Towards the estimation root-zone soil moisture via the simultaneous assimilation of thermal and microwave soil moisture retrievals

Fuqin Li a,b,* , Wade T. Crow a , William P. Kustas a

a USDA-ARS Hydrology and Remote Sensing Lab, Bldg. 007, BARC-West, Beltsville, MD 20705, USA
b National Earth Observation Group, Geoscience Australia, GPO Box 378, ACT 2601, Australia

A R T I C L E   I N F O
Article history:
Received 26 February 2009
Received in revised form 16 November 2009
Accepted 18 November 2009
Available online 22 November 2009

Keywords:
Data assimilation
Microwave remote sensing
Thermal remote sensing
Soil moisture

A B S T R A C T

The upcoming deployment of satellite-based microwave sensors designed specifically to retrieve surface soil moisture represents an important milestone in efforts to develop hydrologic applications for remote sensing observations. However, typical measurement depths of microwave-based soil moisture retrievals are generally considered too shallow (top 2–5 cm of the soil column) for many important water cycle and agricultural applications. Recent work has demonstrated that thermal remote sensing estimates of surface radiometric temperature provide a complementary source of land surface information that can be used to define a robust proxy for root-zone (top 1 m of the soil column) soil moisture availability. In this analysis, we examine the potential benefits of simultaneously assimilating both microwave-based surface soil moisture retrievals and thermal infrared-based root-zone soil moisture estimates into a soil water balance model using a series of synthetic twin data assimilation experiments conducted at the USDA Optimizing Production Inputs for Economic and Environmental Enhancements (OPE3) site. Results from these experiments illustrate that, relative to a baseline case of assimilating only surface soil moisture retrievals, the assimilation of both root- and surface-zone soil moisture estimates reduces the root-mean-square difference between estimated and true root-zone soil moisture by 50% to 35% (assuming instantaneous root-zone soil moisture retrievals are accurate of between 0.020 and 0.030 m$^3$ m$^{-2}$). Most significantly, improvements in root-zone soil moisture accuracy are seen even for cases in which root-zone soil moisture retrievals are assumed to be relatively inaccurate (i.e. retrievals errors of up to 0.070 m$^3$ m$^{-2}$) or limited to only very sparse sampling (i.e. one instantaneous measurement every eight days). Preliminary real data results demonstrate a clear increase in the $R^2$ correlation coefficient with ground-based root-zone observations (from 0.51 to 0.73) upon assimilation of actual surface soil moisture and tower-based thermal infrared temperature observations made at the OPE3 study site.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Soil moisture is a key state variable governing the magnitude and variability of water and energy fluxes along the earth's surface. Given the inherent difficulty of obtaining large-scale measurements of soil moisture using ground-based networks, substantial effort has been dedicated to the development of remote sensing techniques to retrieve soil moisture patterns at continental and global scales. These efforts can be broadly categorized via their use of passive microwave [3,22], active microwave [41] or thermal [1,2,20,21] remote sensing retrieval techniques. Microwave-based approaches exploit the strong dielectric contrast between water and dry soil to directly estimate volumetric water content in the soil surface layer. In contrast, retrievals based on thermal remote sensing attempt to indirectly infer root-zone soil moisture based on an understanding of the thermal response of the vegetation canopy to limitations in available soil water [10,20,31].

Microwave and thermal-based soil moisture estimation techniques each possess their own unique set of advantages and limitations. For example, in the absence of hydrometeors, the atmosphere is approximately transparent at long microwave wavelengths (i.e. C- and L-band). As a result, microwave observations can penetrate through most cloud cover and provide almost daily coverage of the land surface. However, the value of these observations for soil moisture estimation is limited by their shallow penetration (approximately 2 cm for C-band sensors and between 3 and 5 cm for L-band) of the vertical soil column. In addition, the need for microwave observations at long wavelengths and engineering constraints surrounding the maximum size of deployable satellite antennae limits the resolution of passive microwave-based soil moisture retrievals to relatively coarse...
spatial resolutions (>10 km). Active microwave radar can provide finer-scale retrievals but at a generally lower accuracy level.

In contrast, thermal-derived soil moisture estimates are obtainable at lower temporal frequency since retrieval is not possible in the presence of cloud cover. However, because they measure plant response to soil moisture limitation throughout the root-zone, they are capable of inferring soil moisture variations over a deeper portion of the soil column. In addition, thermal remote sensing observations are potentially obtainable at much finer spatial resolutions than microwave observations.

Over the last decade, the assimilation of microwave-derived surface soil moisture into land surface models has been an active area of research [11,13,19,29,35,36,42]. Likewise, a number of past studies have examined the direct assimilation of surface radiometric temperatures (derived from thermal remote sensing observations) into a land model [9,23,26]. Recent work has also focused on the assimilation of energy flux estimates derived from surface radiometric retrievals [34,39] rather than the temperature retrievals themselves. In particular, [10] demonstrates the benefit of assimilating a root-zone soil moisture proxy (based on surface energy flux estimates obtained from thermal remote sensing) relative to the alternative strategy of directly assimilating surface radiometric temperature. However, relatively little work has focused on the simultaneous assimilation of mutually independent soil moisture information obtained from microwave and thermal remote sensing sources.

Using an Ensemble Kalman filter (EnKF), this paper conducts a series of synthetic data assimilation experiments at the USDA Economic and Environmental Enhancement (OPE) experimental site located in Beltsville, Maryland (http://www.ars.usda.gov/Research/docs.htm?docid=8438) to evaluate the relative utility of surface microwave and root-zone thermal soil moisture estimates for constraining model root-zone soil moisture predictions. These experiments represent the appropriate first step in the development of data assimilation systems based on new observation types [14,36,40] and should be completed prior to the assimilation of real measurements. In particular, synthetic experiments allow the performance of data assimilation systems to be assessed in tightly controlled circumstances where the nature of modeling and observation errors is known. For synthetic twin experiments presented here, we place special emphasis on accurately capturing temporal frequency and accuracy limitations in thermal-based soil moisture estimates and, therefore, realistically describing the added value of simultaneously assimilating both microwave and thermal-based soil moisture estimates relative to a baseline case of assimilating only microwave-based surface retrievals. The sensitivities of this added value to variations in the magnitude of modeling and observing errors, local soil texture, root-zone retrieval frequency, and the vertical correlation structure of modeling errors are examined through a series of synthetic experiments. While such synthetic data assimilation results represent an important first step, their results should be subsequently validated via data assimilation experiments involving real observations. As a first step in this direction, we also present results for the assimilation of real observations obtained at the OPE site.

2. The ensemble Kalman filter

The EnKF is based on the generation of an ensemble of model realizations, each perturbed in a Monte Carlo manner, to estimate the state forecast error covariance information required by the standard Kalman filter update equation [4,15]. A decade of application in a variety of earth science fields have demonstrated that it is an effective approach for assimilating observations into moderately non-linear models [16].

Assuming that \( Y(t) \) is vector of land surface state variables at time \( t \), and \( F \) is a potentially non-linear land surface model, the updating of \( Y(t) \) via \( F \) can be expressed as:

\[
\frac{dY(t)}{dt} = F[Y(t) + w]
\]

where \( w \) captures perturbations due to model error arising from any potential combination of inadequate model physics, poorly calibrated parameters, and noisy forcing data. Here such random perturbations are assumed to be mean-zero. The EnKF is based on minimizing the impact of \( w \) via the consideration of independent observations related to land surface states contained in \( Y \). Assuming \( M \) to be the true observation operator, observations \( Z \) (at a discrete time indexed by \( k \)) can be expressed as:

\[
Z_k = M_i[Y(t_k)] + v_k
\]

where the observation perturbation \( v \), a mean-zero, Gaussian random variable with covariance \( C \), represents degradation in the observations due to measurement noise or incomplete knowledge of \( M \). These perturbations are assumed to be statistically independent from the model perturbations introduced in (1).

