assimilation of observations, where the ensemble mean is the best estimate) were simulated using the same perturbed initial states and forcings as the assimilation experiments. Although it is not possible to directly measure the full SM and ST profiles with the current observation systems, there are many monitoring stations that provide in-situ deep soil layer variables (i.e. Oklahoma Mesonet Network). Hence, for the proof of concept, observations through the entire soil column were assimilated (not only the top layer). After open loop simulations were performed and their errors were calculated, observation perturbation variances were selected based on these open loop error variances in order to have comparable open loop and observation realizations. Accordingly, observations were created by adding zero mean Gaussian noise to the truth states for all four soil layers (ST and SM perturbation standard deviations were 0.40K°and 0.004% respectively for all four soil layers). Unconstrained and constrained simulations had the same forcing and initial state perturbations as the open loop.

b. Filter Performance Checks and Analysis of the Results

The simulations were performed for four filters (ETKF, EnKF, WCETKF, and WCEnKF), for two types of assimilated observations (all 4 layers of SM, or all 4 layers of SM and ST together), and for two state update frequencies (3-hourly or once a day) giving a total of 16 sets of experiments. Only the single-stage solutions were used for the constrained filters. State error and the water balance residual statistics were calculated for all 16 sets of experiments. The state error statistics were also calculated for the open loop simulations (open loop simulations have no water balance residual).

1) Innovation Statistics

If the assumptions on which the KF equations were derived are true, then the quadratic form $[(\mathbf{o}-\mathbf{Hx})^T(\mathbf{HP}_f\mathbf{H}^T+\mathbf{R})^{-1}(\mathbf{o}-\mathbf{Hx})]$ should have chi-squared distribution with d.o.f. equal to the size of the observation vector. This chi-squared statistic was calculated at each time step for each pixel and each experiment separately. The percentage of pixels that were within the 2.5 and 97.5 percentiles was calculated for each experiment separately. The 2.5 and 97.5 percentiles of a chi-square distribution are 0.484 and 11.14 for 4 d.of. (for SM only updated scenario); and 2.180 and 17.535 for 8 d.o.f. (for both SM and ST updated scenario).

2) State Errors

Updated states during the assimilation are SM (all 4 layers), ST (all 4 layers), SkT, CMC, and SWE, regardless of the observed variable that is assimilated (SM, or SM and ST). Due to the time interval selection (April-October, no snow), snow related variables were effectively not updated; hence snow related results were not investigated or presented in this study. Mean square error of ensemble means (MSE) for each of 10 states and for each of 16 experiments per pixel were calculated as

$$MSE_{s\ i\ lon\ lat} = \sum_{t} (\boldsymbol{\mu}_{s\ i\ lon\ lat\ t} - \mathbf{ts}_{s\ i\ lon\ lat\ t})^2/(tts - 1)$$

where μ is the ensemble mean state, $\mathbf{ts}$ is the true state, s is each state (total 11), i is each experiment (defined above, total 16 sets), lon is longitude pixel number (total 39), lat is latitude pixel number (total 39), t is each time step, and tts denotes the number of total time steps (total 4501) respectively. Resulting MSE values for each pixel and for all 4 soil layers were then averaged to a single number separately for ST and SM variables and for each experiment.

$$RMSE.SM_i = \sqrt{\sum_{sm}^{4}\sum_{lat}^{39}\sum_{lon}^{39} MSE_{sm\ i\ lon\ lat}/(4 * 39 * 39))}$$

$$RMSE.ST_i = \sqrt{\sum_{st}^{4}\sum_{lat}^{39}\sum_{lon}^{39} MSE_{st\ i\ lon\ lat}/(4 * 39 * 39))}$$

3) Water Balance Residual

The water balance residual was calculated for each ensemble member, at each time step, at each pixel in the study area, and each set of experiments (total 16, defined above). The

variance and the mean of the residuals were calculated using all time step and ensemble member values for each set of experiment and for each pixel in the study area as:

$$r_{i\ lon\ lat\ .\ t} = \sum_{n} r_{i\ lon\ lat\ n\ t}/N$$

$$r_{i\ lon\ lat\ .\ .} = \sum_{t} r_{i\ lon\ lat\ .\ t}/ats$$

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

where the "dot"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 time steps that the observations are assimilated (1500 and 187 for 3-hourly and daily update scenarios respectively), where only the residuals due to assimilation were included in the statistics. Then $\sigma^2 r_{i\ lon\ lat}$ and $r_{i\ lon\ lat\ .\ .}$ values were averaged over the study area into single number ($\sigma^2 r_{i\ .\ .}$ and $r_{i\ .\ .\ .\ .}$) for each experiment separately.

4) Total Column Water Change

Total column water content is defined as the summation of the total soil moisture content (mm) for all 4 soil layers at any given time where its change (due to integration of the model and assimilation of observations) is defined as,

$$\Delta WC_{i\ lon\ lat\ t} = \sum_{N} \sum_{d} (SM_{i\ lon\ lat\ n\ t-1\ d} - SM_{i\ lon\ lat\ n\ t\ d}) * Depth_d/N$$

$$\Delta WC_{i\ lon\ lat\ .} = \sum_{t} \Delta WC_{i\ lon\ lat\ t}/ats$$

$$\sigma^2 Wat_{i\ lon\ lat} = \sum_{t} (\Delta WC_{i\ lon\ lat\ t} - \Delta WC_{i\ lon\ lat\ .})^2/(ats-1)$$

where ΔWC is the total column water content change (mm), d is the soil layer identifier, and $\Delta WC_{i\ lon\ lat\ .}$ and $\sigma^2 Wat_{i\ lon\ lat}$ are the mean and the variance of the total column water change. Calculated $\sigma^2 Wat_{i\ lon\ lat}$ values are then averaged over the study area into a single variance for each experiment ($\sigma^2 Wat_{i\ .\ .}$). For daily update scenarios $\sigma^2 Wat_{i\ .\ .}$ variances, similar to residual variances, were calculated only for the time-steps of the assimilation updates.

5. Results

a. RMS Error of the States

The result of applying a strongly constrained EnKF for a single pixel located at 34.63°N and 94.75°W between May-Oct, 2006 with 3-hourly SM observations is shown in Fig. 1. This figure shows that the strongly constrained filter produces very unrealistic soil temperatures, in the sense that the estimates are well beyond the range of variability of the truth. It is plausible that the unrealistically large increments are caused by large errors in the forcing and observations– instead of "shrinking" the errors, the strongly constrained filter "spreads" the errors in the column in order to conserve the apparent total water balance. We say "apparent" because the water budget terms have errors, so the true water budget is not known. In effect, a strong constraint on the water budget assumes not only that water is conserved, but also that the errors in the budge terms are conserved– a dubious assumption. The strong constraint seems most appropriate when the errors in the forcing and observations are small. Hence the remaining constraint experiments were performed using single stage

weakly constrained filters (WCEnKF or WCETKF).

