An improved approach for estimating observation and model error parameters in soil moisture data assimilation

W. T. Crow[1] and M. J. van den Berg[1,2]

Received 6 April 2010; revised 23 June 2010; accepted 2 July 2010; published 7 December 2010.

[1] The accurate specification of observing and/or modeling error statistics presents a remaining challenge to the successful implementation of many land data assimilation systems. Recent work has developed adaptive filtering approaches that address this issue. However, such approaches possess a number of known weaknesses, including a required assumption of serially uncorrelated error in assimilated observations. Recent validation results for remotely sensed surface soil moisture retrievals call this assumption into question. Here we propose and test an alternative system for tuning a soil moisture data assimilation system, which is robust to the presence of autocorrelated observing error. The approach is based on the application of a triple collocation approach to estimate the error variance of remotely sensed surface soil moisture retrievals. Using this estimate, the variance of assumed modeling perturbations is tuned until normalized filtering innovations have a temporal variance of one. Real data results over three highly instrumented watershed sites in the United States demonstrate that this approach is superior to a classical tuning strategy based on removing the serial autocorrelation in Kalman filtering innovations and nearly as accurate as a calibrated Colored Kalman filter in which autocorrelated observing errors are treated optimally.

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

1. Introduction

[2] During the past decade, the Ensemble Kalman filter (EnKF) has been extensively applied to the assimilation of satellite-based surface soil moisture retrievals into land surface models [see, e.g., *Reichle and Koster*, 2005; *Reichle et al.*, 2007 or *Bolten et al.*, 2010]. An important attribute of the EnKF is its flexibility with regards to land surface modeling errors [*Reichle et al.*, 2002]. Because background error/covariance information in the EnKF is propagated via a Monte Carlo ensemble of model forecasts, almost any type of error can be represented. However, such flexibility exceeds our current knowledge concerning the appropriate type, magnitude and structure of modeling error to assume for land surface model predictions [*Crow and Van Loon*, 2006; *De Lannoy et al.*, 2009]. As a result, most soil moisture EnKF applications are based on ad hoc or overly simplistic error models. Such gaps in knowledge have practical consequences given that inappropriate assumptions regarding the magnitude of modeling and observing errors lead to sub-optimal filter performance and the degradation of EnKF predictions [*Crow and Van Loon*, 2006; *Reichle et al.*, 2008; *Crow and Reichle*, 2008].

[3] A possible solution to this problem is the application of adaptive filtering approaches which attempt to diagnose the correct covariance of both modeling and observation errors based on the statistical analysis of filtering innovations (i.e., the time series of observations minus the model background realized during the analysis cycle). In classical Kalman filtering (KF) theory, optimal assumptions regarding the relative magnitude of modeling and observation error lead to a time series of serially uncorrelated (i.e., "white") innovations [*Mehra*, 1970]. Recent work by *Crow and Reichle* [2008] exploits this constraint by developing a series of land surface adaptive filtering approaches to iteratively whiten EnKF innovations. However, two shortcomings are evident in their results. First, while innovation whitening-based adaptive filter approaches eventually converge to the correct error magnitude; they do so only at relatively large time scales (>5 years) which are incompatible with the expected duration of upcoming satellite missions aimed at the remote estimation of surface soil moisture. Second, the most promising adaptive filtering approaches are based on an assumption that observing errors are temporally uncorrelated. New comparisons with long-term soil moisture times series derived from dense ground-based networks (presented here) call this assumption into question. If not properly accounted for, the presence of autocorrelated observing errors can erode the ability of whitening-based adaptive filtering approaches to identify optimal error parameters.

[4] This paper will address these two shortcomings by exploring an alternative approach to recovering appropriate error information for a KF and/or EnKF. The approach is based the application of a triple collocation (TC) procedure to independently estimate soil moisture retrieval errors in the absence of ground-based soil moisture observations [*Scipal et al.*, 2008; *Miralles et al.*, 2010]. Our particular TC

[1]Hydrology and Remote Sensing Laboratory, USDA Agricultural Research Service, Beltsville, Maryland, USA.
[2]Now at Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium.

approach follows *Scipal et al.* [2008] by estimating error in passive microwave retrievals of surface moisture through comparisons with independent soil moisture data products obtained from an active microwave scatterometer and a land surface model. While the approach requires strict independence between errors in each of the three data products, it functions well even if errors in each product are temporally autocorrelated.

[5] In addition to innovation whiteness, a second theoretical constraint is that, when properly normalized by the sum of observation and forecasting error covariance, KF innovations should have unity variance [*Dee*, 1995]. If the statistical properties of observing errors are determined via TC, this constraint can be used to define the optimal magnitude of modeling errors [*Mitchell and Houtekamer*, 1999]. In addition, *Crow and Reichle* [2008] demonstrate that adaptive tuning of modeling error parameters to match the innovation variance constraint converges much faster than tuning to achieve a serially white innovation sequence. Consequently, a soil moisture adaptive filtering approach based on using TC to estimate observing errors, and the subsequent tuning of modeling error to satisfy an innovation variance constraint, should address key outstanding issues for the design of such systems. The goal of this paper is to test this potential through a series of data assimilation experiments conducted with real remotely sensed soil moisture datasets.

2. Background

[6] As discussed above, our approach is based on an understanding of the impacts of autocorrelated observing errors on KF innovation statistics and the application of a triple collocation (TC) scheme to independently estimate observation error covariance information. This section describes these two topics in greater detail.

2.1. Kalman Filtering and Innovation Statistics

[7] To provide an initial examination of the impact of autocorrelated observing error on KF increments, we will employ a linear modeling strategy. In this simplified case, soil moisture forecasting is based on the well-known Antecedent Precipitation Index (API)

$$\mathrm{API}_i = \gamma \mathrm{API}_{i-1} + P_i, \tag{1}$$

where i is a daily time index and P a daily rainfall accumulation total in mm. For simplicity the unit-less loss parameter γ is held fixed at 0.85.

[8] The assimilation of remotely sensed estimates of surface soil moisture (θ_{RS}) into (1) using a KF requires that assumptions be made concerning the nature of both modeling and observing errors. With regards to modeling error, we assume that during the propagation of API from day i - 1 to i, an additive Gaussian noise term η is introduced into API forecasts which has a mean of zero, a scalar variance of Q and no serial autocorrelation (i.e., a lag-one autocorrelation ρ_1 of zero). This term represents the net impact of stochastic error in P and the overly simplistic representation of soil moisture loss in (1). Note that the individual perturbations η_i are aggregated via the first-order autoregressive structure of (1) to create autocorrelated error in API forecasts. Likewise, individual θ_{RS} retrievals at time i are assumed to be perturbed by an additive error term ξ which is mean-zero, serially white, and has constant variance R.

[9] Prior to assimilation, θ_{RS} retrievals are transformed via cumulative probability function (CDF) matching into a time series ($\theta_{\mathrm{RS}}^{\mathrm{API}}$) possessing the same long-term probability density function as API predictions derived from (1) [*Reichle and Koster*, 2005].

[10] Given the availability of $\theta_{\mathrm{RS}}^{\mathrm{API}}$ at time i, the KF updates (1) via

$$\mathrm{API}_i^+ = \mathrm{API}_i^- + K_i\left(\theta_{\mathrm{RS}_i}^{\mathrm{API}} - \mathrm{API}_i^-\right), \tag{2}$$

where "–" and "+" denote values before and after KF updating. The KF gain K in (2) is given by

$$K_i = T_i^- / \left(T_i^- + R\right), \tag{3}$$

where T^- is the background error variance in API forecasts. When θ_{RS} retrievals are available, T^- is updated following

$$T_i^+ = (1 - K_i) T_i^-. \tag{4}$$

[11] Between measurements, API is forecasted in time using P and (1). Likewise, the model forecast error T^+ is advanced in time following

$$T_i^- = \gamma_i^2 T_{i-1}^+ + Q. \tag{5}$$

[12] For this particular case, the adaptive filtering problem is defined as obtaining scalar estimates of Q and R ($\hat{Q}$ and $\hat{R}$) which accurately describe model and observation error covariance characteristics and therefore optimize the performance of the KF. Classical adaptive filtering approaches are based on examining the temporal statistics of normalized filtering innovations

$$\nu_i = \left(\theta_{\mathrm{RS}_i}^{\mathrm{API}} - \mathrm{API}_i^-\right) / \left(T_i^- + R\right)^{0.5}. \tag{6}$$

[13] A properly constructed linear filter should yield a ν time series that is serially uncorrelated ($\rho_1(\nu) = 0$) with unit variance ($\mathrm{var}(\nu) = 1$) [*Gelb*, 1974]. Hereinafter, these two constraints are referred to as the innovation "whiteness" and the "variance" constraints, respectively.