The EnKF is initialized by the introduction of synthetic Gaussian error into initial conditions and generating an ensemble of model predictions using (1). If \( F \) and \( M \) are linear and all stated assumptions concerning \( v \) and \( w \) are met, then the optimal updating of \( Y \) replicates given the presence of an observation \( Z \) at time \( k \) can be expressed as:

\[
Y_k^e = Y_k^m + K_k[Z_k + v_k - M_i(Y_k^m)]
\]

where:

\[
K_k = [C_{im}(C_{mm} + C_{vv})^{-1}]_{i=k}
\]

State vectors \( Y^m \) and \( Y^e \) refer to the \( i \)th ensemble member before and after updating, respectively. \( K \) in (3) is the Kalman filter gain and defined as a function of: the cross-covariance between \( M_i(Y^m) \) and \( Y^m \) \( C_{im} \), the covariance matrix of \( M_i(Y^m) \) \( C_{mm} \), and \( C_v \) via (4). The extra perturbation term \( v_k \) in (3) is required to ensure the proper spread in the post update ensemble [4]. Both \( C_{im} \) and \( C_v \) are statistically estimated from all individual ensemble realizations and calculated around the ensemble mean. At any point in time, EnKF-based state estimates are obtained by averaging across the resulting ensemble of \( Y \) replicates. Past work has indicated that 30 ensemble members are generally appropriate for the application of the EnKF to one-dimensional land surface models [11,35]. However, in order to be conservative and ensure optimal performance, we use a relatively large ensemble size of 150 for all EnKF simulations presented here.

3. Models

Two separate land surface models are used in this analysis: (i) the thermal remote sensing two-source model (TSM) [25,27,33] and (ii) a one-dimensional water and energy balance (WEB) soil-vegetation–atmosphere transfer model (WEB-SVAT) [10] based on a force-restore concept for the soil water balance [12,32]. The WEB-SVAT model will be used as the basis for the synthetic twin data assimilation experiments described in Section 5. The TSM is used to estimate a root-zone soil moisture proxy based on tower-based thermal infrared temperature observations which are subsequently assimilated into the WEB-SVAT model during the real data analysis described in Section 6.

3.1. The two-source model (TSM)

In the TSM, remotely-sensed surface radiometric temperature \( T_d \) is expressed as the composite temperature of soil and canopy contributions:
$T_s(\phi) \approx \left[ f_c(\phi) T_s^4 + (1 - f_c(\phi)) T_s^4 \right]^{1/4}$

where $T_s$ is observed radiometric surface temperature, $T_c$ canopy temperature, and $T_s$ soil temperature. Soil and canopy contributions to $T_s$ are weighted by a fractional vegetation cover ($f_c$) at view angle $\phi$ estimated as:

$$f_c(\phi) = 1 - \exp\left(-\frac{0.5 \Omega(\phi) \lambda A_l}{\cos(\phi)}\right)$$

where $\Omega$ is a clumping factor to account for non-homogeneous canopy cover [27].

Total surface sensible heat flux ($H$) is separated into vegetated canopy ($H_c$) and soil ($H_s$) components:

$$H = H_s + H_c$$

If soil surface and vegetation canopy fluxes are assumed to be in parallel to each other [27], $H_c$ and $H_s$ can be expressed as:

$$H_c = \rho_c R_c = \frac{T_c - T_A}{R_A}$$

$$H_s = \rho_c R_s = \frac{T_s - T_A}{R_A + R_{s,s}}$$

where $\rho_c$ is the volumetric heat capacity of air, $T_A$ the air temperature, $R_A$ the resistance to heat transfer between the canopy and the reference location where the air temperature and other atmospheric data are measured, and $R_{s,s}$ is the resistance to heat flow in the boundary layer immediately above the soil surface. Both $R_A$ and $R_{s,s}$ are estimated following the approach described in [33].

Other surface energy components are then estimated via energy balance considerations. The surface energy balance can be expressed as:

$$R_n = R_n_S + R_n_C = H + LE + G$$

where $R_n$ is net radiation, $R_n_S$ is net radiation at the soil surface, $R_n_C$ is net radiation at the vegetation canopy, $LE$ is total (soil plus canopy) latent heat flux, and $G$ is soil heat flux. The energy balance for the separate soil and vegetated canopy components are expressed as:

$$R_n_S = H_s + LE_S + G$$

$$R_n_C = H_c + LE_C$$

Following [33], the Priestley–Taylor formula is used to calculate canopy latent heat flux:

$$LE_C = \frac{x_{PT} R_n_C}{\alpha + \gamma}$$

where $x_{PT}$ is the Priestley–Taylor parameter for unstressed conditions (~1.3), $f_c$ the fraction of the LAI that is green (typically $f_c = 1$), $\alpha$ the slope of the saturation vapor pressure versus temperature curve, and $\gamma$ the psychrometric constant (0.0066 kPa C$^{-1}$).

Based on (11), latent heat flux from the soil surface ($LE_S$) is solved for as the residual:

$$LE_S = R_n_S - G - H_S$$

with $G$ estimated as a fraction of $R_n_S$:

$$G = C_G R_n_S$$

and $C_G$ is expressed as a function of time following [37].

Net radiation incident on both the canopy and soil surface is estimated using the physically-based radiation model of [5]:

$$R_n_C = L_n_C + (1 - \tau_s)(1 - \alpha_C)S$$

$$R_n_S = L_n_S + \tau_s(1 - \alpha_S)S$$

where $S$ is incoming solar radiation, $\tau_s$ the solar surface albedo, and $\alpha_C$ the canopy albedo. Terms $L_n_S$ and $L_n_C$ are downward long-wave radiation fluxes incident on the soil and canopy, respectively, and are estimated as:

$$L_n_S = \exp(-k_s OLAI)|L_{sky} + [1 - \exp(-k_s OLAI)](L_C - L_s)$$

$$L_n_C = [1 - \exp(-k_c OLAI)]|L_{sky} + L_s - 2L_C$$

where $L_{sky}$, $L_s$ and $L_C$ are emitted long-wave radiation fluxes from the sky, soil and canopy calculated from air, soil and canopy temperatures, and $k_s$ is an extinction coefficient. All transmittance, albedo and extinction coefficients for both soil and the vegetative canopy are based on parameterizations presented in [5].

The computational order of the TSM goes as follows: (i) $R_n_S$ and $R_n_C$ are calculated via (16) and (17), (ii) $LE_C$ is estimated using (13), (iii) $H_C$ is calculated using the balance equation in (12), (iv) $H_S$ is inverted to obtain $T_c$ (vi) based on estimated $T_C$ and observed $T_s$, $T_A$ is obtained from (5), and (vi) estimated $T_g$ is used to obtain $H_s$ via (9), and $LE_s$ is calculated as the soil surface energy balance residual in (14).

Within the TSM, the impact of soil water stress on canopy transpiration is diagnosed through the prediction of $LE_s$ by (14). In cases of significant water stress, excessive predictions of $LE_s$ made by assuming unstressed canopy conditions (i.e. $x_{PT} = 1.3$ in (13)) propagate through the TSM equations and lead to excessively low predictions of $H_s$ and $T_c$. Because the entire TSM process is constrained by the observed $T_s$, lower $T_c$ leads to higher $T_s$ via (5) and, eventually, lower $LE_s$. In the typical application of the TSM to sunny, mid-afternoon conditions, negative instantaneous $LE_s$ can be reliably taken as evidence that the original assumption of an unstressed vegetation canopy is flawed and $x_{PT}$ is iteratively reduced until $LE_s > 0$ [25]. Consequently, $LE_s$ indirectly reflects the availability of water within the vegetation root-zone. Since $LE_s$ also contains information about the availability of surface soil moisture (due to the direct impact of surface soil moisture on soil evaporation levels), it can be viewed as an effective proxy for the integrated availability of both surface- and root-zone soil water. Work in [10] exploits this potential to define uncorrected $LE_s$ values, calculated assuming $x_{PT} = 1.3$, as a vertically-integrated soil moisture flux ($S_{STSM}$) and presents validation results demonstrating a strong correlation between $S_{STSM}$ and root-zone soil moisture measurements made at the USDA OPE site.