The RMSE of all assimilation experiments, observations, and the open loop runs are shown in Fig 2. In most cases, the RMSEs for the constrained filter were close (within 2%) to the RMSEs for the unconstrained filter. The RMSEs for the constrained filter can be larger than for the unconstrained filter, but in these cases the RMSEs still were much smaller than the RMSEs in observations or the open loop. Not surprisingly, the RMSEs of a variable were much smaller than those of the corresponding observations or the open loop, when observations of that variable were assimilated. However, if the observations of a variable were not assimilated, then the RMSE of that variable can be comparable to that of the open loop, indicating very little benefit from the filter. Three-hourly assimilation of observations has smaller RMSEs than the corresponding daily assimilation, but not by an order of magnitude (even though 3-hourly assimilation was 8 times more expensive than the daily assimilation). In general, the RMSEs for the EnKF, ETKF, WCEnKF, and WCETKF were comparable to each other.

Innovation statistics were analyzed for the filter performance. Observed innovations fell within the 95% confidence interval 92% to 95% of the time, and innovations for all state variables were temporally uncorrelated; both suggesting consistency with the underlying assumptions of the KF.

b. Water Balance Residual

Water balance residual variances for the 16 experiments using the single-stage filters are shown in Fig. 3a. In general, the time mean of the residuals differed only slightly between the

16 sets of experiments (results not shown), where the annual water budget is not conserved on average. The magnitude of the residual bias was orders magnitude smaller (∼2-3% for daily simulations) than the magnitude of the residual variance for all experiments.

Constrained filter residual variances were smaller than unconstrained filter residual variances over all pixels in the study area regardless of the update variable (SM alone, or SM and ST together), filter (WCETKF vs ETKF, or WCEnKF vs EnKF), or update frequency (3-hourly or daily) selection (Fig. 3a). The residual variances of the constrained filters were 14% to 44% less than those for the unconstrained filters.

Two-stage WCETKF was performed using the unconstrained ETKF square root for the first stage. Two-stage WCETKF has consistent tendency to have higher (but not significant) state errors than the single-stage WCETKF errors, whereas two-stage WCETKF residuals were almost identical with the single-stage WCETKF residuals (results not shown).

c. Total Column Water Content Change

Cross comparisons of the variances of the total column water content change were performed for the 16 sets of assimilation experiments, the truth, and the open loop simulations (Fig. 3b).

The water content change variance of the open loop simulations was slightly lower than that of truth simulations. The constrained assimilation experiments had 14%-33% smaller total column water change than the unconstrained experiments regardless of the assimilation frequency, observed variable, or the filter selection (Fig. 3b), supporting the above discussed residual results that the constrained filters were closer to the truth simulations with respect

to their closure of water cycling than unconstrained filters.

Total column water change in an assimilation experiment can be conceived as the summation of the true change plus the residual added due to the assimilation update. Comparison of the residual variances against total water change for the assimilation experiments indicates 70% of the total water change was due to the residual for daily assimilations where this ratio was around 30% for 3-hourly assimilation experiments (Fig. 3a and Fig. 3b); suggesting that in the absence of frequent observations the obtained total soil moisture content change is heavily affected from the residuals along with the true soil moisture change.

d. Sensitivity of φ

Estimation of φ in an objective way from the ensemble of realizations with the above described methodology (16) improves the residuals with little effect on the state errors. The effect of inflating (or deflating) the φ values and using constant φ values rather than being objectively estimated (16) scenarios were investigated (Fig. 4). These simulations were performed for 195 pixels located between 37.0°N – 36.375°N and 96.0°W – 91.0°W with 3-hourly SM and ST observations. An apparent trade-off was found between the state errors and the residuals: the more the φ values were deflated (constraint was applied stronger), the more the state errors were increased and the more the residuals were decreased (Fig. 4). Applying the constraint too strongly (with inflation factor of 0.01 or using constant 0.01 φ values) resulted in state errors equal to observation errors, suggesting no additional benefit from the filter. Applying the constraint even more strongly (with inflation factor of 0.00001 or using constant 0.00001 φ values) resulted in much higher errors than both the observations

and the open loop, which supports the strong constraint results presented above. On the other hand, applying the constraint too weakly (by inflating φ 20 times or using constant φ values of 20) resulted in residuals that are very close to residuals of the unconstrained simulations.

In this sensitivity study, the range of constant (tuned) φ values were chosen based on a priori information obtained from objective estimation (16). SM error–residual trade off performance of WCEnKF was better than the performance of WCETKF. Objective estimation of φ had same performance with the estimation through tuned φ for WCEnKF; whereas for WCETKF using tunable parameter gave better performance than objective estimation. Hence, in this study we conclude there is no universal solution in selection of tuning or inflating φ; for some filters tuning gives better, for some inflation avoids tuning φ.

Optimality of the constrained filter depends on the goal of the specific application; depending on the priority given to the state error or the residual error, φ can be inflated or deflated to improve one error while degrading the other one at a different magnitude (Fig. 4). In general, in hydrological studies, having smaller state error is generally preferred. From this point of view, smaller residuals can be obtained without degrading the state errors noticeably. For example, inflating φ values with factors of 0.50–0.75 gave almost the same state errors with the standard EnKF, while it reduces the residuals to less than half of the standard EnKF. Objectiveness of how a constant φ value can be selected is still questionable; however similar results can be obtained by tuning the φ values prior to the simulations. The objective selection of a tuned φ value or an inflation factor could be less of a problem for reanalysis type of studies; whereas for an operational platform, particularly in a changing system, the selection of tuned φ could be more critical.

6. Conclusions

In land data assimilation systems, the state updates produce a water budget imbalance, called a residual. In this study, a weakly constrained data assimilation solution was introduced to reduce the residual of standard EnKF (Evensen 1994) and ETKF (Bishop et al. 2001). The solutions of these filters for the optimum state estimation can be found by minimizing a cost function which penalizes both the model forecast and the observation errors weighted by their error uncertainty. Similarly, constrained filter solutions (WCEnKF and WCETKF) were derived by minimizing a cost function that is the summation of three terms that represent the forecast errors, observation errors, and the water budget imbalance. These solutions were shown to be obtained in a single stage or in two stages where the first stage is the standard solution and the second stage is the constrained filter update. Two stage solution was shown to be identical to the single stage solution for WCEnKF where the analysis anomaly solutions of WCETKF differ for single and two stage solutions.

The minimization of the constraint cost function requires uncertainty estimates for the water balance elements (φ). This φ term was estimated objectively from the ensembles. Optimality of φ was analyzed by inflating, deflating, and using constant values of φ and comparing the results of these analysis with the flow dependent method. Major outcomes of this study can be summarized as follows:

- In general, the constrained solution affected the state RMSE only slightly when compared to unconstrained solution: constrained filter errors were indistinguishable from the unconstrained filter errors for the majority of the experiments.

- There is little-to-no improvement in ST errors when only SM observations are assimi-

lated. There is also no improvement in SM errors and residuals when ST observations are also assimilated along with SM observations.

- Water balance residual variances of weakly constrained filters (WCEnKF and WCETKF) are smaller than that of unconstrained filters (ETKF or EnKF) regardless of the update frequency (daily or 3-hourly) or the assimilated variable (SM only, or SM and ST together) selection.