[14] Figure 1 illustrates results for a ν-tuning approach based on these two constraints. Results are derived from a synthetic twin data assimilation experiment. Using a daily time series of rainfall for the Little Washita River basin in Oklahoma, "truth" soil moisture values are generated using (1). These values are synthetically degraded by additive Gaussian noise (with variance R) to represent the uncertain retrieval of soil moisture from a satellite-based sensor and then re-assimilated back into (1), using the KF procedure described above, for the case where P is artificially corrupted with multiplicative noise. Consequently, the synthetic experiment is designed to evaluate the degree to which the assimilation of noisy soil moisture retrievals compensates for the impact of uncertain P on API predictions.

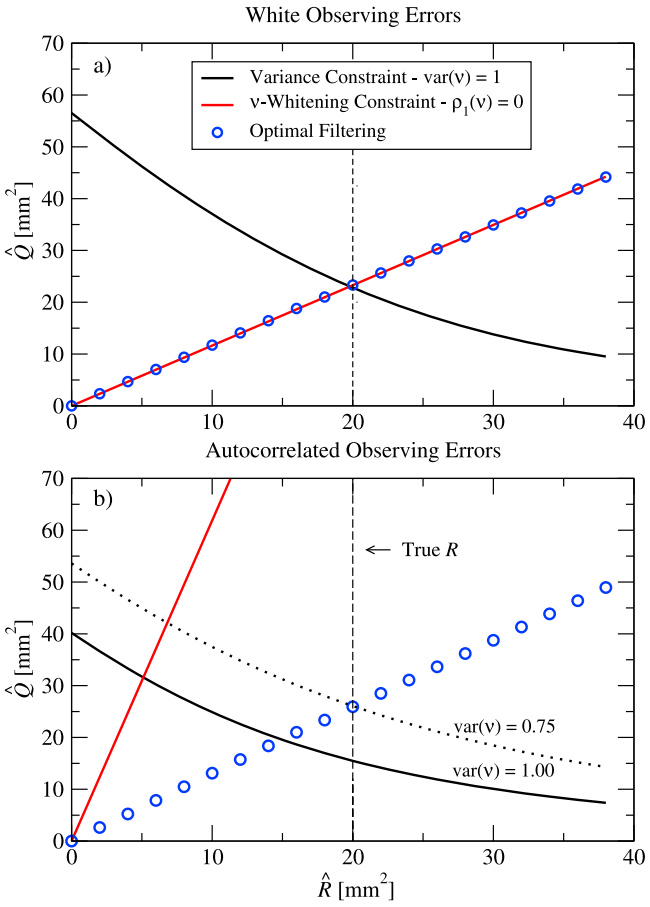

Figure 1. Values of assumed Q and R ($\hat{Q}$ and $\hat{R}$) meeting the innovation whiteness constraint (red line) and the variance constraint (black line) for the case of (a) serially white and (b) autocorrelated observing errors. Also plotted are $\hat{Q}$ and $\hat{R}$ combinations associated with optimal filter performance (blue circles) and the true R used in the synthetic twin experiment.

[15] As with any implementation of the KF, it is incumbent upon the user to provide values of $\hat{Q}$ and $\hat{R}$. Figure 1a shows how various estimates impact ν temporal statistics in the case of serially uncorrelated observing errors ($\rho_1(\xi) = 0$). The black line represents the set of $\hat{Q}$ and $\hat{R}$ which meet the ν-variance constraint (var(ν) = 1) and the red line the set which satisfies the ν-whiteness constraint ($\rho_1(\nu) = 0$). They intersect at the true value of R (20 mm^2) applied in this particular synthetic experiment. Furthermore, the set of $\hat{Q}$ and $\hat{R}$ which produces optimal filtering performance (i.e., the smallest observed root-mean-square error (RMSE) between true and analyzed API) tracks along the $\rho_1(\nu) = 0$ constraint such that all combinations of $\hat{Q}$ and $\hat{R}$ combinations exhibiting serially white ν also provide optimal filter RMSE performance. Note that such alignment is achieved even in this somewhat non-Gaussian case where modeling error is introduced via log-normal multiplicative noise.

[16] However, this alignment is disrupted when autocorrelated observing errors (corresponding to $\rho_1(\xi) = 0.5$) are introduced into the synthetic experiment (Figure 1b). Here the $\hat{Q}/\hat{R}$ ratio ensuring serially white ν increases such that the intersection of the variance and whiteness constraint lines no longer captures the true R or the set of $\hat{Q}$ and $\hat{R}$ ensuring optimal filtering performance. Instead, serially uncorrelated ν are possible only when R is underestimated. The result being an offset between optimal $\hat{Q}$ and $\hat{R}$ combinations and those producing serially white ν.

[17] Relative to the whiteness constraint, the variance constraint and optimal performance lines are relatively insensitive to the presence of autocorrelated observing error (compare Figures 1a and 1b). This suggests an alternative tuning approach for soil moisture data assimilation: independently acquire R and subsequently tune $\hat{Q}$ until the variance constraint is met. In addition to its potential for robustness in the presence of autocorrelated error, this approach has the added advantage of improved adaptive filtering convergence since variance tuning procedures for $\hat{Q}$ (given known R) have been shown to converge much faster than tuning approaches in which $\hat{Q}$ and $\hat{R}$ are simultaneously tuned to achieve white ν [*Crow and Reichle*, 2008]. However, the new approach requires independent estimates of R which are not currently available outside of a small number of heavily instrumented ground sites. To address this need, the following sub-section reviews an alternative methodology for estimating R in the absence of ground-based data.

2.2. Triple Collocation

[18] Unlike ν-whitening, tuning to match the variance constraint ensures optimal filtering performance only if accurate *a priori* information concerning R is available. Such information is generally difficult to obtain [*Scipal et al.*, 2008]. However, the recent application of triple collocation (TC) approaches to soil moisture error estimation offers a potential approach for estimating R over wide continental areas [*Scipal et al.*, 2008; *Miralles et al.*, 2010]. Our particular TC approach is based on the simultaneous availability of three separate surface soil moisture products: a passive microwave retrieval dataset based on Advanced Microwave Scanning Radiometer (AMSR-E) brightness temperature (T_B) observations (θ_{AMSRE}), a model-based product based on (1) (θ_{API}), and an active microwave retrieval product based on European Space Radar (ERS) scatterometer measurements (θ_{ERS}). See section 3 for a detailed description of each product. While estimating the same geophysical quantity, each of these products are impacted by mutually independent errors [*Scipal et al.*, 2008]. The TC approach exploits such mutual independence in redundant data sets in order to obtain RMSE estimates for each individual product. Here, our goal is to estimate the RMSE of θ_{AMSRE}, and use this estimate to obtain the error covariance R for the assimilation of θ_{AMSRE} into a land surface model.

[19] To begin, each of the three soil moisture products (θ_{AMSRE}, θ_{API}, and θ_{ERS}) is individually decomposed into climatology mean and anomaly components

$$\theta_i = \theta_i^{'} + \langle\theta\rangle_{D(i)}^{N}, \tag{7}$$

where $\langle\theta\rangle_{D(i)}^{N}$ is the climatological expectation for soil moisture on the day-of-year (D) associated with daily time-step i, and θ'_i is the actual anomaly relative to this expectation. This expectation is obtained by averaging all available soil moisture estimates (for a given product) located within N days of day-of-year D. Note that this calculation requires access to a

Table 1. Summary of Physical Characteristics and Ground-Based Soil Moisture Instrumentation for the USDA ARS Experimental Watershed Sites

	Little River (LR)	Little Washita (LW)	Walnut Gulch (WG)
Location (lat/long)	31.6/−83.7	34.9/−98.1	31.7/−110.1
Area (km^2)	334	611	150
Soil moisture sites	29	20	21
Climate	humid	subhumid	semiarid
Land cover	crops/forest	rangeland/crops	shrubs/short grass
Relief	moderate	low	moderate

multiyear dataset for each of the three soil moisture products. A choice of $N = 365$ days corresponds to the removal of a constant bias while $N = 31$ days (our default choice) means that anomalies are calculated relative to a seasonally varying soil moisture climatology. The impact of N on our analysis is discussed further in section 4.2. After the application of (7), the resulting soil moisture anomaly time series are rescaled to have the same mean and variance. Here this transformation is based on selecting θ'_{API} as a reference data set and linearly rescaling the other two soil moisture products.

[20] The relationship of our three co-located time series with the hypothetical true anomaly soil moisture (θ'_{TRUE}) time series can now be expressed as

$$\theta'_{i,\mathrm{API}} = \theta'_{i,\mathrm{TRUE}} + \varepsilon_{i,\mathrm{API}}, \tag{8}$$

$$\theta'_{i,\mathrm{AMSRE}} = \theta'_{i,\mathrm{TRUE}} + \varepsilon_{i,\mathrm{AMSRE}}, \tag{9}$$

$$\theta'_{i,\mathrm{ERS}} = \theta'_{i,\mathrm{TRUE}} + \varepsilon_{i,\mathrm{ERS}}, \tag{10}$$

where θ' is the anomaly time series for each product time series (see above) and ε denotes time-varying residual errors in each relative to the unknown truth (θ'_{TRUE}).