Error in $S_{STSM}$ retrievals can arise from a variety of sources (e.g. noise in satellite $T_s$ retrievals, miss-specified LAI or $\Omega$ inputs and uncertain radiation forcing) and is therefore difficult to estimate a priori. For our analysis, the ultimate source of these errors is relatively unimportant; however, their aggregate impact on overall $S_{STSM}$ accuracy is critical for determining the added value afforded by the assimilation of root-zone soil moisture retrievals. To account for uncertainty regarding the typical magnitude of error in $S_{STSM}$ retrievals, we will vary assumed $S_{STSM}$ error over a wide range of possible values in the synthetic data assimilation experiments to follow.

3.2. The water and energy balance soil–vegetation–atmosphere transfer (WEB-SVAT) model

The WEB-SVAT model is based on the merger of the force-store model proposed by [12,32], and subsequently adapted by [30], with a two-layer vegetation/soil energy balance formulation utilizing a vertical canopy structure identical to that employed by the parallel version of the TSM [10].

The model separates the vertical soil column into surface- and root-zone soil layers. These two zones are assumed to be vertically overlapping such that the surface zone constitutes the top fraction of the root-zone. The balance equation for the surface-zone soil moisture is then given as:
\[
\frac{d \theta_{LZ}}{dt} = C_1 \left[ P_e - LE_c (\rho \lambda)^{-1} - C_2 \left( \theta_{LZ} - \theta_{sat} \right) \right]
\]

(20)

where \( P_e \) is precipitation infiltrating into the soil, \( \tau \) the frequency of diurnal variations, \( \rho \) the density of water, \( \lambda \) the latent heat of vaporization for water and \( d_L \) the depth of the surface zone. Variables \( C_1 \) and \( \theta_{sat} \) are also parameterized as functions of soil moisture and soil texture following [30].

The transfer rate coefficient \( C_2 \) controls the degree of coupling between the surface- and root-zone and, following [32], is defined as:

\[
C_2 = C_{2ref} \left( \frac{\theta_{LZ}}{\theta_{sat} - \theta_{LZ} + \theta_i} \right)
\]

(21)

where \( C_{2ref} \) is a soil-dependent parameter used to describe the magnitude of diffusive water flux. Typical \( C_{2ref} \) values for common soil texture types are listed in Table 1. In (21), \( \theta_{sat} \) is the saturated soil moisture content and \( \theta_i \) a small numerical value which limits \( C_2 \) at saturation.

The analogous root-zone balance equation is expressed as:

\[
\frac{d \theta_{LZ}}{dt} = 1 \frac{1}{d_R} \left[ P_e - LE_c (\rho \lambda)^{-1} - LE_s (\rho \lambda)^{-1} - Q \right]
\]

(22)

where \( d_R \) is the root-zone depth and \( Q \) the drainage output the bottom of the root-zone. Drainage flux is modeled as a function of saturated hydraulic conductivity \( K_s \), porosity \( \theta_{sat} \), and the pore size distribution index parameter \( b \) as:

\[
Q = K_s \left( \frac{\theta_{sat} \theta_{LZ}}{\theta_{sat} - \theta_{LZ} + \theta_i} \right)^{2b+3}
\]

(23)

Following modifications made by [10], WEB-SVAT latent heat fluxes from soil and vegetation are expressed as:

\[
LE_s = \rho C_{F} \gamma^{-1} \left[ e_s(T_S) - e_a \right] / (R_{AS} + R_S + R_S)
\]

(24)

\[
LE_c = \rho C_{F} \gamma^{-1} \left[ e_c(T_C) - e_a \right] / (R_c + R_S)
\]

(25)

where \( e_s \) is the saturation vapor pressure and \( e_a \) the near-surface water vapor pressure, \( R_S \) the soil resistance to \( LE_s \), and \( R_c \) the stomatal resistance to \( LE_c \). Soil and stomatal resistance terms are made as functions of surface- and root-zone soil moisture, respectively. In particular, following [38], \( R_S \) is expressed as:

\[
R_S = \exp[ A - B (\theta_{LZ}/\theta_{sat})]
\]

(26)

and \( R_c \) estimated from

\[
R_c = \left( \frac{R_{c,max}}{R_{c,min}} \right)^{\frac{\theta_{LZ} - \theta_{sat}}{\theta_{sat} - \theta_{LZ}}} \left[ \frac{\theta_{sat} - \theta_w}{\theta_{sat} - \theta_{LZ}} + R_{c,max} \right]
\]

\[
\theta_{LZ} < \theta_w
\]

(27)

where \( A \) and \( B \) are empirical constant, and \( \theta_w \) are the volumetric soil moisture levels at which canopy stress and wilting start, respectively. Resistance extremes \( R_{c,max} \) and \( R_{c,min} \) are specified based on typical literature values and consideration of soil texture.

To maximize consistency with the TSM, other components of the WEB-SVAT surface energy balance are calculated using TSM formulations presented earlier in Section 3.1. Specifically, both \( R_T \) and \( R_{AS} \) are estimated following [33]. \( R_T \) and \( R_{AS} \) are calculated using (16) and (17), surface albedo and emissivity values from [5], \( H_c \) and \( H_s \) using (8) and (9), and \( G \) using (15). However, unlike the TSM, \( T_c \) observations are not utilized and component temperatures \( T_C \) and \( T_S \) are diagnosed through the simultaneous solution of the coupled surface and canopy energy balance equations using a numerical root finder. This solution allows for estimation of all components of the soil and canopy energy balance and the subsequent updating of (20) and (22) using observed \( P \) and estimated \( LE_s \) and \( LE_c \).

4. Model and EnKF application

Our analysis is based on a combination of real and synthetic data assimilation studies carried out during four consecutive (2001–2004) growing seasons at the USDA-ARS OPE3 experimental site in Beltsville, Maryland.

4.1. The OPE3 study site

The OPE3 site consists of a series of heavily-instrumented, zero-order watersheds maintained by the United States Department of Agricultural (USDA) Hydrology and Remote Sensing Laboratory. During the growing season, land cover at the site is rain-fed cultivated corn (corn varieties were Pioneer 34K78 (BT), 33A14, 38A25 and 33K81 in 2001, 2002, 2003 and 2004, respectively). Soil texture at the site is 2- to 3-m of sandy loam overlying an impermeable clay lens [17]. Thirty-minute micrometeorological data required by one or both of the TSM and WEB-SVAT models (e.g. solar radiation, air temperature, vapor pressure, wind speed and precipitation) are obtained from instrumentation mounted on a 10-m tower above the corn canopy. A Kipp and Zonen, Inc. CNR-1 radiometer at 4.5 m above ground level (agl) is used to measure incoming and outgoing shortwave and longwave radiation [Trade and company names are given for the benefit of the reader and imply no endorsement by USDA]. Net radiation was also measured with a Kipp and Zonen NR-Lite at 2.5 m agl and a Radiation Energy Balance Systems, Inc. Q7 net radiometer at the same height agl to guarantee a continuous record of net radiation observations. Radiometric surface temperature \( T_S \) is measured using two separate downward-looking Apoge Inc. IRTS-P3 infrared thermocouple sensors mounted at 4.5 m agl. Profile soil moisture measurements are taken from EnviroSCAN capacitance probe sensors at 10, 30, 50 and 80 cm depths installed at five separate locations within 50 m of the instrument tower [17]. A vertically-integrated top 1 m estimate of root-zone soil moisture \( (\theta_{1m}) \) is obtained by:

\[
\theta_{1m} = 0.2 \times \theta_{10 \ cm} + 0.2 \times \theta_{30 \ cm} + 0.2 \times \theta_{50 \ cm} + 0.4 \times \theta_{80 \ cm}
\]

(28)

The resulting \( \theta_{1m} \) measurements are averaged across all five capacitance sites and used to validate real data assimilation results in Section 6.