- There is no major difference found between single-stage WCETKF (with a symmetric square-root) and two-stage WCETKF (with symmetric square-root only in the first stage) when state errors and residuals are compared.

- Estimation of φ in an objective way (16) did not give smaller SM errors and residuals when φ values were selected as a constant.

As with the water balance, land surface models also conserve the energy balance, but an imbalance occurs during assimilation as a result of the temperature state update. Although an energy balance constraint was not performed, the solution implemented in this study for water balance residuals also can be used to reduce the energy balance residuals.

In general, data assimilation of hydrological states results in an inconsistency between the predicted diagnostic variables (i.e. evapotranspiration, runoff) and the updated prognostic variables. Diagnostic variables remain unaffected from the prognostic variable update in current hydrological data assimilation schemes; unaffected diagnostic variables and the updated prognostic variables are model predictions for two different initial conditions. A remedy can be obtained by also updating the diagnostic variables (eg. evapotranspiration and runoff) along with the prognostic variables, where the error covariances for the diagnostic variables

are estimated from the ensembles (Pan and Wood 2006). In this study, an idealized setup was used, where the model errors and the model parameterization errors were not taken into account. An alternative approach could have been a fraternal twin experiment, where the truths are generated in one model and the experiments are performed in another. In this study, flow dependent estimated φ did not give superior results over a constant value for φ. An alternative flow dependent methodology can be obtained where φ can be treated as a parameter to be optimized inside the KF and be solved simultaneously with the estimated state.

In this study the residuals of the standard data assimilation techniques were reduced with a constrained filter. The constrained solution introduced in this study could be very valuable to GEWEX community to obtain a better water and energy cycling information as this study lays a solution to reduce the uncertainty of the water and potentially energy budgets. In general, reanalysis data are used to obtain better analysis of historical data that were not available in the past; NCEP reanalysis (Kalnay et al. 1996) is one of the early examples that produced 40 years of global atmospheric data. Data assimilation offers the ideal platform for reanalysis type of studies as new methods emerge. The introduced weakly constrained filter in this study could be used in reanalysis type of studies to acquire improved water and energy cycles. Weakly constrained assimilation can make the reanalysis products more valuable to the same community without making it less valuable to another community.

Acknowledgments.

We thank Bart Forman for constructive comments as a reviewer, which led numerous clarifications in the final version of the manuscript. This work was partially supported by NASA fellowship NESSF 09-Earth09R-80.

APPENDIX

Constrained Filter

a. Single-Stage Constrained Filter

1) Single-Stage Constrained Kalman Filter

Similar to the traditional KF, a constrained KF solution can be also obtained through minimizing a cost function given in (15)

$$\mathbf{J}_c = (\mathbf{o}-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{H}\mathbf{x}) + (\mathbf{x}-\mathbf{x}_f)^T{\mathbf{P_f}}^{-1}(\mathbf{x}-\mathbf{x}_f) + (\beta-\mathbf{c}_\mathbf{x}^T\mathbf{x})^T\varphi^{-1}(\beta-\mathbf{c}_\mathbf{x}^T\mathbf{x}) \quad \text{(A1)}$$

$$\frac{\partial \mathbf{J}}{\partial \mathbf{x}} = 2(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + {\mathbf{P_f}}^{-1} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)\mathbf{x} - 2(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + {\mathbf{P_f}}^{-1}\mathbf{x}_f + \mathbf{c}_\mathbf{x}\varphi^{-1}\beta) \quad \text{(A2)}$$

Setting derivation (A2) to 0, the solution for the constrained filter can be expressed as

$$\mathbf{x}_{aa} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + {\mathbf{P_f}}^{-1} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)^{-1}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + {\mathbf{P_f}}^{-1}\mathbf{x}_f + \mathbf{c}_\mathbf{x}\varphi^{-1}\beta) \quad \text{(A3)}$$

This equation can be used as a final solution to the constrained KF. However, the analogy with the standard KF is not obvious. Below, a constrained KF filter solution analogous to the standard solution was derived.

To ease the notation, we define $\mathbf{S}^{-1} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T$, then (A3) becomes:

$$\mathbf{x}_{aa} = ({\mathbf{P_f}}^{-1} + \mathbf{S}^{-1})^{-1}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + {\mathbf{P_f}}^{-1}\mathbf{x}_f + \mathbf{c}_\mathbf{x}\varphi^{-1}\beta) \quad \text{(A4)}$$

The notation was eased further by using the second derivation of the cost function, which is equal to the inverse of the analysis error covariance matrix (Lorenc 1986) of the constrained

filter,

$$\frac{\partial^2 J}{\partial \mathbf{x}^2} = \mathbf{P}_{aa}^{-1} = {\mathbf{P_f}}^{-1} + \mathbf{S}^{-1}$$

$$\mathbf{P}_{aa} = ({\mathbf{P_f}}^{-1} + \mathbf{S}^{-1})^{-1}. \tag{A5}$$

Hence, above equation (A4) can be rewritten as

$$\mathbf{x}_{aa} = \mathbf{P}_{aa}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + {\mathbf{P_f}}^{-1}\mathbf{x}_f + \mathbf{c_x}\varphi^{-1}\beta)$$

$$\mathbf{x}_{aa} = \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + \mathbf{P}_{aa}{\mathbf{P_f}}^{-1}\mathbf{x}_f + \mathbf{P}_{aa}\mathbf{c_x}\varphi^{-1}\beta. \tag{A6}$$

Before continuing the derivation from (A6), another equality is introduced

$$\mathbf{P}_{aa} = ({\mathbf{P_f}}^{-1} + \mathbf{S}^{-1})^{-1}$$

$$\mathbf{P}_{aa}({\mathbf{P_f}}^{-1} + \mathbf{S}^{-1}) = \mathbf{I}$$

$$\mathbf{P}_{aa}{\mathbf{P_f}}^{-1} = \mathbf{I} - \mathbf{P}_{aa}\mathbf{S}^{-1}. \tag{A7}$$

Using this equality in (A7), (A6) can be rewritten as

$$\begin{aligned}\mathbf{x}_{aa} &= \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} + (\mathbf{I} - \mathbf{P}_{aa}\mathbf{S}^{-1})\mathbf{x}_f + \mathbf{P}_{aa}\mathbf{c_x}\varphi^{-1}\beta \qquad \text{(A8)}\\ &= \mathbf{x}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{o} - \mathbf{P}_{aa}(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)\mathbf{x}_f + \mathbf{P}_{aa}\mathbf{c_x}\varphi^{-1}\beta\\ &= \mathbf{x}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) + \mathbf{P}_{aa}(\mathbf{c_x}\varphi^{-1}\beta - \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T\mathbf{x}_f)\end{aligned}$$

and the final constrained KF solution is obtained as

$$\mathbf{x}_{aa} = \mathbf{x}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) + \mathbf{P}_{aa}\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f) \tag{A9}$$

where the analogy between the constrained KF and the standard KF solutions becomes more clear when the standard KF solution [$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)$] is equivalently written in the form $\mathbf{x}_a = \mathbf{x}_f + \mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)$.