[21] θ'_{TRUE} can be removed from (8) to (10) through a simple elimination procedure to obtain

$$\theta'_{i,\mathrm{AMSRE}} - \theta'_{i,\mathrm{API}} = \varepsilon_{i,\mathrm{AMSRE}} - \varepsilon_{i,\mathrm{API}}, \tag{11}$$

$$\theta'_{i,\mathrm{AMSRE}} - \theta'_{i,\mathrm{ERS}} = \varepsilon_{i,\mathrm{AMSRE}} - \varepsilon_{i,\mathrm{ERS}}. \tag{12}$$

[22] Assuming that all three error types in (11) and (12) are mutually uncorrelated, error in θ'_{AMSRE} can be isolated by multiplying (11) and (12) and averaging in time (denoted by "$\langle\text{-}\rangle$" brackets)

$$\begin{aligned}\hat{R} &= \left\langle \left(\theta'_{i,\mathrm{AMSRE}} - \theta'_{i,\mathrm{API}}\right)\left(\theta'_{i,\mathrm{AMSRE}} - \theta'_{i,\mathrm{ERS}}\right)\right\rangle \\ &= \left\langle \varepsilon_{i,\mathrm{AMSRE}}{}^2 \right\rangle.\end{aligned} \tag{13}$$

[23] Since the accuracy of $\hat{R}$ from (13) is based the assumed independence of error in $\theta'_{i,\mathrm{AMSRE}}$, $\theta'_{i,\mathrm{API}}$ and $\theta'_{i,\mathrm{ERS}}$, these assumptions can be indirectly verified by comparing $\hat{R}$ from (13) to R values associated with optimal performance of a KF (see section 4). It should also be noted that while (13) requires zero cross-covariance between errors, it produces unbiased $\hat{R}$ even if autocorrelated errors are present in one or more of the three input data products.

3. Approach

[24] All subsequent results are based on data assimilation experiments conducted using real AMSR-E soil moisture products and validated against independent ground-based surface soil moisture observations obtained at three data-rich watershed sites in the United States. Ground data, remote sensing, modeling and data assimilation components of the analysis are described below.

3.1. Study Locations and Ground Data

[25] Our analysis is centered on three heavily instrumented watershed sites within the United States: the Little River (LR) watershed near Tifton, Georgia [*Bosch et al.*, 2007], the Little Washita (LW) watershed in southwest Oklahoma [*Allen and Naney*, 1991; *Cosh et al.*, 2006], and the Walnut Gulch (WG) watershed in southeastern Arizona [*Renard et al.*, 2008; *Cosh et al.*, 2008]. All three sites are maintained as experimental watersheds by the United States Department of Agriculture's Agricultural Research Service (USDA ARS). As part of validation activities for AMSR-E soil moisture products, each USDA ARS watershed has been instrumented with a series of 20 to 30 spatially distributed surface soil moisture probes whose continuous measurements can be aggregated to provide an accurate estimate of watershed-scale surface soil moisture [*Jackson et al.*, 2010]. Here such estimates are withheld as a source of validation information to evaluate the performance of various data assimilation methodologies. Physical details for the watersheds and their soil moisture instrumentation are given in Table 1.

3.2. Remotely Sensed Surface Soil Moisture Products

[26] Three separate surface soil moisture retrievals products derived from AMSR-E T_B observations are used for θ_{AMSRE}. All three products are available starting in mid-June 2002 at a spatial resolution of approximately 40 km. The $\theta_{\mathrm{AMSRE\text{-}NASA}}$ product is the official NASA AMSR-E Level 3 soil moisture product derived from a dual-polarization algorithm applied to H- and V-polarized AMSR-E X band (10.6 GHz) T_B observations [*Njoku*, 2008]. The $\theta_{\mathrm{AMSRE\text{-}USDA}}$ product (developed at the USDA Hydrology and Remote Sensing Laboratory by T. J. Jackson and R. Bindlish) is based on X band T_B observations as well, but uses a single channel (H polarization only) algorithm [*Jackson et al.*, 2010]. The $\theta_{\mathrm{AMSRE\text{-}VU}}$ product (developed at the Vrije University of Amsterdam (VU) by R.A.M. de Jeu and T. Holmes in collaboration with M. Owe at NASA Goddard Space Flight Center) applies the algorithm of *Owe et al.* [2008] to dual-polarized C band (6.9 GHz) T_B and falls back to X band T_B in areas of significant C band radio frequency interference over the United States and Japan. Soil moisture retrievals from ascending (1:30 P.M. local solar time) and descending (1:30 A.M. local solar time) AMSR-E overpasses are combined into a single time series. This results in a daily soil moisture product over each watershed with retrievals on about 4 out of every 5 days.

[27] The θ_{ERS} product used in our TC procedure is based on scatterometer observations obtained from the European Space Radar (ERS-1 and ERS-2) measurements and application of the soil moisture retrieval algorithm described by *Naeimi et al.* [2009]. From mid-2003 onward, a new 50-km resolution θ_{ERS} retrieval is available approximately once every 3 days over each watershed. Time series results for all remotely sensed soil moisture products are obtained by extracting the closest pixel to the center of each watershed from gridded versions of each soil moisture product.

3.3. API and KF

[28] As noted above, initial proof-of-concept results are based on the application of the API model described by (1). Rainfall input (P) are obtained from the 3B42RT rainfall product produced by the Tropical Rainfall Measurement Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) [*Huffman et al.*, 2007]. The TRMM 3B42RT product combines multisatellite thermal and microwave observations to provide 3 hourly rainfall accumulation estimates. As a real-time product, it does not ingest ground-based rain gauge observations. Here 3 hourly 3B42RT products are aggregated into daily (0 to 0 UTC) accumulations to match the time step of (1). Data assimilation of various θ_{AMSRE} products into the API model within each watershed is based on the KF implementation described in section 2.1 and conducted between the onset of AMSR-E soil moisture availability in June 2002 and the end of complete ground-data availability on 27 July 2007.

3.4. NOAH/LIS and EnKF

[29] In addition to the KF/API system described above, additional results are obtained from a more complex system in which a full EnKF is applied to assimilate surface soil moisture retrievals into the nonlinear, multilayer NOAH land surface model (K. Mitchell, The community Noah land-surface model: User Guide Public Release Version 2.7.1, available online at http://www.emc.ncep.noaa.gov/mmb/gcp/noahlsm) implemented within the NASA Land Information System (LIS) framework [*Kumar et al.*, 2006]. The NOAH model is run with four soil layers of thickness: 5 cm, 35 cm, 60 cm and 1 m (from top to bottom). All NOAH surface soil moisture states are initialized as spatially uniform on 1 February 2002 and spun-up (i.e., allowed to develop realistic spatial variability) for approximately five months prior to the assimilation of the first θ_{AMSRE} retrievals in June 2002. NOAH simulations are run on a half-hourly time step for individual quarter-degree grid cells corresponding to each of the three watershed sites (LW, LR and WG). All forcing data except rainfall is acquired from the North American Land Data Assimilation System retrospective forcing data set [*Cosgrove et al.*, 2003]. As in the KF/API system, rainfall data is obtained from the TRMM 3B42RT dataset.

[30] Assimilation into the NOAH model is based on a 150-member ensemble EnKF system using the EnKF module within NASA LIS [*Kumar et al.*, 2008]. This ensemble size is conservatively chosen to be significantly larger than typical ensemble size recommendations of about 20 to 50 [see, e.g., *Crow and Wood,* 2003] for one-dimensional land surface data assimilation problems. To create the forecast ensemble, NOAH rainfall inputs are perturbed via multiplicative noise sampled from a mean-unity, log-normal distribution with a fixed standard deviation of 0.5. Perturbing only rainfall can lead to a collapse of the NOAH EnKF ensemble between rainfall events. Consequently, mean-zero, additive Gaussian noise is also directly applied to the each of the four NOAH soil moisture layers. Based on past experience, these noise values have standard deviations of $\sqrt{Q}$, $\sqrt{Q}/3$, $\sqrt{Q}/6$, and $\sqrt{Q}/10$ (from top to bottom, respectively) and a multilayer cross-correlation matrix of:

$$\begin{pmatrix} 1.0 & 0.6 & 0.4 & 0.2 \\ 0.6 & 1.0 & 0.6 & 0.4 \\ 0.4 & 0.6 & 1.0 & 0.6 \\ 0.2 & 0.4 & 0.6 & 1.0 \end{pmatrix}.$$

[31] All random model perturbations are constructed to have a temporal correlation length of 2 h (i.e., four NOAH time steps). As noted above, the requirement for the user to make (largely) arbitrary assumptions regarding the statistical structure of modeling perturbations is a general problem in the application of EnKF to complex land surface models. In addition, by making fixed assumptions concerning the variation of error magnitude and cross-correlation with depth, we are reducing the potentially complex NOAH model error parameterization problem to the estimation of a single scalar parameter Q.