4.2. Model parameterization at the OPE3 site

Fig. 1 contains daily time series (plotted against days after planting) of leaf area index (LAI) values assumed for WEB-SVAT model. The times series are based on yearly planting/emergence information (Table 2) and (at least) one or two plot measurements conducted at the OPE3 site during peak canopy conditions. Plant canopy height \( h_C \) are assumed to follow the same piecewise-linear
model as LAI except vertically rescaled to have a minimum of 0.1 m and a maximum of 2 m. Likewise, roughness lengths are assumed equal to \( h_c/8 \), and zero plane displacement height was set to \( 2h_c/3 \). Canopy greenness \( (f_c) \) and clumping fraction \( (\Omega) \) values are set to a constant value of one. Stomatal resistance values \( R_{c,min} \) and \( R_{c,max} \) are assumed to be 100 and 1000 \( \text{s m}^{-1} \). Following [24], the unit-less parameters \( A \) and \( B \) in (26) are assigned values of 8.2 and 4.3, respectively. At the start of each growing season, WEB-SVAT soil moisture variables are initialized to match capacitance probes measurements at the OPE\(^3\) site, and layer depths \( d_a \) and \( d_z \) in (20) and (22) are set to constant values of 10 cm and 1 m in order to match capacitance probe vertical measuring depths. Soil hydraulic parameters (e.g. \( \theta_{sat}, \theta_m, \theta_c^* \) and \( K_s \)) are set to match sandy loam lookup table values presented in [7,32]. However, in order to improve the performance of WEB-SVAT soil moisture predictions, a calibrated \( b \) value of 2.8 was substituted for the generic sandy loam value recommended by Cosby et al. [7]. Sensitivity results demonstrate that this modification has no discernible impact on subsequent data assimilation results. The volumetric soil moisture at which canopy stress and wilting begins are calculated assuming matrix potentials of \(-2.1\) and \(-210 \text{ mb} \) for the onset of stress and wilting. Unless otherwise, meteorological and radiation forcing for the WEB-SVAT model are obtained from local instrumentation at the OPE\(^3\) site. The sole exception being the use of the coarser-scale rainfall product for the real data assimilation analysis presented in Section 6.

While aspects of this parameterization are relatively simple, validation results for WEB-SVAT \( \theta_c, T_s \), and total-system latent heat flux \( (L_E + L_E) \) predictions at the OPE\(^3\) site [10] suggest that it is adequately representing key water and energy processes. Here, the WEB-SVAT model is run on a 30-minute time step during the 2001–2004 growing seasons. Based on these parameter values, Fig. 2 plots daily WEB-SVAT root-zone soil moisture predictions obtained at 2 pm local standard time (LST) for the OPE\(^3\) site. The 2002 growing season was relatively dry and followed by a wetter than average 2003 growing season. In contrast, the 2001 and 2004 growing seasons exhibit relatively normal levels of rainfall. For the same set of growing season periods, Fig. 3 plots \( S_{TSM} \) root-zone soil moisture proxies derived from the TSM. Note that these TSM-based results are withheld until analysis with real data (Section 6).

### 4.3. Synthetic twin experiments

Our synthetic twin data assimilation experiment is based on defining output from WEB-SVAT model simulations (forced with local meteorological, radiation and rainfall observations) during the 2001–2004 growing seasons as “truth” at the OPE\(^3\) site. Fig. 2 shows the “truth” 2 pm LST root-zone soil moisture time series derived in this manner. These baseline estimates of both surface- and root-zone soil moisture are then perturbed via the artificial introduction of additive Gaussian noise to generate realistic “retrieved” values which reflect the reduced accuracy at which such quantities can be remotely estimated. Using the EnKF, these retrieved values...
are then re-assimilated back into a version of the WEB-SVAT model which has been degraded via the introduction of additive noise (described below) applied to its internal water balance calculations of surface- and root-zone soil moisture. The added value of assimilating a particular set of observations can then be assessed based on the degree to which the data assimilation product is able to accurately recover “truth” soil moisture results.

Modeling error – expressed by \( w \) in (1) – is incorporated via the direct application of mean-zero additive noise perturbations to both WEB-SVAT surface- and root-zone soil moisture predictions at each half-hourly time step. This noise is meant to capture the aggregate impact of all potential model error sources. Specific random perturbations applied to the surface layer are sampled from a mean-zero Gaussian distribution with a standard deviation of 0.009 m\(^3\) m\(^{-3}\). All model perturbations are uncorrelated in time; however, a cross-correlation of 0.50 is introduced between soil moisture perturbations applied to the surface- and root-zone model layers. Such correlation reflects a tendency for modeling errors to demonstrate vertical correlation due to hydraulic communication between individual soil layers. To reflect increased modeling uncertainty with depth, the standard deviation of root-zone perturbations is obtained by rescaling the standard deviation of surface zone perturbations by the factor \( 1.5 \times \frac{d_s}{d_z} \). Since many of these model error assumptions are relatively arbitrary, sensitivity results will be presented to clarify their impact on key results.

Unless otherwise noted, observation error – expressed by \( \nu \) in (2) – is introduced via the addition of mean-zero, Gaussian noise with a standard deviation of 0.030 m\(^3\) m\(^{-3}\) (for the surface-zone retrievals) and 0.020 m\(^3\) m\(^{-3}\) (for root-zone retrievals) to each observation prior to their assimilation. As with the modeling error parameters described above, these observation error parameters are somewhat arbitrary and will be systematically varied in sensitivity tests to follow. Observational errors are specified to be uncorrelated in time and mutually independent. The statistical properties of perturbations applied to the model and soil moisture retrievals are assumed to be perfectly known and used to optimally parameterize the EnKF during subsequent synthetic twin experiments.

All root-zone soil moisture retrievals are assumed to occur at 2 pm local solar time (LST). Past sensitivity tests indicate comparable data assimilation results are obtainable for any retrieval time between about 11 am and 2:30 pm LST [10]. Unless otherwise noted, surface soil moisture retrievals are initially assumed to occur at 6 am LST to reflect the expected overpass time of upcoming satellite missions. However, the impact of modifying this time-of-day is examined below.

5. Synthetic twin experiment results

As noted above, results will be presented for a series of synthetic twin data assimilation experiments in which assimilated observations are synthetically generated using the WEB-SVAT model, and real data experiments (presented in Section 6) in which assimilated observations are obtained from actual field observations at the OPE\(^3\) site. All data assimilation experiments are conducted for three separate cases. The first case is an “open loop” case in which Gaussian model error is added to WEB-SVAT surface- and root-zone soil moisture predictions with no subsequent attempt to correct this error via data assimilation. The second “surface-only” case is based on correcting the open loop case by assimilating only surface soil moisture retrievals into the WEB-SVAT model using a 150-member ensemble Kalman filter (EnKF). The third “surface- and root-zone” case is similar to the second one except that both surface- and root-zone soil moisture retrievals are simultaneously assimilated into the WEB-SVAT model. In the following sub-sections, synthetic twin experiment results derived from these three cases will be used to describe the impact of retrieval and modeling attributes on the added value of assimilating both surface- and root-zone soil moisture retrievals into the WEB-SVAT model versus a baseline case of assimilating only surface soil moisture retrievals.