The best estimate and the analysis anomaly of WCEnKF is found as

$$\boldsymbol{\mu}_{aa} = \boldsymbol{\mu}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\boldsymbol{\mu}_f) + \mathbf{P}_{aa}\mathbf{c}_\mathbf{x}\varphi^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\boldsymbol{\mu}_f) \tag{A10}$$

$$\mathbf{X}_{aa} = \mathbf{X}_f + \mathbf{P}_{aa}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{O}' - \mathbf{H}\mathbf{X}_f) + \mathbf{P}_{aa}\mathbf{c}_\mathbf{x}\varphi^{-1}(\mathbf{B}' - \mathbf{c}_\mathbf{x}^T\mathbf{X}_f). \tag{A11}$$

where $\mathbf{O}'$ and $\mathbf{B}'$ are matrices holding the observation anomalies (namely random numbers used for the perturbations) and the constraint anomalies respectively.

Solution of the standard KF requires a single inverse, computation of the Kalman gain. Similarly, the solution of the constrained filter can be obtained with a single inverse, through the computation of $\mathbf{P}_{aa}$,

$$\mathbf{P}_{aa} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + {\mathbf{P}_\mathbf{f}}^{-1} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)^{-1} \tag{A12}$$

$$\mathbf{P}_{aa} = \mathbf{P}_f\big(\mathbf{I} + (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)\mathbf{P}_f\big)^{-1} \tag{A13}$$

provided that the observation error covariance matrix ($\mathbf{R}$) is assumed diagonal, hence its inverse is trivial.

Whitaker and Hamill (2002) showed that without the perturbation of observations, the analysis error covariance of EnKF is underestimated by a term of $\mathbf{KRK}^T$. The term β holds the prognostic variables of the previous time-step analysis, fluxes, and the forcing data, where β is obtained from ensembles ($\mathbf{B}' \neq 0$). Hence, construction of perturbed constraints is not needed for the constrained filters.

2) Single-Stage Constrained Ensemble Transform Kalman Filter

The best estimate of the state for the WCETKF is found using (A10). Similar to the traditional ETKF solutions, analysis anomaly solution of WCETKF can also be obtained by

using analysis error covariance matrix of the constrained filter.

$$\begin{aligned}
\mathbf{P}_{aa} &= (\mathbf{P_f}^{-1} + \mathbf{S}^{-1})^{-1} \\
&= \mathbf{P}_f \mathbf{P_f}^{-1} (\mathbf{P_f}^{-1} + \mathbf{S}^{-1})^{-1} \\
&= \mathbf{P}_f (\mathbf{P_f}^{-1} \mathbf{P}_f + \mathbf{S}^{-1} \mathbf{P}_f)^{-1} \\
&= \mathbf{P}_f (\mathbf{I} + \mathbf{S}^{-1} \mathbf{P}_f)^{-1} \\
&= \mathbf{X}_f \mathbf{X}_f^T (\mathbf{I} + \mathbf{S}^{-1} \mathbf{X}_f * \mathbf{I} * \mathbf{X}_f^T)^{-1} \qquad \text{(A14)}
\end{aligned}$$

Using the Sherman-Morrison-Woodbury formula [a reminder for the reader $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$, (A14) can be rewritten as

$$\begin{aligned}
\mathbf{P}_{aa} &= \mathbf{X}_f \mathbf{X}_f^T [\mathbf{I} - \mathbf{I} * \mathbf{S}^{-1} \mathbf{X}_f (\mathbf{I} + \mathbf{X}_f^T * \mathbf{I} * \mathbf{S}^{-1} \mathbf{X}_f)^{-1} \mathbf{X}_f^T * \mathbf{I}] \\
&= \mathbf{X}_f [\mathbf{I} - \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f (\mathbf{I} + \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f)^{-1}] \mathbf{X}_f^T \\
&= \mathbf{X}_f [(\mathbf{I} + \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f - \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f)(\mathbf{I} + \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f)^{-1}] \mathbf{X}_f^T \\
\mathbf{P}_{aa} &= \mathbf{X}_f (\mathbf{I} + \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f)^{-1} \mathbf{X}_f^T \qquad \text{(A15)} \\
\mathbf{P}_{aa} &= \mathbf{X}_f \mathbf{D} \mathbf{X}_f^T \qquad \text{(A16)}
\end{aligned}$$

where $\mathbf{D} = (\mathbf{I} + \mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f)^{-1}$. Using eigenvalue decomposition of $\mathbf{X}_f^T \mathbf{S}^{-1} \mathbf{X}_f$ ($\mathbf{U}$ is eigenvectors and $\mathbf{\Lambda}$ is diagonal) and defining its square root as $\mathbf{D} = \mathbf{A}_{aa} \mathbf{A}_{aa}^T$, this square root can be found $\mathbf{A}_{aa} = \mathbf{U}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{V}^T$, where $\mathbf{V}^T$ is unitary. These equalities imply the solution for the anomaly of the analysis for the constrained filter can be rewritten as

$$\mathbf{X}_{aa} = \mathbf{X}_f \mathbf{A}_{aa} \qquad \text{(A17)}$$

where this solution is also consistent with (A16).

b. Two-Stage Constrained Filter

In this section it is shown that the single-stage solution in (A9) can equivalently be performed in two-recursive stages where the first stage is the standard KF equations and the second stage is the constrained filter adjustment.

1) Two-Stage Constrained Kalman Filter

Expending the terms, (A9) becomes

$$\mathbf{x}_{aa} = \mathbf{x}_f + (\mathbf{P_f}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) + (\mathbf{P_f}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)^{-1}\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f). \quad \text{(A18)}$$

Substituting inverse of the standard KF analysis error covariance $\mathbf{P}_a^{-1} = \mathbf{P_f}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, (A18) becomes

$$\mathbf{x}_{aa} = \mathbf{x}_f + (\mathbf{P}_a{}^{-1} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) + (\mathbf{P}_a{}^{-1} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)^{-1}\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f) \quad \text{(A19)}$$

Using the Sherman-Morrison-Woodbury formula,$(\mathbf{P}_a{}^{-1} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)^{-1} = \mathbf{P}_a - \mathbf{P}_a c_\mathbf{x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}\mathbf{c_x}^T\mathbf{P}_a$, (A19) becomes

$$\mathbf{x}_{aa} = \mathbf{x}_f + \big(\mathbf{P}_a - \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}\mathbf{c_x}^T\mathbf{P}_a\big)\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) +$$

$$\big(\mathbf{P}_a - \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}\mathbf{c_x}^T\mathbf{P}_a\big)\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}{}^T\mathbf{x}_f)$$

$$\mathbf{x}_{aa} = \mathbf{x}_f + \big[\mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) - \mathbf{P}_a c_\mathbf{x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}\mathbf{c_x}^T\mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)\big] +$$

$$\big[\mathbf{P}_a\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f) - \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}\mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f)\big]$$

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c_x}\big[\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f) - (\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}$$

$$\mathbf{c_x}^T\mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) - (\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}\mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f)\big]$$

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}$$

$$\big[(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f) - \mathbf{c_x}^T\mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) - \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f)\big]$$

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}$$

$$\big[(\beta - \mathbf{c_x}^T\mathbf{x}_f) + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f) - \mathbf{c_x}^T\mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f) - \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x}\varphi^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_f)\big]$$

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}\big[(\beta - \mathbf{c_x}^T\mathbf{x}_f) - \mathbf{c_x}^T\mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)\big]$$

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}\big[\beta - \mathbf{c_x}^T\big(\mathbf{x}_f - \mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)\big)\big] \tag{A20}$$

The two-stage solution can be written as

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a c_\mathbf{x})^{-1}(\beta - \mathbf{c_x}^T\mathbf{x}_a) \tag{A21}$$

where $\mathbf{x}_a = \mathbf{x}_f + \mathbf{P}_a\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o} - \mathbf{H}\mathbf{x}_f)$ is the standard KF solution without the constraint. Above equation (A21) can also be used for the best estimate and the analysis anomaly solution of the two-stage WCEnKF. It is also noted that for $(lim_\varphi \to \infty)$, the second term in (A21) vanishes, and the constrained filter solution equals to the standard KF solution.