[32] Prior to any data assimilation, $\theta_{AMSRE\text{-}USDA}$ retrievals are linearly rescaled to have the same temporal mean and variance as updated NOAH surface soil moisture predictions at each individual watershed site. These rescaled retrievals are then assimilated into NOAH at their corresponding acquisition times (i.e., 1:30 or 13:30 local solar time). EnKF updating is limited to the four vertical soil moisture states in NOAH, and validation is performed by comparing the EnKF surface soil moisture analysis, calculated as the mean of the EnKF ensemble for the top NOAH soil moisture layer, to watershed-average surface soil moisture observations obtained at each of the watershed sites (section 3.1). As in KF/API analysis, the analysis is conducted between mid-June 2002 and July 2007.

[33] Unless otherwise specified, all references to tuning $\hat{Q}$ and/or $\hat{R}$ values (for either the API/KF or the NOAH/EnKF system) are based on a batch calibration approach where the analysis is run repeatedly from start to finish. At the end of each individual analysis run, temporally fixed values of $\hat{Q}$ and/or $\hat{R}$ are modified using a simple tangent linear search algorithm to iteratively acquire a ν-sequence that satisfies the ν-variance and/or whiteness constraints.

4. Results

[34] As discussed above, preliminary results are based on the assimilation of θ_{AMSRE} retrievals into the API model in (1) using a standard KF (section 3.3). In addition, supplemental results are obtained for the case of assimilating θ_{AMSRE} into the multilayer NOAH model using a full EnKF (section 3.4).

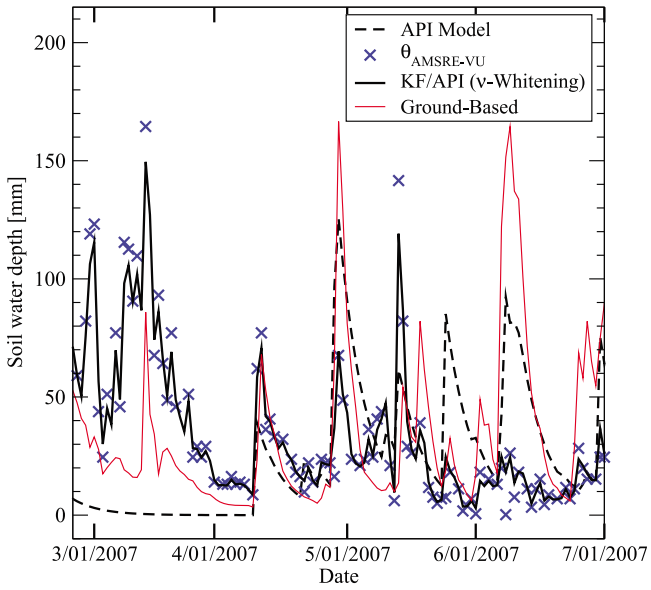

Figure 2. For the assimilation of the $\theta_{\text{AMSRE-VU}}$ soil moisture product into the API model over the LW watershed, the mid-2007 time series for the open loop API simulation, rescaled $\theta_{\text{AMSRE-VU}}$ observations, the KF surface soil moisture analysis, and the ground-based estimate of watershed-scale soil moisture.

4.1. Impact of Autocorrelated Observing Error

[35] Figure 2 provides illustrative daily time series results for the KF-based assimilation of the $\theta_{\text{AMSRE-VU}}$ product into (1) within the LW watershed for the case where $\hat{Q}$ and $\hat{R}$ are obtained via ν-whitening (i.e., tuning $\hat{Q}$ and $\hat{R}$ until $\rho_1(\nu) = 0$ and $\text{var}(\nu) = 1$). Also plotted is the time series of watershed-averaged soil moisture obtained from averaging ground observations within the watershed. All values are plotted after transformation into the API model's soil moisture climatology. RMSE comparisons to the ground-based observations allow for the evaluation of various data assimilation approaches.

[36] Figure 3 illustrates the impact of $\hat{R}$ on the RMSE accuracy of KF surface soil moisture predictions for the case of assimilating $\theta_{\text{AMSRE-VU}}$ into the API model over the LW watershed. The plotted black curve is derived by examining a range of assumed $\hat{R}$ and tuning $\hat{Q}$ such that $\text{var}(\nu) = 1$. The resulting curve is bounded by API modeling results obtained without assimilation (the labeled "open loop" line) and optimal data assimilation results obtained from applying a Colored Kalman filter (the "ColKF" line). Unless otherwise noted, all subsequent RMSE and $\hat{R}$ values are reported in the soil moisture climatology of the ground-based observations.

[37] Unlike the standard KF, the ColKF is designed to account for the presence of autocorrelated observing and/or modeling errors [*Chui and Chen*, 1991]. Here, model noise η and observation error ξ are modeled as first-order, autoregressive processes at a discrete daily time interval i

$$\eta_i = \Sigma\eta_{i-1} + \zeta_i \tag{14}$$

$$\xi_i = \Theta\xi_{i-1} + \mu_i, \tag{15}$$

where Σ and Θ are constant scalar values and ζ and μ are mean-zero, random Gaussian variables with respective variances Q and R and no autocorrelation or cross correlation.

[38] Based on these error models and a state-augmentation procedure, the ColKF generalizes the KF implementation in (1) to (6) to explicitly account for autocorrelated modeling and observing errors (see *Chui and Chen* [1991] for details). The ColKF line in Figure 3 represents the lowest ColKF RMSE achievable via automated batch tuning of all four parameters in (14) and (15) (i.e., R, Q, Σ, and Θ) to optimize the fit of ColKF surface soil moisture results to ground-based observations. While such tuning provides an optimal analysis, it is possible only in rare data-rich sites possessing adequate ground-based soil moisture monitoring networks.

[39] Between these two bounds, the performance of the KF depends on $\hat{R}$. A large value of $\hat{R}$ leads to reduced weight on assimilated observations and to results that converge back to the open loop case of running the model without assimilation. Likewise, very small values of $\hat{R}$ converge to a sub-optimal direct insertion case where no weight is applied to the background API forecast. As expected, an optimal choice for $\hat{R}$ lies in between. Due to the degrading effects of autocorrelated observing errors on a KF filter implementation [*Daley*, 1992], no choice of $\hat{R}$ can match the calibrated ColKF. However, a proper choice of $\hat{R}$ produces nearly optimal results even in the presence of autocorrelated observing errors. The challenge lies in obtaining $\hat{R}$ values associated with the RMSE minimum in Figure 3.

[40] As illustrated in Figure 1a, the classical approach for finding this minimum is adjusting $\hat{R}$ until $\rho_1(\nu) = 0$. Unfortunately, ν-whitening in Figure 3 leads to a $\hat{R}$ value that is clearly too low (Figure 3). Based on our earlier discussion of Figure 1b, such underestimation implies the

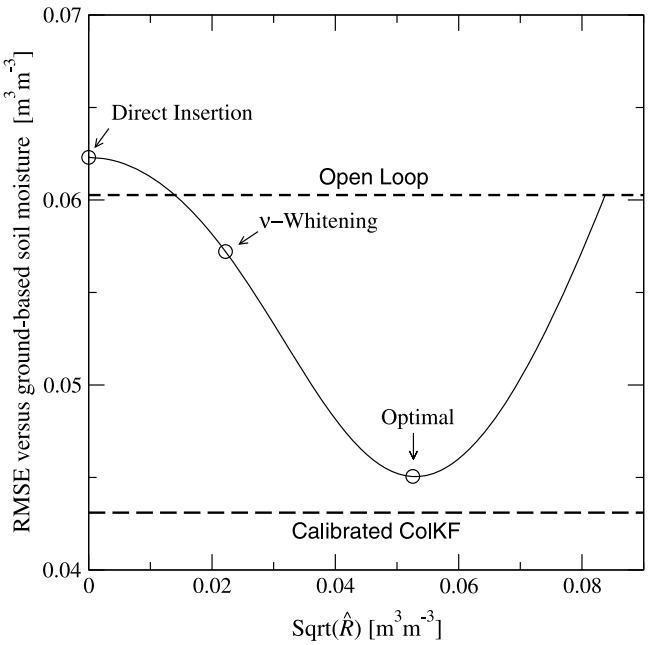

Figure 3. For the assimilation of the $\theta_{\text{AMSRE-VU}}$ soil moisture product into the API model over the LW watershed, the impact of estimated R ($\hat{R}$) on the RMSE accuracy of KF surface soil moisture estimates versus ground data.