5.1. Impact of root-zone soil moisture retrieval accuracy

For 2001, Fig. 4 shows 2 pm LST root-zone soil moisture predictions derived from the open loop case, the case of assimilating only surface soil moisture and the case of simultaneously assimilating both surface- and root-zone soil moisture retrievals. Subsequent figures will be based on evaluating the root-mean-square difference (RMSD) between these results and baseline “truth” soil moisture. Combining all four years (2001–2004) of results, the open loop case leads to a volumetric RMSD of 0.030 m\(^3\) m\(^{-3}\) between truth and open loop root-zone soil moisture (Fig. 5a). WEB-SVAT/EnKF root-zone soil moisture RMSD is reduced to 0.019 m\(^3\) m\(^{-3}\) when 6 am surface soil moisture retrievals (with a root-mean-square (RMS) error of 0.030 m\(^3\) m\(^{-3}\)) are assimilated into the WEB-SVAT model using a 150-member EnKF and compared to comparable 6 am truth soil moisture results.

For the combined surface- and root-zone data assimilation case, variations in the assumed time-of-day for surface soil moisture (6 am) and root-zone soil moisture (2 pm) retrievals make it difficult to define the correct time-of-day for comparisons against baseline “truth” soil moisture. Because EnKF performance degrades as you move away from update times, choosing a comparison time that differs from the retrieval time may artificially enhance the value of one observation type and undercut our stated goal of objectively comparing the relative value of each. To solve this problem, Fig. 5a also plots results associated with the assimilation of a 2 pm LST surface soil moisture product. Relative to assimilation of 6 am LST surface soil moisture retrievals, the 2 pm LST product leads to only a slight (on the order of 0.001 m\(^3\) m\(^{-3}\)) reduction in root-zone RMSD (when compared to 2 pm truth predictions). The impact is even smaller when evaluated in term of surface-zone RMSD (Fig. 5b). Consequently all subsequent analyses are based on an assumed retrieval acquisition time of 2 pm for all soil moisture estimates and comparison against 2 pm LST truth values.

The simultaneous assimilation of both 2 pm LST surface- and root-zone soil moisture estimates into the model leads to further improvement in the accuracy of the WEB-SVAT/EnKF root-zone
soil moisture predictions. At an assumed root-zone retrieval accuracy of 0.030 m$^3$ m$^{-3}$, the addition of root-zone soil moisture estimates reduces the RMSD in EnKF root-zone predictions by approximately 35% (from 0.017 m$^3$ m$^{-3}$ to 0.011 m$^3$ m$^{-3}$) relative to the assimilation of only 2 pm LST surface soil moisture estimates. Naturally, this improvement is particularly large when highly accurate root-zone soil moisture retrievals are assimilated. However, a ~20% relative reduction in RMSD (from 0.017 to 0.014 m$^3$ m$^{-3}$) is found even when error in root-zone soil moisture retrievals is increased to 0.070 m$^3$ m$^{-3}$. Past tower-based work at the OPE site indicates that root-zone soil moisture estimates derived using the retrieval technique in Section 3.1 can reproduce ground-based root-zone soil moisture observations to within 0.030 m$^3$ m$^{-3}$ [10], and the application of similar approaches to geostationary satellite data leads to errors of between 0.040 and 0.050 m$^3$ m$^{-3}$ [18]. Consequently, Fig. 5 suggests that the assimilation of root-zone soil moisture estimates is beneficial even in cases where retrievals are of comparably low quality relative to current expectations for thermal remote sensing algorithms. Fig. 5b is analogous to Fig. 5a except for it contains surface-zone (and not root-zone) soil moisture results. In contrast to Fig. 5a, the assimilation of root-zone soil moisture estimates (regardless of their accuracy) has little or no impact on the subsequent accuracy of WEB-SVAT/EnKF surface soil moisture predictions.

Better root-zone soil moisture predictions should also lead to the improved growing season representation of total system (plant transpiration plus soil evaporation) evapotranspiration ($E_T$). This potential is examined in Fig. 6 which replicates Fig. 5 for the case of daily total $E_T$ RMSD. As in the preceding soil moisture analysis, true $E_T$ is obtained by the “truth” simulation of the WEB-SVAT model (Section 4.3). Compared to the truth simulation, the synthetically perturbed open loop case leads to a daily $E_T$ RMSD of 0.89 mm. For comparison, the average daily total $E_T$ observed at the site is about 3 mm. This RMSD is reduced to 0.58 mm upon the assimilation of surface soil moisture retrievals. The dual assimilation of both surface- and root-zone soil moisture retrievals leads to a further reduction in $E_T$ RMSD. However, the magnitude of this reduction is sharply reduced at moderate and high levels of root-zone soil moisture retrieval error. For instance, the assimilation of root-zone retrievals with a (root-mean-square) error of 0.040 m$^3$ m$^{-3}$ still leads to a substantial reduction in root-zone soil moisture RMSD (see Fig. 5a) but has only a minor positive impact on $E_T$ (Fig. 6).

There are several potential reasons for the reduced impact of root-zone retrieval assimilation on $E_T$ RMSD. First, periods of energy-controlled canopy transpiration are relatively common at the site during the growing season. Fig. 2 plots the critical soil moisture $\theta^*$ at which the WEB-SVAT model transitions between energy and water control of surface energy fluxes. Substantial periods of time during the 2001, 2003 and 2004 growing seasons are spent above this line. During these periods, accuracy gains in root-zone soil moisture do not translate into improved WEB-SVAT $E_T$ predictions and are squandered from an energy flux prediction perspective. Periods of water-limited direct soil evaporation, as opposed to total system $E_T$, are much more common at the site. However, the positive impact of soil moisture assimilation on these predictions is already captured by the assimilation of surface soil moisture and is not enhanced when root-zone soil moisture retrievals are also assimilated. Finally, the abrupt non-linear transition between water- and energy-limited canopy resistance at $\theta^*$ in (27) poses a challenge for estimating $E_T$ using the EnKF. This nonlinearity produces a non-Gaussian $E_T$ ensemble and, subsequently, sub-optimal WEB-SVAT/EnKF flux predictions. Therefore, improvements in $E_T$ should not be trivially regarded as direct consequence of improved root-zone soil moisture prediction.

Synthetic twin data assimilation experiments can also be conducted to examine the sensitivity of EnKF results to uncertain model and observation characteristics. For example, results in Fig. 5 are based on an assumed root-mean-square error of 0.030 m$^3$ m$^{-3}$ and 10-cm vertical measurement depth for microwave-based surface

![Fig. 5. The impact of root-zone soil moisture retrievals error on the RMSD between truth soil moisture and (a) root-zone and (b) surface-zone predictions obtained from the open loop, surface assimilation only and dual surface- and root-zone assimilation EnKF cases.](image_url)

![Fig. 6. The impact of root-zone soil moisture retrieval error on the RMSD between true evapotranspiration ($E_T$) and predictions obtained from the open loop, surface assimilation only and dual surface and root-zone assimilation EnKF cases.](image_url)
soil moisture retrievals. Increasing the assumed error in surface soil moisture retrievals reduces the ability of the EnKF to correct surface soil moisture but has a relatively minor impact on root-zone soil moisture corrections (Table 3). Consequently, results shown in Fig. 5 are representative of results for assumed errors in surface soil moisture retrievals between about 0.020 and 0.060 m$^3$ m$^{-3}$. Likewise, results in Table 4 demonstrate relatively little sensitivity to variations in the assumed vertical measurement depth for surface soil moisture retrievals (dsz). In order to match the vertical support of ground-based surface soil moisture measurements at the OPE3 site, the default value of dsz is set equal to 10 cm. However, results for this default choice are generally representative for any choice of dsz between 5 and 20 cm (Table 4).

Another uncertain quantity is the standard deviation of random perturbations assumed to represent WEB-SVAT modeling error. Fig. 5 is generated using an assumed standard deviation of 0.009 m$^3$ m$^{-3}$ for perturbations applied to half-hourly WEB-SVAT surface-zone soil moisture predictions (the magnitude of root-zone predictions is determined by the approach described in Section 4.3). Table 5 examines the impact of either raising or lowering this value. As expected, error in EnKF root-zone soil moisture predictions increases as assumed modeling error rises. In addition, relative to the surface-only data assimilation cases, the performance of dual surface- and root-zone data assimilation deteriorates less with increasing model error. Nevertheless, patterns seen in Fig. 5 are not qualitatively affected by these variations.