Moreover, setting $\varphi = 0$ in (A21), strongly constrained KF solution is obtained as

$$\mathbf{x}_{aa} = \mathbf{x}_a + \mathbf{K}_s(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a) \tag{A22}$$

where $\mathbf{K}_s = \mathbf{P}_a\mathbf{c}_\mathbf{x}(\mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}$. This strongly constrained solution in (A22) is identical with the maximum probability method constrained solution of (Simon and Chia 2002, eq. 25).

The two-stage solution (A21) implies that the constraint can be performed in two sequential stages: the first stage is the standard KF ($\mathbf{x}_a$) without the constraint and the second stage is the constrained filter $\mathbf{P}_a\mathbf{c}_\mathbf{x}(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a)$.

A comparison of the residual terms $(\beta - \mathbf{c}_\mathbf{x}^T x)$ of the constrained and standard filters can be performed using the two-stage solution in (A21):

$$\begin{aligned}
\mathbf{x}_{aa} &= \mathbf{x}_a + \mathbf{P}_a\mathbf{c}_\mathbf{x}(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a) \\
\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_{aa} &= \beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a - \mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a) \\
\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_{aa} &= [\mathbf{I} - \mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}](\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a) \\
\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_{aa} &= \left[(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x}) - \mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}\right](\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a) \\
\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_{aa} &= \varphi(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a c_\mathbf{x})^{-1}(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a)
\end{aligned} \tag{A23}$$

For scalar $\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x} > 0$, $\varphi(\varphi + \mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x})^{-1} < 1$ therefore $(\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_{aa}) < (\beta - \mathbf{c}_\mathbf{x}^T\mathbf{x}_a)$; hence the constraint shrinks the residual of KF toward zero by a rate that depends on φ.

2) Two-stage Constrained Ensemble Transform Kalman Filter

Similar to the two stage WCEnKF, the best estimate of the two-stage WCETKF can be found using two-stage constrained KF solution (A21). Two stage solution of the state anomalies for the WCETKF can be found using the inverse of the analysis error covariance

of the constrained filter. Using inverse of (A12)

$$\mathbf{P}_{aa}{}^{-1} = \mathbf{P_f}{}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T$$

$$\mathbf{P}_{aa}{}^{-1} = \mathbf{P}_a{}^{-1} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T$$

$$\mathbf{P}_{aa} = (\mathbf{P}_a{}^{-1} + \mathbf{c_x}\varphi^{-1}\mathbf{c_x}^T)^{-1}.$$

Using the Sherman-Morrison-Woodbury formula,

$$\mathbf{P}_{aa} = \mathbf{P}_a - \mathbf{P}_a\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}\mathbf{c_x}^T\mathbf{P}_a$$

$$\mathbf{X}_{aa}\mathbf{X}_{aa}{}^T = \mathbf{X}_a(\mathbf{I} - \mathbf{X}_a{}^T\mathbf{c_x}(\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}\mathbf{c_x}^T\mathbf{X}_a)\mathbf{X}_a{}^T$$

$$\mathbf{X}_{aa}\mathbf{X}_{aa}{}^T = \mathbf{X}_a(\mathbf{I} - \mathbf{z}\alpha\mathbf{z}^T)\mathbf{X}_a{}^T$$

where $\mathbf{X}_{aa}$ is the analysis anomaly of the constrained filter, $\alpha = (\varphi + \mathbf{c_x}^T\mathbf{P}_a\mathbf{c_x})^{-1}$ is a scalar, and $\mathbf{z} = \mathbf{X}_a{}^T\mathbf{c_x}$. A square root can be found analytically by finding a scalar (δ) such that

$$(\mathbf{I} + \delta\mathbf{z}\mathbf{z}^T)(\mathbf{I} + \delta\mathbf{z}\mathbf{z}^T)^T = \mathbf{I} - \mathbf{z}\alpha\mathbf{z}^T$$

and rearranging the terms on both sides as

$$(\gamma\delta^2 + 2\delta + \alpha)\mathbf{z}\mathbf{z}^T = 0$$

where $\gamma = \mathbf{z}^T\mathbf{z}$ is a scalar and the solution is found as

$$\delta_\pm = \frac{-1 \pm \sqrt{1 - \alpha\gamma}}{\gamma}$$

This quadratic form gives two solutions, but only one of them produces a positive definite square root. To determine the correct choice, we choose the solution that renders

$$\mathbf{z}^T(\mathbf{I} + \delta_\pm z z^T)\mathbf{z} > 0$$

where the above quadratic form checks for the positive definiteness of $(\mathbf{I}+\delta_{\pm}\mathbf{z}\mathbf{z}^T)$ for vector $\mathbf{z}$. Rearranging the above equation and replacing $\delta_{\pm}$

$$\mathbf{z}^T\mathbf{z}+\delta_{\pm}\mathbf{z}^T\mathbf{z}\mathbf{z}^T\mathbf{z}>0$$

$$\gamma(1-1\pm\sqrt{1-\alpha\gamma})>0$$

$$\pm\sqrt{1-\alpha\gamma}>0$$

Hence the positive root is selected:

$$\begin{aligned}\mathbf{X}_{aa}&=\mathbf{X}_a(\mathbf{I}+\delta_{+}\mathbf{z}\mathbf{z}^T)\\&=\mathbf{X}_a\Big[+\frac{-1+\sqrt{1-(\varphi+\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x})^{-1}\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}}}{\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}}{\mathbf{X}_a}^T\mathbf{c}_\mathbf{x}\mathbf{c}_\mathbf{x}^T\mathbf{X}_a\Big]\end{aligned}$$

The final solution for the two-stage WCETKF analysis anomaly can be found as

$$\mathbf{X}_{aa}=\mathbf{X}_a\Big[\mathbf{I}+{\mathbf{X}_a}^T\mathbf{c}_\mathbf{x}\mathbf{c}_\mathbf{x}^T\mathbf{X}_a\Big(-1+\sqrt{\varphi(\varphi+\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x})^{-1}}\Big)/\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x}\Big] \tag{A24}$$

which can be also rewritten as

$$\mathbf{X}_{aa}=\mathbf{X}_a\mathbf{E}=\mathbf{X}_f\mathbf{A}\mathbf{E} \tag{A25}$$

where $\mathbf{A}$ is the square-root multiplier matrix that is estimated from the standard ETKF equations and $\mathbf{E}$ is the matrix obtained from the operations within the square-brackets on the rhs of (A24). This equation implies the two-stage analysis anomaly of WCETKF ($\mathbf{X}_{aa}$) can be obtained by first solving for the analysis anomaly of the standard ETKF ($\mathbf{X}_a$) and then multiplying it by the matrix $\mathbf{E}$.