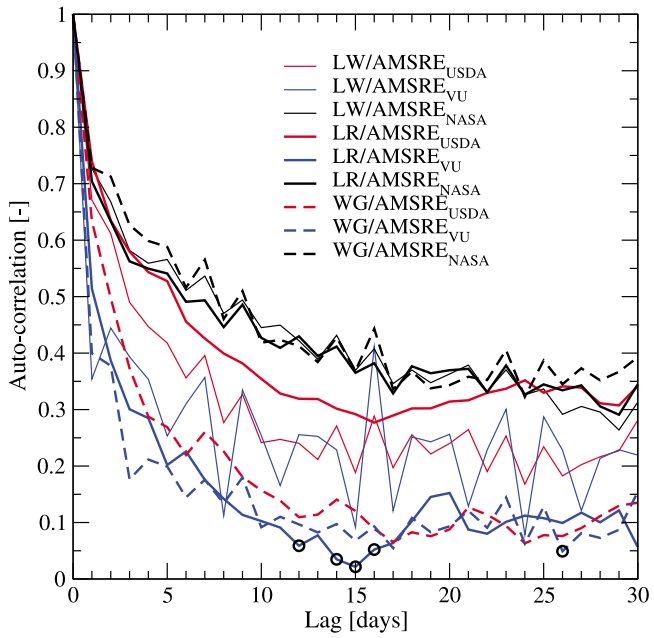

Figure 4. Temporal autocorrelation functions for error in all three θ_{AMSRE} products (USDA, VU, and NASA) within all three watershed sites (LW, LR, and WG). Circles indicate points on the autocorrelation functions which are not statistically significant at 95% confidence.

presence of autocorrelated errors in assimilated $\theta_{AMSRE\text{-}VU}$ retrievals.

[41] Figure 4 directly examines this issue by calculating the autocorrelation function for error in all three AMSR-E surface soil moisture products ($\theta_{AMSRE\text{-}USDA}$, $\theta_{AMSRE\text{-}VU}$ and $\theta_{AMSRE\text{-}NASA}$) at each of the three watershed sites (LW, LR and WG). Here, error is explicitly calculated as the difference between ground-based observations and retrievals after a 31 day moving average climatology is removed from each product. Even after the removal of such seasonality, results demonstrate the presence of significant autocorrelation in retrieval errors for all three θ_{AMSRE} products at all three watershed sites.

[42] Such autocorrelation appears to have a detectable impact on $\hat{R}$ results obtained from ν-whitening. In particular, black symbols in Figure 5 plot the relationship between estimated (via ν-whitening) and optimal $\hat{R}$ for all three θ_{AMSRE} products at all three watershed sites (open circles) and demonstrates that $\hat{R}$ obtained via ν-whitening consistently underestimates values of $\hat{R}$ associated with optimal KF performance. These real data assimilation results are therefore consistent with earlier theoretical predictions in Figure 1b.

4.2. TC-based *R* Estimates

[43] Results in Figures 3 to 5 illustrate that ν-whitening does a poor job of identifying values of $\hat{R}$ associated with the optimal assimilation of θ_{AMSRE} into (1). An alternative approach for obtaining $\hat{R}$ for θ_{AMSRE} is the TC procedure described in section 2.2. However, before the procedure is applied, it is important to clarify the relationship between optimal $\hat{R}$ and various types of RMSE measures for soil moisture retrieval accuracy.

[44] Since our data assimilation approach is based on rescaling surface soil moisture retrievals prior to their assimilation (see section 2.1), the correct RMSE formulation should be bias-free (i.e., calculated after the relative long-term bias between satellite-based surface moisture retrievals and ground-based observations has been removed). However, it is unclear whether such correction should be based on removing a single long-term soil moisture mean or a seasonally varying climatology. In the former case, RMSE values, or "total RMSE," reflect the ability of the retrieval product to capture both a seasonal soil moisture climatology *and* anomalies relative to this climatology. In the latter case, the "anomaly RMSE" is sensitive only to the representation of anomalies.

[45] Figure 6 examines this issue by plotting both types of RMSE (calculated using the ground-data available at each of the three USDA ARS watershed sites) against values of Sqrt($\hat{R}$) which produce optimal RMSE performance (i.e., the lowest RMSE fit to ground-based soil moisture observations). The comparisons reveal that RMSE values containing seasonal climatology errors (i.e., "total RMSE") slightly overestimate optimal Sqrt($\hat{R}$) values and calculating RMSE with anomaly values in which a seasonal climatology has been removed (i.e., "anomaly RMSE") provides a better representation of optimal Sqrt($\hat{R}$). The implication is that a TC approach will provide more useful $\hat{R}$ results when applied to AMSR-E soil moisture data sets in which a seasonal climatology has been explicitly removed.

[46] Based on this analysis, the TC approach is applied to (1) one of the θ_{AMSRE} products, (2) a soil moisture proxy value predicted by the API model, and (3) soil moisture retrievals from the ERS scatterometer after each product has been decomposed into anomaly products using (7) and N = 31. Figure 7 plots the resulting anomaly time series

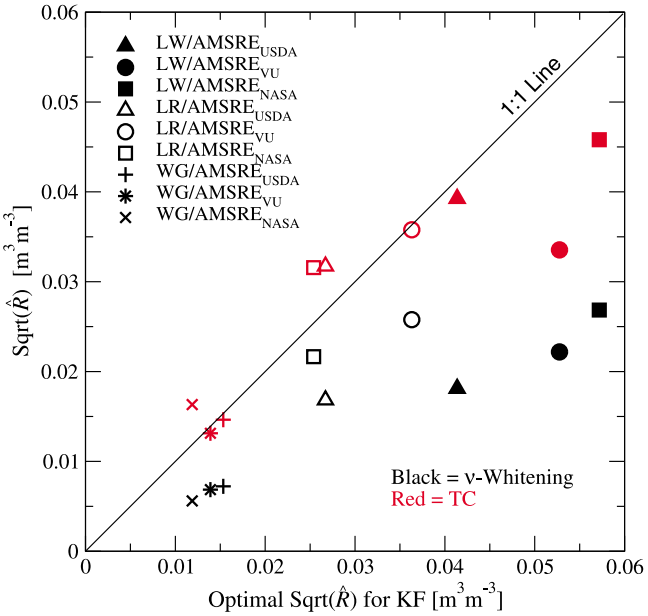

Figure 5. Relationship between optimal KF Sqrt($\hat{R}$) and Sqrt($\hat{R}$) derived from ν-whitening (in black) and TC (in red). Points are shown for all three θ_{AMSRE} products (USDA, VU, and NASA) within all three watersheds (LW, LR, and WG).

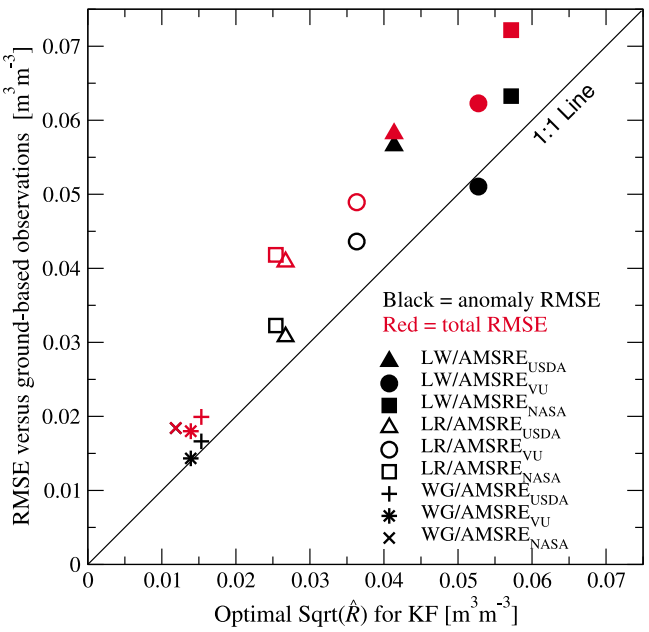

Figure 6. Relationship between optimal KF Sqrt($\hat{R}$) and the total (in black) and anomaly (in red) RMSE values for θ_{AMSRE} retrievals calculated versus ground-based soil moisture measurements. Points are shown for all three θ_{AMSRE} products (USDA, VU, and NASA) within all three watersheds (LW, LR, and WG).

for θ'_{API}, $\theta'_{AMSRE\text{-}VU}$ and θ'_{ERS} (after transformation into the API climatology) at the LW site. These three anomaly data products are then inputted into (13) to obtain $\hat{R}$ values for use during the KF-based assimilation of θ_{AMSRE} products into the API model. The process is repeated for all three θ_{AMSRE} products in all three watersheds. For each of these nine total cases, red symbols in Figure 5 indicate the relationship between TC-based $\hat{R}$ and optimal $\hat{R}$ which minimize the RMSE of KF predictions relative to ground-based surface soil moisture measurements. In each case, the TC approach provides a better estimate of optimal $\hat{R}$ than ν-whitening (i.e., tuning $\hat{Q}$ and $\hat{R}$ until $\rho_1(\nu) = 0$ and var(ν) = 1). In particular, it avoids the low bias observed in $\hat{R}$ obtained from ν-whitening in the presence of autocorrelated observing errors (Figure 5). The close fit between optimal and TC-based $\hat{R}$ in Figure 5 also indirectly verifies the independent error assumptions underlying TC, since cross correlation in θ'_{API}, $\theta'_{AMSRE\text{-}VU}$ and θ'_{ERS} errors would induce bias in $\hat{R}$ obtained from (13).