### 5.2. Inter- and intra-annual variability

Synthetic data assimilation results can also be parsed to examine both within and between season variability in results. Note that open loop results in Fig. 4 for 2001 are based on the synthetic introduction of error in the “truth” simulation. Therefore, real seasonal variability is obscured by introduced randomness. Relative to the single year shown in Fig. 4, more robust results are obtainable by averaging results across all four years of available growing season data. Using this approach, Table 6 breaks down aggregated results for all four growing seasons as a function of corn crop growth stage (see Table 2). Results for all growth stages are constricted to a very narrow range of RMSD (between 0.017 and 0.019 m$^3$ m$^{-3}$ for the surface-only assimilation case and between 0.010 and 0.011 m$^3$ m$^{-3}$ for the surface and root-zone case) – indicating little or no seasonal variation in the effectiveness of either data assimilation approach within a particular growing season. Likewise, breaking down total growing season data assimilation RMSD results on a year-by-year basis leads to only minor inter-annual variability (on the order of 0.001–0.002 m$^3$ m$^{-3}$ – see Table 7). However, it should be noted, that temporal processes not explicitly accounted for in our synthetic analysis (e.g. the impact of variations in canopy structure on the accuracy of root-zone soil moisture retrievals) could introduce important seasonal and inter-annual temporal variability. For this reason, this analysis is repeated for the real data assimilation case examined later in Section 6.

---

### Table 3

For an assumed root-zone soil moisture observation error of 0.020 m$^3$ m$^{-3}$, the impact of surface soil moisture retrieval error on the RMSD between truth soil moisture and open loop, surface-only and surface- and root-zone soil moisture data assimilation results.

<table>
<thead>
<tr>
<th>Surface soil moisture observing error [m$^3$ m$^{-3}$]</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface RMSD [m$^3$ m$^{-3}$]</td>
<td>Root-zone RMSD [m$^3$ m$^{-3}$]</td>
<td>Surface RMSD [m$^3$ m$^{-3}$]</td>
<td>Root-zone RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>0.010</td>
<td>0.048</td>
<td>0.030</td>
<td>0.010</td>
</tr>
<tr>
<td>0.020</td>
<td>0.048</td>
<td>0.030</td>
<td>0.018</td>
</tr>
<tr>
<td>0.030</td>
<td>0.048</td>
<td>0.030</td>
<td>0.025</td>
</tr>
<tr>
<td>0.040</td>
<td>0.048</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td>0.050</td>
<td>0.048</td>
<td>0.030</td>
<td>0.034</td>
</tr>
<tr>
<td>0.060</td>
<td>0.048</td>
<td>0.030</td>
<td>0.037</td>
</tr>
<tr>
<td>0.070</td>
<td>0.048</td>
<td>0.030</td>
<td>0.039</td>
</tr>
</tbody>
</table>

### Table 4

The impact of assumed surface layer depth on the RMSD between truth soil moisture and open loop, surface-only and surface and root-zone soil moisture data assimilation results.

<table>
<thead>
<tr>
<th>Surface layer depth [cm]</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface RMSD [m$^3$ m$^{-3}$]</td>
<td>Root-zone RMSD [m$^3$ m$^{-3}$]</td>
<td>Surface RMSD [m$^3$ m$^{-3}$]</td>
<td>Root-zone RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>2</td>
<td>0.041</td>
<td>0.030</td>
<td>0.023</td>
</tr>
<tr>
<td>5</td>
<td>0.046</td>
<td>0.030</td>
<td>0.025</td>
</tr>
<tr>
<td>10</td>
<td>0.048</td>
<td>0.030</td>
<td>0.025</td>
</tr>
<tr>
<td>20</td>
<td>0.051</td>
<td>0.030</td>
<td>0.025</td>
</tr>
</tbody>
</table>

### Table 5

The impact of modeling error magnitude on the RMSD between truth soil moisture and open loop, surface-only and surface and root-zone soil moisture data assimilation results. Standard deviations refer to surface-zone perturbations which are transformed into root-zone perturbations via the approach described in Section 4.3.

<table>
<thead>
<tr>
<th>Standard deviation of modeling noise [m$^3$ m$^{-3}$]</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface RMSD [m$^3$ m$^{-3}$]</td>
<td>Root-zone RMSD [m$^3$ m$^{-3}$]</td>
<td>Surface RMSD [m$^3$ m$^{-3}$]</td>
<td>Root-zone RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>0.005</td>
<td>0.028</td>
<td>0.016</td>
<td>0.019</td>
</tr>
<tr>
<td>0.009</td>
<td>0.048</td>
<td>0.030</td>
<td>0.025</td>
</tr>
<tr>
<td>0.013</td>
<td>0.069</td>
<td>0.043</td>
<td>0.027</td>
</tr>
</tbody>
</table>
Table 6
For the open loop, surface-only and surface and root-zone synthetic twin soil moisture data assimilation cases, variation in root-zone soil moisture RMSD (m$^3$ m$^{-3}$) and $R^2$ with corn growth stage. Results are based on pooling similar stages during the 2001–2004 growing seasons (Table 2).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
<td>$R^2$ [-]</td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>Planting to R1</td>
<td>0.028</td>
<td>0.68</td>
<td>0.018</td>
</tr>
<tr>
<td>R1–R6</td>
<td>0.026</td>
<td>0.74</td>
<td>0.017</td>
</tr>
<tr>
<td>R6 to simulation end</td>
<td>0.034</td>
<td>0.66</td>
<td>0.019</td>
</tr>
<tr>
<td>Entire season</td>
<td>0.028</td>
<td>0.77</td>
<td>0.018</td>
</tr>
</tbody>
</table>

Table 7
For the open loop, surface-only and surface and root-zone synthetic twin soil moisture data assimilation cases, yearly variations in root-zone soil moisture RMSD (m$^3$ m$^{-3}$) and $R^2$.

<table>
<thead>
<tr>
<th>Year</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
<td>$R^2$ [-]</td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>2001</td>
<td>0.024</td>
<td>0.73</td>
<td>0.017</td>
</tr>
<tr>
<td>2002</td>
<td>0.034</td>
<td>0.58</td>
<td>0.020</td>
</tr>
<tr>
<td>2003</td>
<td>0.024</td>
<td>0.72</td>
<td>0.016</td>
</tr>
<tr>
<td>2004</td>
<td>0.026</td>
<td>0.56</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Fig. 7. The impact of root-zone soil moisture observation frequency on the RMSD between truth soil moisture and root-zone predictions obtained from the open loop, surface assimilation only and dual surface- and root-zone assimilation EnKF cases.

5.3. Impact of temporal retrieval frequency

In addition to accuracy limitations described above, temporal gaps in retrieval coverage due to cloud cover pose an obvious limitation on root-zone soil moisture estimates derived from thermal infrared remote sensing. To test how the temporal frequency of observations influences the accuracy of WEB-SVAT/EnKF root-zone soil moisture predictions, a series of synthetic twin data assimilation experiments were carried out in which the frequency of root-zone soil moisture retrievals was systematically varied. In parallel, it was assumed that, due to the all-weather capabilities of microwave remote sensing, daily retrievals of surface soil moisture are always available. Fig. 7 shows how root-zone soil moisture retrieval frequency affects the accuracy of WEB-SVAT/EnKF root-zone soil moisture predictions. For a root-mean-square error of 0.02 m$^3$ m$^{-3}$ in volumetric root-zone soil moisture retrievals (Fig. 7), root-zone RMSD increases from 0.009 to 0.016 m$^3$ m$^{-3}$ when the availability of root-zone soil moisture retrievals is reduced from everyday (at 2 pm LST) to once every 8 days. Assuming daily 2 pm LST observations are available from a polar orbiting or geostationary satellite, and a very conservative minimum threshold of 800 W m$^{-2}$ in surface solar radiation to define “cloud free” 2 pm LST conditions, root-zone soil moisture retrievals are available at the OPE3 size with at an average frequency of at least one per four days. Therefore, a restriction of retrievals to once every eight days can be considered highly conservative. Nevertheless, even after such a pessimistic reduction in the assumed temporal availability of root-zone retrievals, RMSD levels for the dual assimilation case remain below those found for the case of assimilating only daily surface soil moisture estimates. This lack of sensitivity to the frequency of root-zone soil moisture retrievals is likely caused by the relatively large water storage capacity of the root-zone layer since variables possessing substantial memory generally require less frequent updating via sequential data assimilation [42].