Similar to the strongly constrained KF solution, a strongly constrained ETKF solution can be estimated by setting φ in (A24) into 0 as

$$\mathbf{X}_{aa}=\mathbf{X}_a-\mathbf{P}_a\mathbf{c}_\mathbf{x}\mathbf{c}_\mathbf{x}^T\mathbf{X}_a/\mathbf{c}_\mathbf{x}^T\mathbf{P}_a\mathbf{c}_\mathbf{x} \tag{A26}$$

It is stressed that the WCETKF analysis anomaly of single-stage (A17) and two-stage (A24) solutions differ, although they have identical solutions for the analysis error covariance matrix $\mathbf{P}_{aa}$. In fact, these single-stage and two-stage solutions are two different square root filters with the same error covariance matrices but with different state analysis anomalies. It is fairly easy to make the single-stage WCETKF square-root $\mathbf{A}_{aa}$ (A17) symmetric with the selection of $\mathbf{V}^T = \mathbf{U}$; whereas for the two-stage filter, it is not immediately clear which selection for the $\mathbf{V}^T$ matrix would make the $\mathbf{AE}$ term in (A25) symmetric. On the other hand, it is emphasized that WCEnKF solutions are identical for both single-stage (A9) and two-stage (A21) constrained filters.

Computationally, both standard (2) and two-stage constrained (A21) KF solutions require single inverse $(\mathbf{HP}_f\mathbf{H}^T + \mathbf{R})$, where the single-stage constrained KF solution (A9) requires two inverses $[(\mathbf{I} + \mathbf{S}^{-1}\mathbf{P}_f)$ and $\mathbf{R}]$. Although the inverse of $\mathbf{R}$ can be avoided by a diagonal observation error covariance matrix assumption, the dimension of the term to be inverted is higher for the single-stage constrained KF solution than it is for other two solutions (assuming not all state variables are observed). Hence computationally, the two-stage solution is similar to the standard KF whereas the single-stage KF solution is more expensive. The load for the square root filters is the same for all standard ETKF, single-stage WCETKF, and two-stage WCETKF solutions. They all require single inverse $(\mathbf{R})$ and single eigenvalue decomposition. Standard ETKF and two-stage WCETKF solutions require the eigenvalue decomposition of the term $\mathbf{X}_f^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_f$ (8); this term for the single-stage WCETKF solution is $\mathbf{X}_f^T(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{c}_\mathbf{x}\varphi^{-1}\mathbf{c}_\mathbf{x}^T)\mathbf{X}_f$ (A15).

REFERENCES

Alsdorf, D. E., E. Rodriguez, and D. P. Lettenmaier, 2007: Measuring surface water from space. *Rev. Geophys.*, **45 (2)**, RG2002.

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Monthly Weather Review*, **129**, 2884–2903.

Bishop, C. H., B. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review*, **129**, 420–436.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, **126**, 1719–1724.

Cosgrove, B. A., et al., 2003: Real-time and retrospective forcing in the north american land data assimilation system (NLDAS) project. *Journal of Geophysical Research*, **108 (D22)**, 8842.

Crow, W. T. and D. Ryu, 2009: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrol. Earth Syst. Sci.*, **13**, 1–16.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, K. R. Arsenault, N. E. C. Verhoest, and V. R. N. Pauwels, 2010: Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. *Journal of Hydrometeorology*, **11 (2)**, 352–369.

DelSole, T., M. Zhao, and P. Dirmeyer, 2009: A new method for exploring coupled land-atmosphere dynamics. *Journal of Hydrometeorology*, **10 (4)**, 1040–1050.

Dirmeyer, P. A., 2003: The role of the land surface background state in climate predictability. *Journal of Hydrometeorology*, **4**, 599–610.

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. *Journal of Geophysical Research*, **108 (D22)**, 8851.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99 (C5)**, 10 143–10 162.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, **34 (12)**, 3405–3420.

Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77 (3)**, 437–471.

Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009: Role of subsurface physics in the assimilation of surface soil moisture observations. *Journal of Hydrometeorology*, **10 (6)**, 1534–1547.

Lakshmi, V., 2000: A simple surface temperature assimilation scheme for use in land surface models. *Water Resources Research*, **36 (12)**, 3687–3700.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112**, 1177–1194.

Maybeck, P. S., 1982: *Stochastic models, estimation, and control*, Vol. 2. Academic Press.

Ott, E., et al., 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **A56**, 415–428.

Pan, M. and E. F. Wood, 2006: Data assimilation for estimating the terrestrial water budget using a constrained ensemble kalman filter. *Journal of Hydrometeorology*, **7 (3)**, 534–547.

Pauwels, V. R. N., R. Hoeben, N. E. C. Verhoest, F. P. De Troch, and P. A. Troch, 2002: Improvement of toplats-based discharge predictions through assimilation of ers-based remotely sensed soil moisture values. *Hydrological Processes*, **16**, 995–1013.

Reichle, R. H., W. T. Crow, and C. L. Keppenne, 2008: An adaptive ensemble kalman filter for soil moisture data assimilation. *Water Resources Research*, **44 (3)**, W03 423.

Roads, J., et al., 2003: GCIP water and energy budget synthesis (WEBS). *Journal of Geophysical Research*, **108 (D16)**, 8609.

Simon, D. and T. L. Chia, 2002: Kalman filtering with state equality constraints. *IEEE Transactions on Aerospace and Electronic Systems*, **38 (2)**, 128–136.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square-root filters. *Monthly Weather Review*, **131**, 1485–1490.

WCRP JSC Report, 2010: WCRP joint scientific committee 31st session report. Tech. rep., WCRP.

Wei, J., P. A. Dirmeyer, and Z. Guo, 2010: How much do different land models matter for climate simulation? part ii: A decomposed view of the landatmosphere coupling strength. *Journal of Climate*, **23 (11)**, 3135–3145.

Whitaker, J. and T. M. Hamill, 2002: Ensemble Data Assimilation Without Perturbed Observations. *Monthly Weather Review*, **130**, 1913–1924.

Zaitchik, B. F., M. Rodell, and R. H. Reichle, 2008: Assimilation of grace terrestrial water storage data into a land surface model: Results for the mississippi river basin. *Journal of Hydrometeorology*, **9 (3)**, 535–548.