4.3. API/KF Analysis Results

[47] The improved accuracy of $\hat{R}$ estimates in Figure 5 should lead to better KF surface soil moisture predictions. Table 2 details the impact of various strategies for obtaining $\hat{R}$ on the accuracy of the subsequent KF surface soil moisture analysis. A range of approaches are examined: (1) a direct insertion (DI) case where $\hat{R} = 0.0$, (2) a ν-whitening case where $\hat{R}$ and $\hat{Q}$ are tuned until $\rho_1(\nu) = 0.0$ and var(ν) = 1, and (3) a batch TC case where $\hat{R}$ is obtained via the TC approach described above and $\hat{Q}$ is batch-tuned until the var (ν) = 1. Also listed are open loop (OL) API modeling results for the case of no assimilation and optimal ColKF data assimilation results. Note that all results to this point are based on a "batch" tuning methodology where all TC error estimation and KF calibration is conducted by repeatedly running the KF analysis for the entire June 2002 to July 2007 time period. More operationally relevant "adaptive" tuning techniques where estimates are obtained during a single run through the dataset are examined in section 4.3.1.

[48] For all nine cases in Table 2 (the three θ_{AMSRE} products over three different watersheds) the batch TC-based calibration approach either matches or outperforms the batch ν-whitening case. In addition, the batch TC case is typically only marginally worse than the theoretically optimal ColKF case, suggesting that our approach is effective at obtaining approximately optimal filter performance.

[49] Despite the use of a var(ν) = 1 constraint, a close examination of synthetic data results in Figure 1b suggests that a lower ν-variance target may actually be advantageous in the presence of autocorrelated observing errors. Specifically, the intersection between true R and a var(ν) = 0.75 constraint in Figure 1b is closer to the optimal Q/R ratio than a var(ν) = 1.00 constraint. This implies that an advantageous ad hoc correction would be to redefine the variance constraint as var(ν) = 1.00 − δ where $\delta >= 0.00$. To test this possibility, API/KF TC results in Table 2 were regenerated for a range of δ between 0.00 and 0.50. However, in contrast to the synthetic experiments in Figure 1, real data results in Table 2 are not consistently improved by any choice of $\delta > 0.00$. Consequently, retaining the original var(ν) = 1 constraint appears justified.

4.3.1. Adaptive Filtering

[50] As noted above, all results to this point have been based on batch optimizing techniques in which the KF analysis is run repeatedly end-to-end in order to tune νstatistics and acquire TC-based $\hat{R}$. Such reanalysis-type approaches are generally difficult to implement in operational environments. In particular, an operational implementation presents two new challenges. First, at the onset of the implementation, sampling errors in TC-based $\hat{R}$ will be

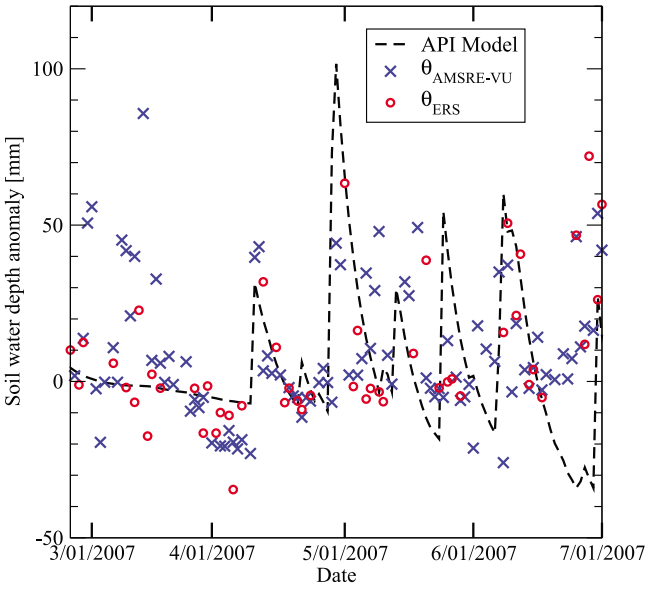

Figure 7. The mid-2007 time series of θ_{API}, $\theta_{AMSRE\text{-}VU}$, and θ_{ERS} anomalies for the LW watershed site.

Table 2. For the KF/API Data Assimilation Case, RMSE Fit Ground-Based Observations for the Open Loop (OL), Direct Insertion (DI), ν-Whitening, Triple Collocation (TC), and ColKF Cases[a]

Site	Product	OL ($m^3\ m^{-3}$)	DI ($m^3\ m^{-3}$)	ν-Whitening (Batch) ($m^3\ m^{-3}$)	TC (Adaptive) ($m^3\ m^{-3}$)	TC (Batch) ($m^3\ m^{-3}$)	ColKF ($m^3\ m^{-3}$)
LW	USDA	0.06028	0.05819	0.05564	0.05079	0.05048	0.04922
	VU	0.06027	0.06230	0.05721	0.04633	0.05139	0.04311
	NASA	0.06053	0.07219	0.06619	0.05657	0.05702	0.05213
LR	USDA	0.05602	0.04092	0.03904	0.03950	0.03842	0.03761
	VU	0.05588	0.04895	0.04302	0.04122	0.04074	0.04042
	NASA	0.05697	0.04180	0.03955	0.04092	0.04003	0.03923
WG	USDA	0.02417	0.01998	0.01884	0.01781	0.01694	0.01678
	VU	0.02396	0.01797	0.01685	0.01635	0.01525	0.01508
	NASA	0.02267	0.01847	0.01786	0.01756	0.01786	0.01676

[a]Results are broken down by AMSR-E soil moisture data product (USDA, VU, or NASA) and site (LW, LR, or WG).

large due to averaging (13) over short time periods. Second, the tuning of $\hat{Q}$ to ensure var(ν) = 1 will have to be performed on-line and updated continuously as real-time data becomes available.

[51] To examine these challenges, the TC-based calibration approach was modified to run in an adaptive mode based on dividing the total run time into 150 day windows and assuming no data available prior to the start of the analysis in June 2002. At the end of each window, all data acquired since June 2002 are used to obtain $\hat{R}$ from (13). Since ERS observations (and thus TC inferences) are not available until mid-2003, $\hat{R}$ is held fixed at an initial value of 0.04 $m^3\ m^{-3}$ prior to this point.

[52] In concert, $\hat{Q}$ is adaptively adjusted at the end of each 150 day window using an ad hoc nudging rule

$$\hat{Q}'_{j+1} = \left\{ \begin{array}{ll} 3/2\ \hat{Q}_j & \mathrm{var}(\nu)_j > 1 \\ 3/4\ \hat{Q}_j & \mathrm{var}(\nu)_j <= 1 \end{array} \right\}$$

where

$$\hat{Q}_{j+1} = \left(\hat{Q}_j + \hat{Q}'_{j+1}\right)/2 \qquad (16)$$

and var(ν)$_j$ is the sampled ν-variance within each non-overlapping window j. Since adaptive filtering results are somewhat sensitive to the initial specification of error magnitudes, the entire adaptive analysis is repeated ten times using different starting values of $\hat{Q}$ for June 2002. Reported results reflect the average RMSE obtained for these ten runs.

[53] Despite this change, transitioning from a batch calibration approach to an adaptive framework is associated with little or no degradation in KF performance (Table 2). This result is consistent with *Crow and Reichle* [2008] who demonstrate that, unlike adaptive approaches tasked with the simultaneous estimation of both $\hat{Q}$ and $\hat{R}$, adaptive approaches like (16) can quickly recover $\hat{Q}$ when provided an accurate external estimate of R.

[54] Expanding on earlier results from Figure 3 (based solely on the assimilation of $\theta_{\mathrm{AMSRE\text{-}VU}}$ over the LW site), Figure 8 summarizes results in Table 2 by averaging normalized RMSE results across for all nine possible soil moisture product/watershed combinations (i.e., all three θ_{AMSRE} products assimilated over all three watershed sites). Normalization is based on dividing the RMSE for KF results by the RMSE of the corresponding open loop case. Consequently, a value of one represents no net improvement relative to the case of no data assimilation. Simple direct insertion eliminates (on average) about 12% of the open loop error and another 6% (for a total reduction of 18%) is eliminated by batch ν-whitening. In contrast, our TC-based tuning eliminates either 23% or 24% of the open loop error (depending on whether tuning is performed in batch or adaptive mode). These values are only slightly less than the theoretical maximum (27% of open loop error) realized via the application of a tuned ColKF.