5.4. Impact of soil texture

In the WEB-SVAT model, vertical moisture coupling between surface- and root-zone soil layers is largely controlled by the parameter $C_2$ in (20) which, in turn, is parameterized as a function of root-zone soil moisture and the soil texture dependent parameter $C_{2ref}$ in (21). Table 1 lists typical values of $C_{2ref}$ obtained from Noilhan and Planton [32] for a range of soil texture types. Values of $C_{2ref}$ range between 0.3 for clay to 3.9 for sand. The magnitude of $C_{2ref}$ determines the degree to which variations in root-zone soil moisture are communicated to the surface zone, and, consequently, the degree to which the root-zone can be accurately constrained via the assimilation of only surface soil moisture estimates. To illustrate this, Fig. 8 shows data assimilation results for a variety of $C_{2ref}$ choices. For a clay soil with low $C_{2ref}$, and therefore limited vertical coupling of the root- and surface-zones, the assimilation of surface soil moisture retrievals is ineffective at correcting root-zone soil moisture model predictions (a RMSD of 0.025 m$^3$ m$^{-3}$ versus 0.030 m$^3$ m$^{-3}$ for the open loop). Consequently, the assimilation of both surface- and root-zone retrievals leads to relatively large added improvement in root-zone soil moisture RMSD (from 0.025 to 0.010 m$^3$ m$^{-3}$). In contrast, when soil is sandy and possesses a larger $C_{2ref}$, strong vertical coupling between the two soil layers substantially improves the ability of surface soil moisture observations to constrain soil moisture in deeper model layers (a RMSD of 0.015 m$^3$ m$^{-3}$ versus 0.029 m$^3$ m$^{-3}$ for the open loop). Consequently, reduced added value is associated with the subsequent inclusion of root-zone soil moisture retrievals in our dual assimilation case (from a RMSD of
cross-correlation between surface- and root-zone soil moisture, as captured by $C_{vm}$ in [4], is reduced and, therefore, the size of update increments applied by the filter to the root-zone. These low-coupling cases present opportunities for significant added improvement upon the dual assimilation of both surface- and root-zone soil moisture retrievals. At zero assumed vertical correlation, Fig. 9 demonstrates a substantial reduction in WEB-SVAT root-zone soil moisture RMSD (from 0.021 m$^3$ m$^{-3}$ to 0.010 m$^3$ m$^{-3}$) between the surface-only and surface- and root-zone data assimilation cases. Conversely, increasing the assumed vertical correlation of modeling errors reduces such opportunities since strong coupling leads to larger values of $C_{vm}$ and more confident updating of the root-zone with only surface-zone retrievals. In such cases, substantial information regarding root-zone soil moisture errors is already available from surface-zone retrievals, and our surface-only data assimilation case is associated with lower RMSD (0.008 m$^3$ m$^{-3}$). Consequently, assimilating root-zone soil moisture retrievals offers little added value for estimating root-zone soil moisture.

While any level of coupling can be accommodated within our synthetic experiments, relatively little is known about what constitutes a realistic level of assumed vertical correlation between soil moisture predictions attained from a layered soil water balance model. Presumably, strong coupling is more likely during relatively wet periods in which precipitation errors actively vertically propagate in highly saturated (and thus highly hydraulically conductive) soils. Conversely, vertical decoupling of errors might be more prevalent for dry periods in which root-zone errors are dominated by modeling errors in evapotranspiration which do not strongly propagate upward into the surface layer (see, e.g. [6,28]). Since root-zone soil moisture retrievals are of greatest value for $E_T$ predictions during dry periods (due to the increased likelihood of water control on surface processes) such decoupling may represent a critical barrier for accurately constraining model $E_T$ estimates using only surface soil moisture retrievals.

6. Real data assimilation results

All results presented to this point are based on a synthetic identical twin methodology in which assimilated observations are synthetically generated by the WEB-SVAT model. However, the data-rich nature of the OPE$^3$ site also allows for real data cases in which actual observations are assimilated and subsequent WEB-SVAT/EnKF results are validated using comparisons with independent ground-based soil moisture measurements. For these experiments, surface-zone soil moisture retrievals are based on 10-cm soil moisture capacitance probe measurements available at the OPE$^3$ site (Section 4.1), and root-zone soil moisture retrievals are based on the $STSM$ root-zone soil moisture proxy derived from tower-based IRT measurements at the OPE$^3$ site and the application of the TSM (Fig. 3 and Section 3.1). Error estimates for subsequent WEB-SVAT/EnKF root-zone soil moisture predictions are derived via comparisons with integrated, top-1-m soil moisture observations created using (28).

Prior to their assimilation, $\sigma_{STSM}$ root-zone soil moisture proxy values are rescaled so their climatology matches that of comparable WEB-SVAT soil moisture predictions. This rescaling is based on the linear transformation:

$$\theta_{z,i}^{STSM} = \mu_{\theta_i} + (S_{TSM} - \mu_{STSM}) \frac{\sigma_{\theta_i}}{\sigma_{STSM}}$$

(29)

where $\mu$ is the temporal mean of soil moisture and $\sigma$ the temporal standard deviation. Subscripts $\theta_i$ and $STSM$ in (29) refer to statistics sampled from WEB-SVAT root-zone predictions and $STSM$ time series results during the 2001–2004 growing seasons. Sampling uncer-
tainty in these statistics can induce errors into $\theta_{\text{TSM}}^{\text{root}}$; however, previous work has shown that a single growing season is sufficient to adequately estimate $\mu$ and $\sigma$ [10]. Time series values of $\theta_{\text{TSM}}^{\text{root}}$ obtained from (29) are subsequently assimilated into the WEB-SVAT model as estimates of root-zone soil moisture.

To maximize the realism of the experiment for regional-scale modeling applications, WEB-SVAT rainfall forcing is obtained from the continental-scale, real-time North American Land Data Assimilation (NLADAS) precipitation product derived from a merger of ground-based rain gauge and weather radar observations [8]. This is in contrast to previous synthetic twin data assimilation results in Section 5 where rainfall forcing data was obtained from a single rain gauge at the OPE3 site. In addition, initial WEB-SVAT soil moisture (both surface- and root-zone) is intentionally underestimated as 0.20 m$^3$ m$^{-3}$ and $b$ is set equal to a default value of 4.3 (versus its calibrated value of 2.8). In order to simulate the impact of their uncertain remote estimation, capacitance probe measurements of 10-cm soil moisture at the OPE3 are artificially perturbed with additive noise sampled from a mean-zero, Gaussian distribution with a standard deviation of 0.030 m$^3$ m$^{-3}$. Using (29), these perturbed surface soil moisture retrievals are then rescaled to match the WEB-SVAT surface soil moisture climatology. Model perturbations applied to create the EnKF ensemble are identical to those employed in the synthetic twin experiments (see Section 4.3). Surface soil moisture retrievals are assimilated daily, but $\theta_{\text{TSM}}^{\text{root}}$ is assimilated only on sufficiently cloud-free days in which instantaneous downward solar radiation measured at the ground is larger than 600 W m$^{-2}$ at 2 pm LST. This results in an average assimilation frequency of about once every two days.

As in the synthetic data assimilation results, real data assimilation experiments are conducted for three separate cases. The first case is an “open loop” case where Gaussian model error is added to WEB-SVAT surface- and root-zone soil moisture predictions with no subsequent attempt to correct this error via data assimilation. The second “surface-only” case is based on correcting the open loop case by assimilating only surface soil moisture observations, and the third “surface- and root-zone” case is similar to the second one except that both surface- and root-zone soil moisture retrievals ($\theta_{\text{TSM}}^{\text{root}}$) are simultaneously assimilated into WEB-SVAT. Data assimilation results are then evaluated based on direct comparisons with actual root-zone soil moisture measurements made at the OPE3 site. Fig. 10 shows open loop, surface and root-zone assimilation and observed soil moisture variations for each for the four examined growing seasons, and Tables 8 and 9 break
However, the inclusion of root-zone soil moisture is generally associated with more added skill (an increase in $R^2$ from 0.23 for the surface-only case to 0.34 for the surface- and root-zone case) during the relatively wet 2003 growing season than during the much drier 2002 season, where $R^2$ actually decreases from 0.88 to 0.80 upon inclusion of root-zone soil moisture in the EnKF. This trend appears to be at odds with previous synthetic twin results, which suggested that the assimilation of root-zone soil moisture estimates was relatively more valuable during dry periods (Section 5.5).

### 7. Summary and conclusions

Using a series of synthetic twin data assimilation experiments conducted at the USDA OPE site, this analysis examines conditions under which the dual assimilation of both surface- and root-zone soil moisture retrievals can improve the prediction of root-zone soil moisture by a land surface model relative to the baseline case of assimilating only surface soil moisture retrievals. Since root-zone soil moisture estimates are based on the inversion of cloud-free surface radiometric temperature retrievals to indirectly infer root-zone soil moisture status, a critical issue is the impact of inherent accuracy and frequency limitations on their subsequent utility. Synthetic data assimilation results demonstrate that direct root-zone soil moisture retrievals are able to improve the representation of root-zone soil moisture estimates (relative to a baseline case of assimilating only surface soil moisture retrievals) even in circumstances in which they are obtained relatively infrequently (once every eight days) or degraded to relatively low accuracy levels ($0.070 \text{ m}^3 \text{ m}^{-3}$). These results suggest that, despite known limitations in thermal-based estimates of root-zone soil moisture, they should provide added value to hydrologic and water resource applications requiring vertically-integrated root-zone soil moisture information.

This general conclusion is relatively stable over the range of conditions encountered during the 2001–2004 growing seasons and during various corn crop growth stages. It also appears robust to variations in assumed surface soil moisture observation error, modeling noise standard deviations and surface soil moisture observing depths. However, the added value of root-zone soil moisture retrievals does vary with soil texture and the assumed vertical correlation structure of soil moisture profile errors. The added value of assimilating root-zone soil moisture estimates is particularly large for clay and silt soils demonstrating reduced diffusive interaction between surface- and root-zone model layers. This lack of coupling hampers the efficient constraint of root-zone

---

### Table 8

<table>
<thead>
<tr>
<th>Stage</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
<td>$R^2$ [-]</td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>Planting to R1</td>
<td>0.037</td>
<td>0.18</td>
<td>0.035</td>
</tr>
<tr>
<td>R1–R6</td>
<td>0.050</td>
<td>0.53</td>
<td>0.048</td>
</tr>
<tr>
<td>R6 to simulation end</td>
<td>0.052</td>
<td>0.57</td>
<td>0.046</td>
</tr>
<tr>
<td>Entire season</td>
<td>0.044</td>
<td>0.51</td>
<td>0.041</td>
</tr>
</tbody>
</table>

---

### Table 9

For the open loop, surface-only and surface and root-zone real soil moisture data assimilation cases, yearly variations in root-zone soil moisture RMSD and $R^2$. Results are based on pooling data from the entire season.

<table>
<thead>
<tr>
<th>Year</th>
<th>Open loop</th>
<th>Surface-only</th>
<th>Surface- and root-zone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
<td>$R^2$ [-]</td>
<td>RMSD [m$^3$ m$^{-3}$]</td>
</tr>
<tr>
<td>2001</td>
<td>0.034</td>
<td>0.28</td>
<td>0.035</td>
</tr>
<tr>
<td>2002</td>
<td>0.050</td>
<td>0.72</td>
<td>0.046</td>
</tr>
<tr>
<td>2003</td>
<td>0.049</td>
<td>0.05</td>
<td>0.045</td>
</tr>
<tr>
<td>2004</td>
<td>0.040</td>
<td>0.13</td>
<td>0.037</td>
</tr>
</tbody>
</table>
soil moisture using surface observations alone and provides an opportunity for enhancing the estimation of root-zone soil moisture using thermal infrared remote sensing retrieval techniques. This opportunity is less viable for sandy soils in which vigorous vertical coupling enables surface soil moisture observations to be extrapolated beyond the root-zone with high levels of confidence. In a similar way, strong vertical correlation between assumed modeling errors in the surface- and root-zone reduces the added benefit of assimilating a root-zone soil moisture proxy. High amounts of correlation mean that information concerning root-zone errors can be accurately inferred from surface observations alone. Conversely, low vertical cross-correlation between soil moisture errors dictates that surface observations are of comparatively less value and direct retrievals of root-zone soil moisture are required to effectively filter root-zone soil moisture modeling errors.

In addition to the set of synthetic experiments described above, the assimilation approach was applied to measurements obtained at the OPE site for a preliminary analysis using real data. As in the synthetic data assimilation case, strong improvement (i.e. a long-term $R^2$ improvement from 0.51 to 0.73 with respect to observed root-zone soil moisture) is found relative to the open loop case upon the dual assimilation of both surface-zone observations and root-zone soil moisture estimates acquired from thermal infrared measurements. However, most of the increase is obtainable via the assimilation of only surface soil moisture retrievals, and, relative to synthetic data assimilation results, more modest added value is associated with the additional assimilation of thermal infrared-based root-zone soil moisture retrievals (e.g. the long-term $R^2$ with actual root-zone soil moisture observations increases only from 0.70 to 0.73). In particular, strong added value was restricted to an improved correlation with observed soil moisture between the start of the corn’s reproductive phase (R1) and physiological maturity (R6) and little, if any, improvement is noted in RMSD. The lack of improvement in RMSD is likely associated with the inability of our data assimilation approach to compensate for model biases contributing to RMSD. In addition, the lack of more consistent correlation-based improvement may be attributable to very sandy soils at the OPE site which reduce the added impact of assimilating root-zone soil moisture retrievals by enhancing the degree of vertical coupling existing between the surface- and root-zones. Additional study at a range of other field sites will likely be required to fully understand quantitative differences observed between synthetic and real data assimilation results. Future work on this topic will be aimed in this direction.

Acknowledgements

Research was supported by NASA Terrestrial Hydrology Program Grant NNX06AG07G entitled “Monitoring of Root-Zone Soil Moisture Using Multi-Frequency Observations of Surface Soil Moisture and Evapotranspiration.” We are grateful to Dr. Timothy Gish from the USDA-ARS Hydrology and Remote Sensing Laboratory (HRSL) and Dr. John Prueger from the USDA-ARS National Laboratory for Agriculture and the Environment, Ames IA who provided expertise and technical oversight of the soil moisture and eddy flux tower measurements at the OPE site. We also would like to thank Mr. Andrew Russ from HRSL who was responsible for maintaining the flux tower instruments, collecting and processing the agronomic data and provided quality assurance of the weather and eddy covariance data.

References

[31] Moran MS. Thermal responses as indicators of biophysical system health and integrity. In: Quattrochi DA, Luvall JC, editors. Thermal infrared remote


[38] Sellers PJ, Heiser MD, Hall FG. Relations between surface conductance and spectral vegetation indices at intermediate (100 m² to 15 km²) length scales. J Geophys Res 1992;97(D17):19033–50.