List of Figures

4 SM error and residual relation for varying Phi values, where 3-hourly SM and ST observations are assimilated using 50 ensemble members. Each line represents series of simulations using 24 different constant φ values or 24 different φ inflation factors (Both inflation and constant values were selected as 0.00001, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 1.0, 1.2, 1.5, 2, 5, 10, 20, 50, 100, 200, 500, 1000). Single points represent single simulations of constrained filters with un-inflated φ values or of unconstrained filters. For both constant and inflated φ experiments, higher residuals are result of higher φ values and lower residuals are result of lower φ values (Inflation or constant φ values increase from left to right for green and red lines). Observation error is also marked with a black diamond. The residuals and the errors asymptotically approached to those of unconstrained simulations or strongly constrained simulations as the inflation factors for φ or the constant φ value was increased to ∞ or decreased to 0 respectively. 48

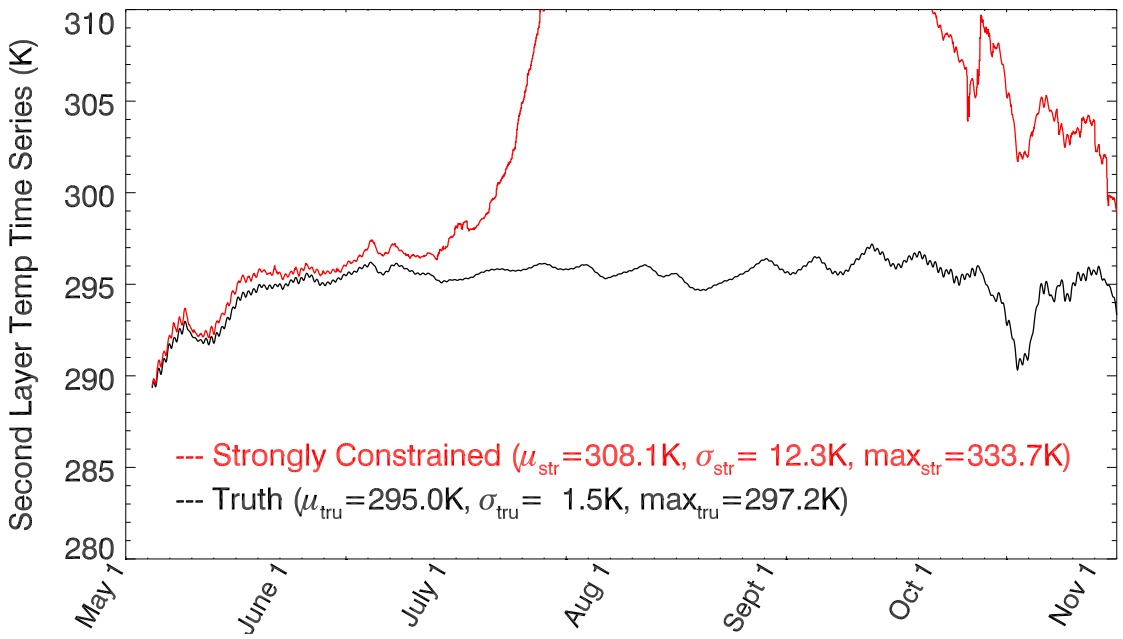

FIG. 1. Second soil layer temperature errors of strongly constrained EnKF simulations and truth run.

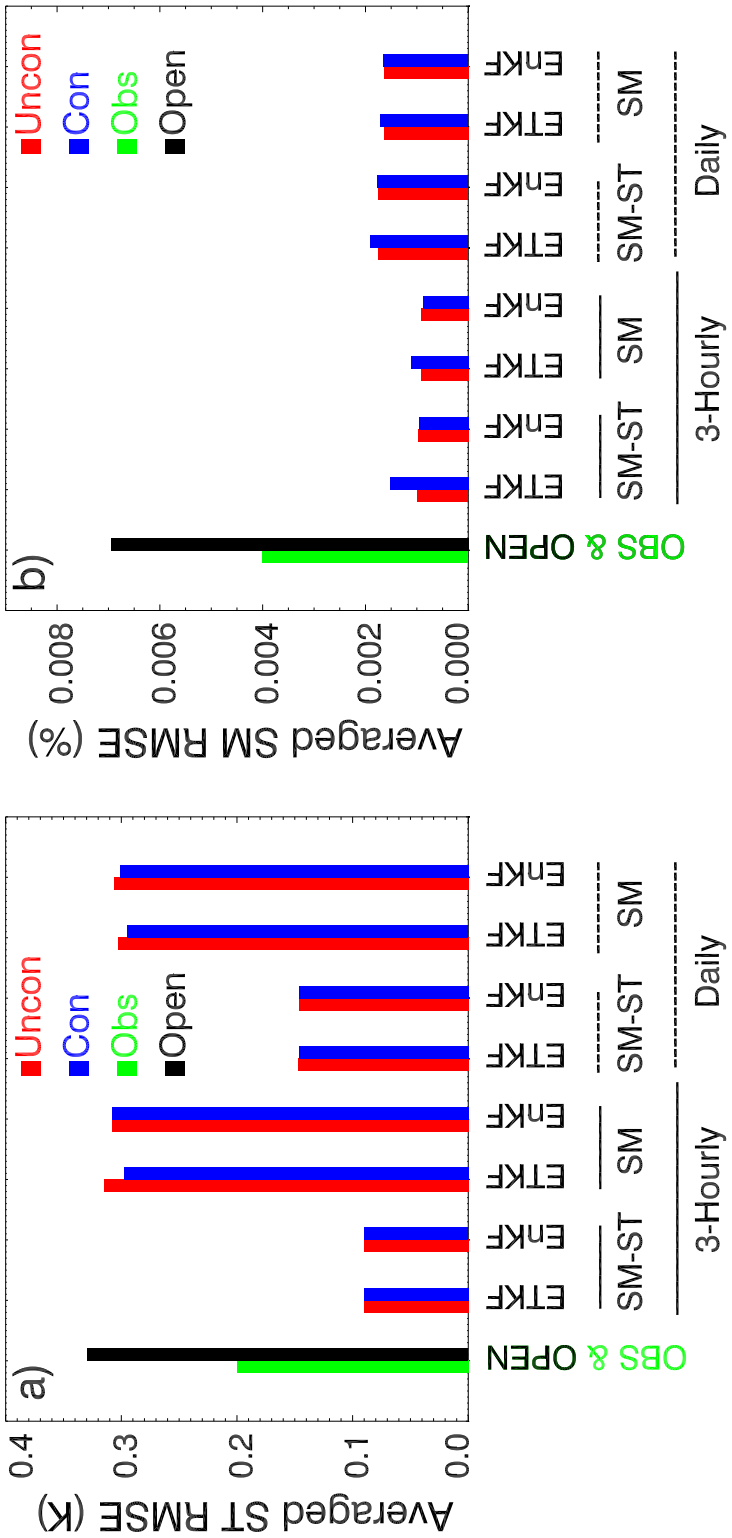

FIG. 2. (a) Soil temperature and (b) soil moisture errors averaged across soil layers. Horizontal axis: "OBS" refers to observation errors (in green color); "OPEN" refer to open loop errors (in black color); and "Uncon" and "Con" refer to unconstrained and constrained filters respectively.

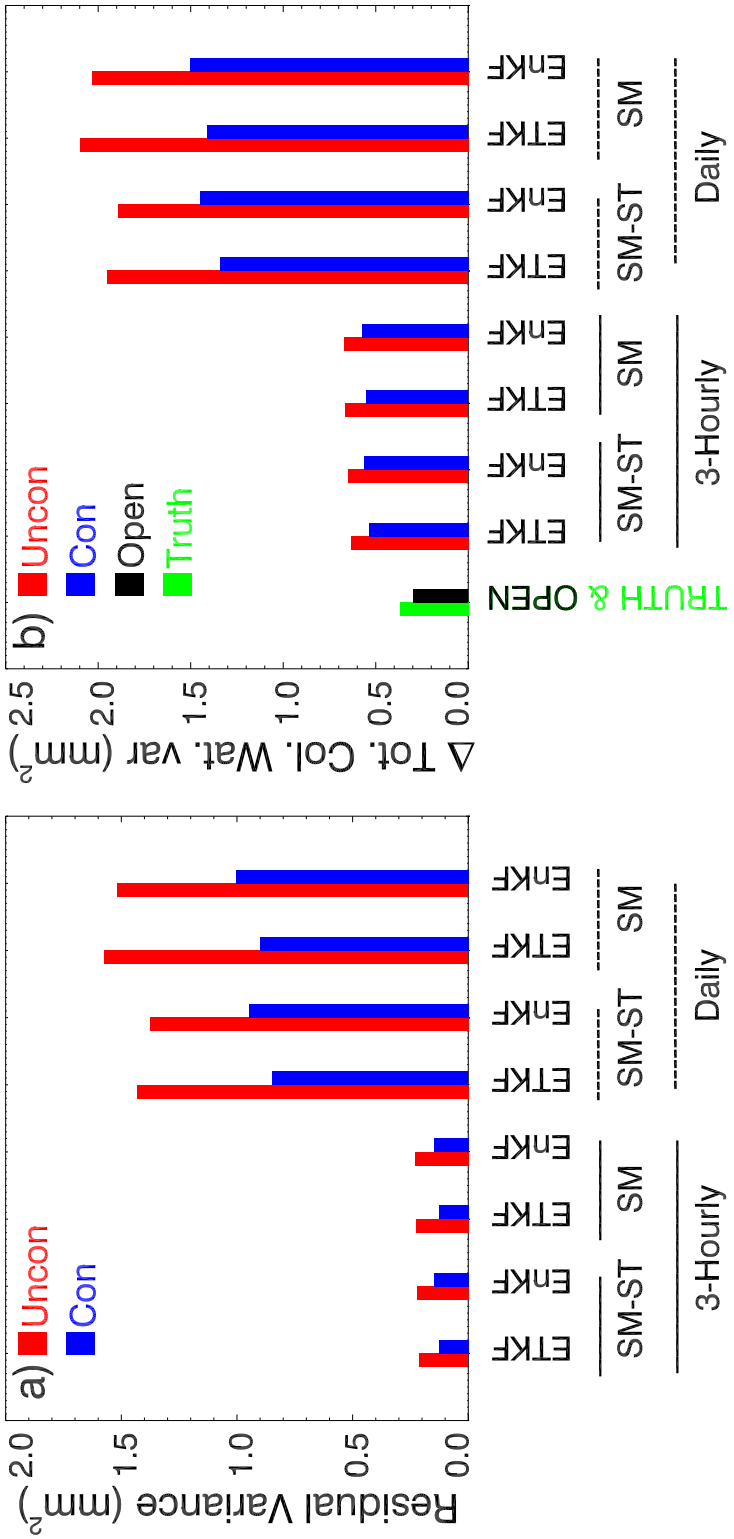

FIG. 3. (a) Water budget residual variance and (b) Total Column Water Content Change variance of constrained and unconstrained experiments. True total column water content change is shown in green bar, open loop water content change in black, unconstrained filter results in red, and constrained filter results in blue.

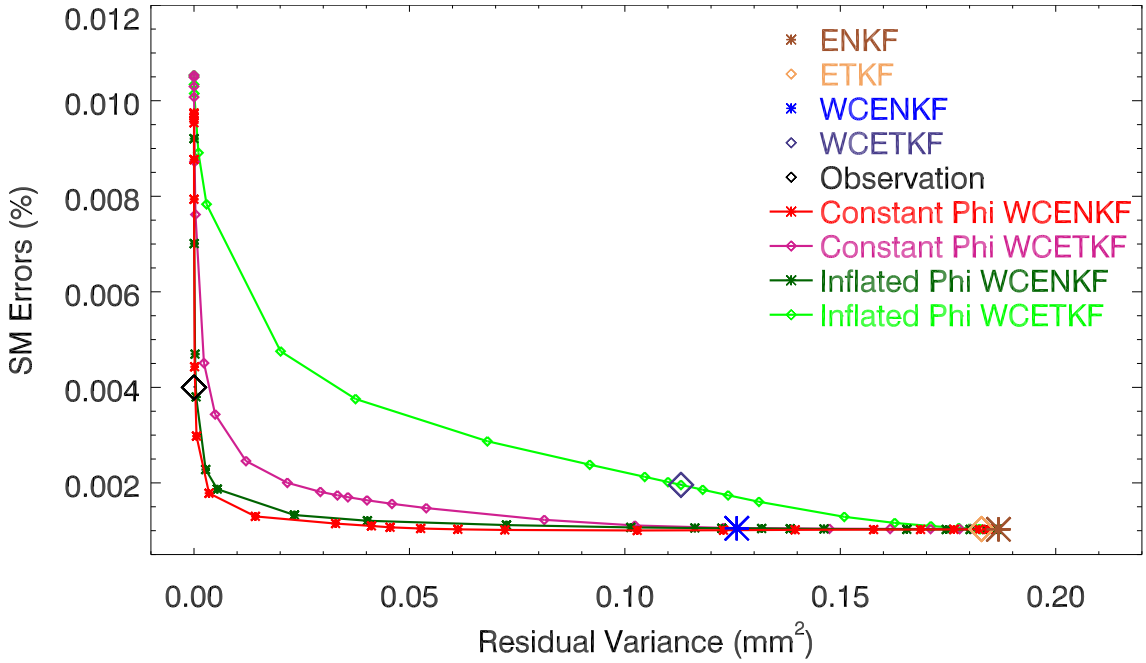

FIG. 4. SM error and residual relation for varying Phi values, where 3-hourly SM and ST observations are assimilated using 50 ensemble members. Each line represents series of simulations using 24 different constant φ values or 24 different φ inflation factors (Both inflation and constant values were selected as 0.00001, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 1.0, 1.2, 1.5, 2, 5, 10, 20, 50, 100, 200, 500, 1000). Single points represent single simulations of constrained filters with un-inflated φ values or of unconstrained filters. For both constant and inflated φ experiments, higher residuals are result of higher φ values and lower residuals are result of lower φ values (Inflation or constant φ values increase from left to right for green and red lines). Observation error is also marked with a black diamond. The residuals and the errors asymptotically approached to those of unconstrained simulations or strongly constrained simulations as the inflation factors for φ or the constant φ value was increased to ∞ or decreased to 0 respectively.