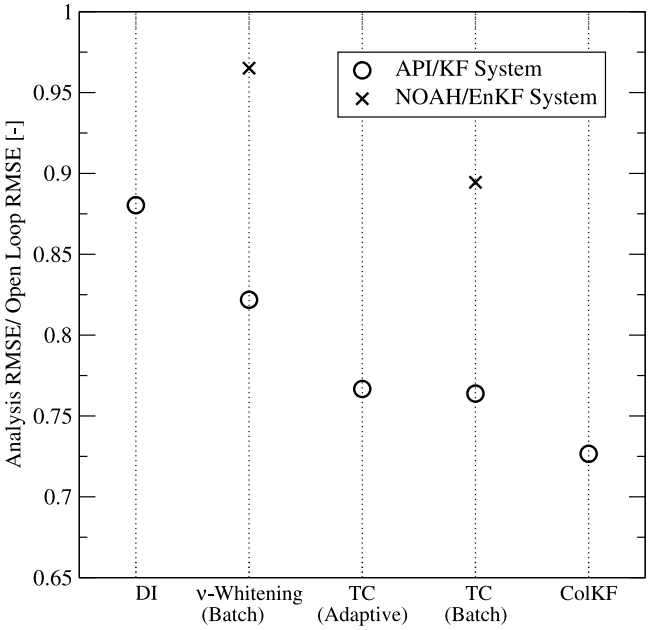

Figure 8. The impact of direct insertion (DI), ν-whitening, and triple collocation (TC) strategies for obtaining $\hat{R}$ on accuracy of KF/API and EnKF/NOAH surface soil moisture predictions (averaged across all soil moisture products and watershed sites). All results are normalized by the RMSE of the no-assimilation open loop case to reflect the fraction of modeling error remaining after assimilation. Also shown are Colored Kalman filter (ColKF) results in which error parameters have been tuned to optimize filter performance.

Table 3. For the API/KF Data Assimilation System, the Comparison Between Sqrt($\hat{R}$) Obtained From ν-Whitening and Optimal Values of Sqrt($\hat{R}$) for Applying the KF Directly to the Raw API Model in (1) or the Anomaly-Based API Model in (17)

Site	Product	Optimal Sqrt(R) Raw (m^3 m^{-3})	ν-Whitening Sqrt($\hat{R}$) Raw (m^3 m^{-3})	Optimal Sqrt(R) Anomaly (m^3 m^{-3})	ν-Whitening Sqrt($\hat{R}$) Anomaly (m^3 m^{-3})
LW	USDA	0.04137	0.01811	0.04388	0.01966
	VU	0.05276	0.02219	0.04281	0.03166
	NASA	0.05718	0.02685	0.05139	0.02608
LR	USDA	0.02671	0.01683	0.02176	0.01736
	VU	0.03632	0.02578	0.03447	0.03569
	NASA	0.02540	0.02166	0.02011	0.01279
WG	USDA	0.01533	0.00723	0.01157	0.00668
	VU	0.01390	0.00686	0.01035	0.00945
	NASA	0.01189	0.00560	0.01195	0.00299

4.3.2. Potential Impact of Seasonality

[55] Another neglected factor in our data assimilation analysis (for both the API/KF and NOAH/EnKF systems) is the impact of differences in seasonal climatologies between modeled and observed surface soil moisture. While θ_{AMSRE} retrievals are rescaled to match the observed probability density function of the model prior to data assimilation, such rescaling is based on a single bulk transformation which does not rectify potential seasonal differences between the two products. If present, such differences likely contribute to autocorrelation in retrieval errors.

[56] In order to examine the magnitude of this contribution, ν-whitening was repeated for the API/KF case where a 31 day moving average climatology is subtracted from θ_{AMSRE} retrievals, the P time series used to force the API model, and ground-based soil moisture measurements used for verification.

[57] Due to the linearity of (1), it can be trivially modified to function in this new anomaly space

$$\mathrm{API}'_i = \gamma_i \mathrm{API}'_{i-1} + P'_i, \tag{17}$$

where the superscript "′" indicates anomalies relative to a 31 day moving window climatology. The decomposition of θ_{AMSRE} retrievals into corresponding anomalies (θ'_{AMSRE}) allows the entire KF-based assimilation process in (1) to (6) to be repeated in an anomaly space where seasonality has been explicitly removed [*Crow et al.*, 2010]. Results from these experiments can then be used to clarify the impact of soil moisture seasonality on previous results.

[58] Experimental results for the KF-based assimilation of θ'_{AMSRE} anomalies into (17) are given in Table 3. "Raw" results reflect earlier $\hat{R}$ results shown in Figure 5 and "anomaly" results duplicate the API/KF data assimilation experiment after the removal of a seasonal climatology from all data sets. While the gap between optimal values of $\hat{R}$ and those obtained from ν-whitening shrinks, a clear underestimation of the $\hat{R}$ persists even for this new "anomaly" case. This suggests that the impact of autocorrelated errors on $\hat{R}$ obtained from ν-whitening remains even after the explicit removal of seasonal cycles from soil moisture data products.

4.4. NOAH/EnKF Analysis Results

[59] While appropriate for initial proof-of-concept results, the API model has a number of well-known shortcomings which hamper its physical interpretation. For instance, since it lacks a maximum storage capacity, time series results for the API model, and other products which have been scaled into the API model climatology, show unrealistic levels of storage for the top 5-cm of the soil column (see, e.g., Figure 2). In addition, the API/KF system cannot capture the impact of non-Gaussian precipitation errors or nonlinear land surface processes. Consequently, it is not obvious whether results presented up to this point are relevant for the assimilation of surface soil moisture retrievals into a more realistically complex (and nonlinear) land surface model using an EnKF. To address this issue, the relative performance of our TC-based calibration approach versus ν-whitening was also assessed for the case of assimilating $\theta_{\mathrm{AMSRE\text{-}USDA}}$ into the NOAH model using an EnKF.

Table 4. For the NOAH/EnKF Data Assimilation System, RMSE Fit to Ground-Based Soil Moisture Observations for the Open Loop (OL), ν-Whitening, and TC Cases

Site	Product	OL (m^3 m^{-3})	ν-Whitening (m^3 m^{-3})	TC (m^3 m^{-3})
LW	USDA	0.05155	0.05049	0.04692
LR	USDA	0.03953	0.04124	0.03900
WG	USDA	0.02185	0.01907	0.01719

[60] Table 4 summarizes EnKF/NOAH results (in terms of the RMSE fit to observed surface soil moisture in all three watersheds) for the NOAH open loop case (lacking any data assimilation) as well as the case of obtaining $\hat{R}$ using both batch ν-whitening and our new TC-based approach (i.e., obtaining $\hat{R}$ from TC and tuning $\hat{Q}$ until var(ν) = 1). Relative to the API model, the NOAH model produces a more accurate open loop simulation and therefore reduced relative improvement upon the assimilation of $\theta_{\mathrm{AMSRE\text{-}USDA}}$. However, as in the KF/API system, better relative surface soil moisture predictions are obtained when applying the TC-based calibration approach versus ν-whitening (Table 4).

[61] Averaged NOAH/EnKF results across all three watersheds in Figure 8 show that switching from batch ν-whitening to our TC-based approach increases the fraction of filtered RMSE in the NOAH/EnKF from about 3% to about 11% of the open loop RMSE. As in the in KF/API case, this improvement is associated with a clear increase in TC-based $\hat{R}$ relative to excessively low values obtained from ν-whitening (Table 5).

[62] One reason for the poor performance of the ν-whitening NOAH/EnKF case in Table 4 and Figure 8 may be unique problems associated with the overestimation of Q when applying an EnKF to a nonlinear land surface model. Excessively low values of $\hat{R}$ obtained from ν-whitening tend to be compensated by an increase in $\hat{Q}$ (Figure 1b). Large $\hat{Q}$, in turn, requires large mean-zero Monte Carlo model perturbations (η) during the EnKF forecast step. When

Table 5. For the NOAH/EnKF Data Assimilation System, Sqrt($\hat{R}$) Acquired From ν-Whitening and TC

Site	Product	ν-Whitening (m^3 m^{-3})	TC (m^3 m^{-3})
LW	USDA	0.01620	0.02880
LR	USDA	0.01520	0.02390
WG	USDA	0.02140	0.03370

applied to nonlinear models, mean-zero perturbations with sufficient variance can cause the soil moisture forecast ensemble to become biased, resulting in degraded EnKF performance [*Ryu et al.*, 2009]. This problem is not encountered when assimilating into a linear model. Therefore, the degrading effects of ν-whitening in the presence of autocorrelated observing errors may be amplified by nonlinear model physics.

5. Summary and Future Work

[63] The accurate estimation of modeling and observing error parameters represents a remaining barrier to the widespread implementation of land data assimilation systems. Here, we describe the theoretical impact of autocorrelated observing errors on techniques for defining such parameters using temporal ν statistics (Figure 1). Given the apparent serial autocorrelation of error present in currently available soil moisture retrieval data sets (Figure 4), these impacts are expected to be relevant for on-going attempts to assimilate surface soil moisture retrievals into land surface models. The theoretical signature associated with autocorrelated observing errors, the systematic underestimation of $\hat{R}$ acquired from ν-whitening (Figure 1b), is observed in real data cases using an API/KF data assimilation system (Figures 3 and 5) and is shown to have a discernible impact on the accuracy of a KF surface soil moisture analysis (Table 2 and Figure 8). An alternative strategy based on using a triple collocation (TC) procedure to independently estimate R, and a ν-variance tuning approach to subsequently obtain $\hat{Q}$, consistently provides better R estimates (Figure 5) and enhances KF surface soil moisture predictions (Table 2 and Figure 8). While this approach does not produce a fully optimal analysis, it is shown to be only slightly worse than a batch-tuned ColKF approach in which autocorrelated errors are optimally considered and appears readily adaptable to an adaptive filtering framework in which all error parameter estimation is done on-line (Table 2). The relative superiority of a TC-based calibration approach to tuning via ν-whitening is also confirmed for the case of assimilating $\theta_{\text{AMSRE-USDA}}$ into the nonlinear, multilayer NOAH land surface model using an EnKF (Table 4 and Figure 8).

[64] A potential extension of this work is addressing its relevance for the direct assimilation of low-frequency (<10 GHz) radiometer T_B observations instead of soil moisture products derived from these observations. The presence of autocorrelated errors in soil moisture retrievals obtained from inverse radiative transfer models implies that forward versions of these same models will likewise produce autocorrelated error in T_B predictions. Consequently, when such forward models are applied during direct T_B assimilation, the autocorrelation problem is simply shifted over to the modeling side of the data assimilation system. Therefore, in order to examine the potential relevance of this work for T_B assimilation, future research is required to examine the possibility of correcting for the impact of autocorrelated modeling errors using a similar set of tools.

[65] Our approach is also somewhat simplified in the sense that the entire modeling error perturbation structure of the NOAH model is reduced to a single scalar parameter Q. Other important model error attributes (e.g., the variance of multiplicative rainfall perturbations or the vertical correlation structure of errors within various soil layers) are fixed and presumed known. Consequently the approach provides only a partial constraint on potentially complex modeling error structure. Future work is required to fully address this issue and evaluate data assimilation performance for each technique over a wider range of model outputs (e.g., root zone soil moisture or surface energy fluxes).

[66] Finally, it should be stressed that results in this analysis are particularly relevant for preparatory activities aimed at the design of data assimilation systems to ingest soil moisture products from the NASA Soil Moisture Active/Passive (SMAP) mission [*Entekhabi et al.*, 2010]. The improved accuracy and spatial resolution of soil moisture retrievals obtained from L band SMAP measurements (versus existing X and C band AMSR-E measurements) should improve the overall performance of data assimilation systems relative to the no-assimilation open loop case. In addition, SMAP's ability to provide both active and passive soil moisture products will feed directly into the TC requirement for multiple independent soil moisture data sources.

[67] **Acknowledgment.** Support for this study was provided by the NASA Terrestrial Hydrology Program through grant NNG05GB61G.

References

Allen, P. B., and J. W. Naney (1991), Hydrology of the Little Washita River Watershed, Oklahoma: Data and Analyses, *USDA Tech. Rep. ARS-90*, U.S. Dep. of Agric., Washington, D. C.

Bolten, J. D., W. T. Crow, T. J. Jackson, X. Zhan, and C. A. Reynolds (2010), Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring, *IEEE J. Sel. Top. Earth Obs. Appl. Remote Sens.*, *3*, 57–66.

Bosch, D. D., J. M. Sheridan, and L. K. Marshall (2007), Precipitation, soil moisture, and climate database, Little River Experimental Watershed, Georgia, United States, *Water Resour. Res.*, *43*, W09472, doi:10.1029/2006WR005834.

Chui, C. K., and G. Chen (1991), *Kalman filtering with Real-Time Applications ColKF*, 2nd ed., 195 pp., Springer, New York.

Cosgrove, B. A., et al. (2003), Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118.

Cosh, M. H., T. J. Jackson, P. J. Starks, and G. Heathman (2006), Temporal stability of surface soil moisture in the Little Washita River Watershed and its applications in satellite soil moisture product validation, *J. Hydrol.*, *323*, 168–177.

Cosh, M. H., T. J. Jackson, S. Moran, and R. Bindlish (2008), Temporal persistence and stability of surface soil moisture in a semiarid watershed, *Remote Sens. Environ.*, *122*, 304–313.

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

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

Crow, W. T., and R. H. Reichle (2008), Comparison of adaptive filtering techniques for land surface data assimilation, *Water Resour. Res.*, *44*, W08423, doi:10.1029/2008WR006883.

Crow, W. T., D. G. Miralles, and M. H. Cosh (2010), A quasi-global evaluation system for satellite surface soil moisture retrievals, *IEEE Trans. Geosci. Remote Sens.*, *48*, 2516–2527.

Daley, R. (1992), The effect of serially correlated observation and model error on atmospheric data assimilation, *Mon. Weather Rev.*, *120*, 164–177.

Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, *123*, 1128–1145.

De Lannoy, G. J. M., P. R. Houser, N. E. C. Verhoest, and V. R. N Pauwels (2009), Adaptive soil moisture profile filtering for horizontal information propagation in the independent column-based CLM2.0, *J. Hydrometeorol.*, *10*, 766–779.

Entekhabi, D., et al. (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proc. IEEE*, *98*, 704–716.

Gelb, A. (1974), *Applied Optimal Estimation*, 374 pp., MIT Press, Cambridge, Mass.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM multisatellite precipitation analysis: Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale, *J. Hydrometeorol.*, *8*, 28–55.

Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. D. Bosch, M. Seyfried, D. Goodrich, M. S. Moran, and J. Du (2010), Validation of advanced microwave scanning radiometer soil moisture products, *IEEE Trans. Geosci. Remote Sens.*, *99*, 1–17, doi:10.1109/TGRS.2010.2051035.

Kumar, S. V., et al. (2006), Land information system—An interoperable framework for high resolution land surface modeling, *Environ. Modell. Software*, *21*, 1402–1415.

Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B. Eylander, and P. R. Houser (2008), A land surface data assimilation framework using the land information system: Description and application, *Adv. Water Resour.*, *31*, 1419–1432.

Mehra, R. K. (1970), On identification of variances and adaptive Kalman filtering, *IEEE Trans. Automat. Control*, *15*(2), 75–184.

Mitchell, H. L., and P. L. Houtekamer (1999), An adaptive ensemble Kalman filter, *Mon. Weather Rev.*, *128*, 416–433.

Miralles, D. G., W. T. Crow, and M. H. Cosh (2010), Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations, *J. Hydrometeorol.*, doi:10.1175/2010JHM1285.1, in press.

Naeimi, V., K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner (2009), An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations, *IEEE Trans. Geosci. Remote Sens.*, *47*, 1999–2013, doi:10.1109/TGRS.2008.2011617.

Njoku, E. G. (2008), *AMSR-E/Aqua daily L3 surface soil moisture, interpretive parameters, and QC EASE-Grids*, July 2002 to December 2007, Boulder, CO, USA: National Snow and Ice Data Center, daily updated digital media.

Owe, M., R. de Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, *113*, F01002, doi:10.1029/2007JF000769.

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

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

Reichle, R. H., R. D. Koster, P. Liu, S. P. Mahanama, E. G. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from AMSR-E and SMMR, *J. Geophys. Res.*, *112*, D09108, doi:10.1029/2006JD008033.

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

Renard, K. G., M. H. Nichols, D. A. Woolhiser, and H. B. Osborn (2008), A brief background on the U.S. Department of Agriculture Agricultural Research Service Walnut Gulch Experimental Watershed, *Water Resour. Res.*, *44*, W05S02, doi:10.1029/2006WR005691.

Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson (2009), Correcting unintended perturbation biases in hydrologic data assimilation using Ensemble Kalman filter, *J. Hydrometeorol.*, *10*, 734–750, doi:10.1175/2008JHM1038.1.

Scipal, K., T. Holmes, R. A. M. de Jeu, V. Naeimi, and W. Wagner (2008), A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophys. Res. Lett.*, *35*, L24403, doi:10.1029/2008GL035599.

W. T. Crow, Hydrology and Remote Sensing Laboratory, USDA Agricultural Research Service, Rm. 104, Bldg. 007, BARC-W, Beltsville, MD, USA. (wade.crow@ars.usda.gov)

M. J. van den Berg, Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium.