Benchmarking a Soil Moisture Data Assimilation System for Agricultural Drought Monitoring

EUNJIN HAN

SSAI, Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland

WADE T. CROW AND THOMAS HOLMES

Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland

JOHN BOLTEN

NASA Goddard Space Flight Center, Greenbelt, Maryland

ABSTRACT

Despite considerable interest in the application of land surface data assimilation systems (LDASs) for agricultural drought applications, relatively little is known about the large-scale performance of such systems and, thus, the optimal methodological approach for implementing them. To address this need, this paper evaluates an LDAS for agricultural drought monitoring by benchmarking individual components of the system (i.e., a satellite soil moisture retrieval algorithm, a soil water balance model, and a sequential data assimilation filter) against a series of linear models that perform the same function (i.e., have the same basic input/output structure) as the full system component. Benchmarking is based on the calculation of the lagged rank cross correlation between the normalized difference vegetation index (NDVI) and soil moisture estimates acquired for various components of the system. Lagged soil moisture/NDVI correlations obtained using individual LDAS components versus their linear analogs reveal the degree to which nonlinearities and/or complexities contained within each component actually contribute to the performance of the LDAS system as a whole. Here, a particular system based on surface soil moisture retrievals from the Land Parameter Retrieval Model (LPRM), a two-layer Palmer soil water balance model, and an ensemble Kalman filter (EnKF) is benchmarked. Results suggest significant room for improvement in each component of the system.

1. Introduction

In water limited ecosystems, soil water content in the root zone is a strong predictor of future vegetation condition (Adegoke and Carleton 2002; Wang et al. 2007). Therefore, characterization of root-zone soil moisture plays a critical role for crop yield forecasting and agricultural drought monitoring systems (Bolten et al. 2010). The availability of satellite-based remote sensing data has accelerated the development of drought early warning systems by providing continuous information in space and time (Hayes et al. 2012). Nonetheless, the shallow sensing depth (top few centimeters) and uncertain accuracy of currently available satellite soil moisture retrievals has necessitated the integration of hydrologic models and surface soil moisture observations from satellites through data assimilation techniques to obtain more accurate root-zone soil moisture estimates. Previous studies have demonstrated that the assimilation of surface soil moisture retrievals can improve the estimation of root-zone soil moisture by a land surface model (LSM; Reichle et al. 2002; Crow and Wood 2003; Reichle and Koster 2005). Specifically for agricultural drought monitoring, Bolten et al. (2010) and Bolten and Crow (2012) describe the benefit of assimilating surface soil moisture retrievals from the Advanced Microwave Scanning Radiometer for Earth Observing System (EOS; AMSR-E) into the U.S. Department of Agriculture (USDA) modified Palmer soil moisture model.

Corresponding author address: Eunjin Han, Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture, 10300 Baltimore Ave., BARC-WEST, Bldg. 007, Rm. 104, Beltsville, MD 20705.
E-mail: eunjin.han@ars.usda.gov

The above-mentioned soil moisture data assimilation systems encompass three major components: 1) surface soil moisture retrieval from primary observations, 2) land surface model prediction, and 3) update (analysis) of the model-predicted soil moisture using a sequential data assimilation filter. First, remotely sensed soil moisture estimates are obtained through appropriate retrieval algorithms from primary measurements (e.g., brightness temperature from passive microwave sensors and backscattering coefficients from active microwave sensors). Second, LSMs take meteorological inputs (e.g., precipitation and air temperature) and produce soil moisture state outputs based on the physics represented in the model. Finally, the model-predicted soil moisture is constrained by the assimilation of observed soil moisture to minimize errors in model state and flux predictions.

Although a number of previous studies have demonstrated the benefits of assimilating satellite-based surface soil moisture into LSMs for root-zone soil moisture prediction, relatively little is known about the relative merits of particular retrieval, modeling, and/or data assimilation strategies. In particular, it remains unclear what level of complexity and/or nonlinearity is appropriate for the system components described above. Traditionally, the performance of these components has been evaluated by comparing their output with a set of independent ground-based data using metrics such as root-mean-square error (RMSE) or correlation coefficient. However, this typical evaluation approach does not take into account the minimum level of expected performance that stems from the inputs into the process. As a consequence, it is difficult to objectively determine the efficiency of each system component.

A possible way to overcome this limitation is by examining the relative change of a target metric against a benchmark set by a competing simpler approach. While the typical evaluation scheme computes RMSE against observations, for example, the benchmarking approach shows a "change" in RMSE versus a competing simpler model. In this way, benchmarking allows us to directly assess the value of the more complex model. Therefore, in order to objectively evaluate the net benefit of any nonlinear processes, it is desirable to compare the complex nonlinear model against a benchmark established with a simpler linear competing model (a minimum reference level). When this benchmarking approach is applied to each component of a data assimilation system, the relative weaknesses/strengths of a specific component can be diagnosed.

Recently, the LSM community has paid increasing attention to benchmarking approaches for more systematic and objective evaluation of model performance (Abramowitz 2005; Abramowitz et al. 2008; Abramowitz 2012; Kumar et al. 2012; Luo et al. 2012). For instance, Abramowitz et al. (2008) adopted two statistical models as benchmarks against which the performances of different LSMs were evaluated. The statistical models took four meteorological inputs (downward shortwave radiation, air temperature, specific humidity, and wind speed) and were trained to produce the output fluxes (sensible heat, latent heat, and net CO_2 exchange) just as the typical LSMs do. Since the statistical models do not involve any physical mechanisms, they can be used to examine how much impact the input variables have on the output fluxes and how efficient the nonlinear model physics are in augmenting this value (Abramowitz 2005).

Here we apply a similar benchmarking approach to evaluate the performance of each component of a soil moisture data assimilation system in current operational use for global agricultural drought monitoring. In particular, we attempt to assess individual components of a drought-monitoring soil moisture data assimilation system and benchmark the efficiency of these components relative to simpler, linear retrieval, modeling, and data integration strategies. In this way, we hope to improve our understanding of skill contributed by various components of the system and ultimately target specific aspects of such systems for improvement.

The performance of the data assimilation components and their benchmarks will be compared using an evaluation metric, the lagged rank correlation between the output of each component and the normalized difference vegetation index (NDVI), in particular, the 1-month lagged correlation. This evaluation approach was proposed by Peled et al. (2010) and was applied by Bolten and Crow (2012) and Crow et al. (2012b) to quantify the skill of different LSMs in predicting variations in vegetative health associated with water stress. The advantage of this particular evaluation approach is that it overcomes the spatial limitations of traditional evaluation approaches for LSMs or retrieval algorithms against in situ observations. For instance, satellite-based soil moisture products are typically evaluated based on comparisons with a small number of point-scale in situ measurements within a coarse-scale retrieval pixel (e.g., Draper et al. 2009; Brocca et al. 2011). These kind of direct comparisons between satellite-based and in situ (point) observations are necessarily limited in spatial extent (Crow et al. 2012a) and have issues such as differences in spatial resolution and vertical support (Owe et al. 2001). While using the correlation with NDVI is somewhat indirect, the benefit of this evaluation approach is that it can be applied over wide spatial areas and is not limited to the relatively small number of sparse sites where high-quality soil moisture information can readily be upscaled to match a satellite remote sensing footprint (Crow et al. 2012a).

Here, remotely sensed surface soil moisture and vegetation optical depth retrievals are obtained from the Land Parameter Retrieval Model (LPRM) developed at VU University Amsterdam and the National Aeronautics and Space Administration (NASA; Owe et al. 2008). Likewise, the two-layer Palmer water balance model (Bolten et al. 2010) is used as a representative LSM. While more modern (and complex) LSMs are available, it is worth noting that the two-layer Palmer model remains the operational water balance tool applied by the USDA Foreign Agricultural Service (FAS) for global agricultural drought forecasting. In addition, past research suggests that when evaluated by our proposed NDVI rank correlation metric, increased LSM complexity does not necessarily lead to improved root-zone soil moisture predictions (Crow et al. 2012b). Finally, an ensemble Kalman filter (EnKF) approach is applied to assimilate LPRM surface soil moisture retrievals into the Palmer two-layer model.

The main objectives of this study are 1) to identify where strengths or deficiencies exist among each component of an integrated soil moisture data assimilation system, 2) to test whether nonlinearity in the model (or algorithm) is adding information in the context of drought monitoring, and 3) to explore opportunities for model improvement by assessing various linear combinations of available input data.

2. Data and model

Our general methodological approach is based on reproducing (and then benchmarking) the remote sensing, modeling, and data assimilation components of an existing operational drought monitoring system. Details on individual components of the system are described below.

a. LPRM

In this study, surface soil moisture observations are derived from AMSR-E through LPRM. AMSR-E is a microwave sensor that operated on board the *Aqua* satellite from 2002 to 2011. Since it was the first satellite mission to produce soil moisture as a standard product and with a prescribed accuracy goal (Njoku et al. 2003; Brocca et al. 2011), its soil moisture products have been widely validated against in situ observations over various regions (Wagner et al. 2007; Draper et al. 2009; Rüdiger et al. 2009; Jackson et al. 2010). Among different retrieval algorithms, LPRM has been proven to perform well, producing relatively high correlations with in situ measurements (Wagner et al. 2007; Draper et al. 2009; Rüdiger et al. 2009).

The LPRM uses H- and V-polarized brightness temperature ($T_{b,H}$ and $T_{b,V}$) from either C band (6.9 GHz)

or X band (10.7 GHz) to retrieve soil moisture (θ_{LPRM}) and vegetation optical depth (τ) simultaneously. The retrieval methodology consists of an iterative optimization of the nonlinear forward radiative transfer model to select the θ_{LPRM} and τ that minimize the difference between predicted and measured $T_{b,H}$. The radiative transfer model is constrained by parameterizing τ as a function of the soil dielectric constant, and the microwave polarization difference index MPDI = $(T_{b,H} - T_{b,V})/(T_{b,H} + T_{b,V})$ (Owe et al. 2001; Meesters et al. 2005). The nonlinear relationship between soil moisture and soil dielectric constant is parameterized based on soil texture data according to Wang and Schmugge (1980). Furthermore, the surface soil temperature (T_s), required for the radiative transfer equation, is linearly related to 36.5-GHz V-polarized data.

C-band data are more sensitive to soil moisture than X-band observations, but are more affected by radio frequency interferences (RFIs). In this study, C-band observations are used for soil moisture retrievals by default, except for regions such as the United States where C-band RFI is problematic. For a more detailed description of the LPRM algorithm, see Owe et al. (2001), De Jeu and Owe (2003), and Meesters et al. (2005).

In this paper, we focus on data from the descending (nighttime) orbit that have generally demonstrated higher correlation with in situ data (Wagner et al. 2007; Draper et al. 2009; Rüdiger et al. 2009; Gruhier et al. 2010). All the input and output data are in gridded 0.25° format. The output soil moisture product is expressed as normalized volumetric water content ($m^3\,m^{-3}$) ranging from 0.00 to 1.00. However, LPRM output is known to be biased high in many locations (e.g., Wagner et al. 2007), and its optimization loop is therefore purposefully not capped at expected saturation soil moisture values, as this would adversely affect correlation with actual soil moisture values. Consequently, its soil moisture output is more accurately interpreted as a relative soil moisture index, rather than an absolute value. Since τ is a linear function of vegetation water content (De Jeu and Owe 2003), it should also be useful for monitoring vegetation conditions. Typical values of the τ range between 0 and 1.3 at C band, and values above 0.75 are large enough to decrease the sensitivity of C band to soil moisture variation for practical purposes (Owe et al. 2001).

b. Hydrologic models

The modified two-layer Palmer model developed initially by Palmer (1965) is used to test the efficiency of nonlinear hydrologic models. The two-layer Palmer model is a relatively simple LSM compared to other modern LSMs, but it captures key nonlinear LSM characteristics due to finite water holding capacity of the

soil layers, water movement between two layers, and water loss due to evapotranspiration (ET). Therefore, the Palmer model has been used operationally by USDA-FAS for agricultural drought monitoring. In addition, it is reasonable to take it as a baseline for soil moisture assimilation studies considering the marginal added skills of other advanced LSMs in predicting vegetation conditions (Crow et al. 2012b). For more objective evaluation, the performance of the Palmer model is compared with a linear soil water accounting model [the antecedent precipitation index (API) model] as well as a statistical benchmark model explained in section 3b.

The Palmer model is based on a simple bookkeeping method: precipitation replenishes soil water content within its layers, and water loss from each layer is derived by actual ET. The actual ET depends on potential ET, the initial water content, and available water capacity of the soil layers. The first soil layer of the model is assumed to contain 2.54 cm of available water content at field capacity. The available water capacity of the second layer is determined based on soil texture, depth to bedrock, and soil type from the Food and Agriculture Organization (FAO) Digital Soil Map of the World (Bolten et al. 2010). The second layer is assumed to have a ''no flow'' boundary. The modified Palmer model currently operated by the USDA-FAS has an additional diffusion term for stronger vertical coupling and gradual soil moisture gradients between two layers, which contributes to better assimilation results (Bolten et al. 2010).

The potential ET is calculated from the modified FAO Penman–Monteith equation using observed daily minimum and maximum temperature (Bolten et al. 2010). Therefore, the major meteorological forcing inputs of the Palmer model are daily precipitation and air temperature. Those daily meteorological inputs are derived from the National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS) at 0.25° resolution. The model is run from July 2002 to June 2011 when AMSR-E observations are available. The initial condition is set up after spinning up the model three times for the simulation period. Palmer model output could possibly be enhanced by adding additional processes to the model and/or improved calibration of its existing processes. However, at global scales, such improvements are hard to implement and validate. Instead, our focus here is on evaluating the existing Palmer model version currently in use operationally at USDA-FAS.

While the Palmer model serves as a baseline for nonlinear hydrologic models, a simpler linear API model expressed in Eqs. (1) and (2) is also used for benchmarking purposes:

$$\text{API}_{i,j} = \gamma_{i,j}\text{API}_{i-1,j} + P_{i,j} \qquad (1)$$

$$\gamma_{i,j} = \alpha - \beta(T_{\max,i,j} - 270), \qquad (2)$$

where $P_{i,j}$ is accumulated precipitation on day i and grid j and $T_{\max,i,j}$ is the climatological air temperature (K). The global constants α and β are assigned values of 0.99 (no unit) and $0.0002\,\text{K}^{-1}$, respectively, based on manual calibration to maximize the lagged correlation (-1 month) of API versus NDVI (explained in section 3a).

The loss coefficient γ of the API model is modified as shown in Eq. (2) to reflect varying depletion rates of soil water content with daily maximum temperature $T_{\max,i,j}$. This modification allows the API model to run with the same meteorological forcing (i.e., daily precipitation and maximum air temperature) as the Palmer model. Both the Palmer and the API model are prognostic models reflecting previous soil moisture conditions (memory), while the benchmark statistical model (explained in section 3b) is a fully diagnostic model. Therefore, additional comparison with the API model enables us to evaluate the efficiency of nonlinear model physics in the Palmer model more objectively than comparing only with the benchmark model.

c. EnKF

The EnKF is a well-known sequential data assimilation technique and has been demonstrated as an effective technique for soil moisture assimilation by a number of studies (Reichle et al. 2002; Crow and Wood 2003; Reichle and Koster 2005; Zhou et al. 2006). It is based on a statistical Monte Carlo approach in which forecast error covariance information is sampled from an ensemble of model realizations. In this study, a 30-member ensemble is initially created by directly adding perturbations to a state vector $\boldsymbol{\theta}$ (consisting of both surface and root-zone soil moisture values). Each ensemble member of the state vector $\boldsymbol{\theta}_{k-1}^i$ is forecasted through a nonlinear model operator $f_{k-1}(\cdot)$ at time step $k - 1$:

$$\boldsymbol{\theta}_k^{i-} = f_{k-1}(\boldsymbol{\theta}_{k-1}^{i+}, w_{k-1}^i), \qquad (3)$$

where the plus and minus superscripts refer to the updated and forecasted states, respectively, for each ensemble member i. The error term w represents all uncertainties in the forcing data, model formulation, and/or parameterization. It conforms to a Gaussian distribution with zero mean and has covariance $\mathbf{Q}$:

$$\mathbf{Q} = \begin{bmatrix} Q & \alpha Q\rho \\ \alpha Q\rho & \alpha^2 Q \end{bmatrix}, \qquad (4)$$

where ρ indicates the vertical correlation between perturbations applied to each soil layer and is assumed to be equal to one in this study. The scalar term α reflects the ratio of standard deviations of the root-zone soil moisture perturbations to the one for surface soil moisture perturbations. Here, this ratio is assumed equal to the ratio between the available water holding capacity of the surface layer and the root-zone layer. We also assume the uncertainty in the model forecast is dominated primarily by the accuracy of the main forcing variable, precipitation data. Therefore, Q in Eq. (4) is expressed in Eq. (5) as a function of the average distance (D) to the three closest World Meteorological Organization (WMO) rain gauges used to correct satellite-based precipitation data:

$$Q = \begin{cases} 0.02^2 \text{ m}^6 \text{ m}^{-6} & \text{if} \quad D < 100 \text{ km} \\ 0.04^2 \text{ m}^6 \text{ m}^{-6} & \text{if} \quad 100 \text{ km} \le D < 150 \text{ km} \\ 0.06^2 \text{ m}^6 \text{ m}^{-6} & \text{if} \quad 150 \text{ km} \le D < 200 \text{ km} \\ 0.08^2 \text{ m}^6 \text{ m}^{-6} & \text{if} \quad D \ge 200 \text{ km} \end{cases}.$$

(5)

The EnKF updates model-predicted soil moisture based on relative uncertainties in observations and model predictions. Since AMSR-E provides daily soil moisture observations, the model predictions are updated on a daily basis. Before assimilation, the LPRM soil moisture retrievals are rescaled through a linear transformation to match the temporal mean and standard deviation of model-predicted surface soil moisture for the 9-yr simulation period in order to put them in the same climatology with the model. The predicted state variable (θ_k^{i-1}) is then updated by optimally integrating observations ($\theta_{\text{LPRM},k}$) and model forecasts via

$$\theta_k^{i+} = \theta_k^{i-} + \mathbf{K}_k[\theta_{\text{LPRM},k} - \mathbf{H}\theta_k^{i-} + v_k^i], \qquad (6)$$

where the observation operator $\mathbf{H} = [1\ 0]$ and the observation error term v_k^i is a Gaussian noise with mean zero and variance R. The Kalman gain ($\mathbf{K}_k$) is determined by the cross correlation between the forecasted observations and the forecasted state variables ($\mathbf{P}_k^- \mathbf{H}_k^{\text{T}}$) and covariance matrix of the forecasted observations ($\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\text{T}}$):

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^{\text{T}} [\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^{\text{T}} + R]^{-1}. \qquad (7)$$

In this study, the uncertainties of the surface soil moisture retrievals from LPRM are specified based on the land cover type because of the critical effect of vegetation density on above-canopy brightness temperature measurements (Owe et al. 2001). Here, we follow typical

TABLE 1. Assumed observation errors ($R^{1/2}$) for different land cover.

Land cover classification	Std dev of obs error (m^3 m^{-3})
Forests or woodlands (ENF, EBF, DNF, DBF)	0.06
Wooded grasslands/shrubs (WGS) or closed bushlands or shrublands (CBS)	0.05
Open shrublands (OS) or grasses (GRS)	0.04
Croplands (CRP)	0.03
Bare (BAR)	0.02
Water	0.99

practice by assigning time constant values of R as a function of land cover type. Land cover information is obtained from the 8-km Moderate Resolution Imaging Spectroradiometer (MODIS) land cover classification dataset produced by the University of Maryland (http://glcf.umd.edu/). Standard deviations of observation errors, $R^{1/2}$, for each different land cover type are shown in Table 1. It should be noted that more complex approaches have been proposed to define Q and R in data assimilation systems (Crow and Van den Berg 2010; Dorigo et al. 2010; Parinussa et al. 2011). However, the impact of these new approaches on large-scale data assimilation results has not yet been verified (Draper et al. 2013), nor have they been widely applied yet in operational systems. As a result, we choose to apply the simpler approaches described above as a baseline for our benchmarking procedure.

From a benchmarking point of view, the EnKF takes inputs from the outputs of the hydrologic model (surface and root-zone soil moisture) and the observation algorithm (surface soil moisture) and produces new (updated) root-zone soil moisture estimates (Fig. 1). Because the weighting underlying these new estimates (supposedly) reflects our understanding of time/space variations in model and observation errors, EnKF estimates should outperform competing approaches based on arbitrary averaging. Therefore, the efficiency of the EnKF can be evaluated by quantifying the added skill of the updated soil moisture compared to a benchmarking model consisting of a simple (spatially fixed) linear combination of background model predictions and surface soil moisture observations.

d. Evaluation data

The performance of each process is evaluated using the lagged rank correlation with vegetation condition reflected in NDVI. Monthly NDVI data are derived from MODIS MOD13C2 products and are aggregated to 0.25° resolution from its initial 0.05° resolution.
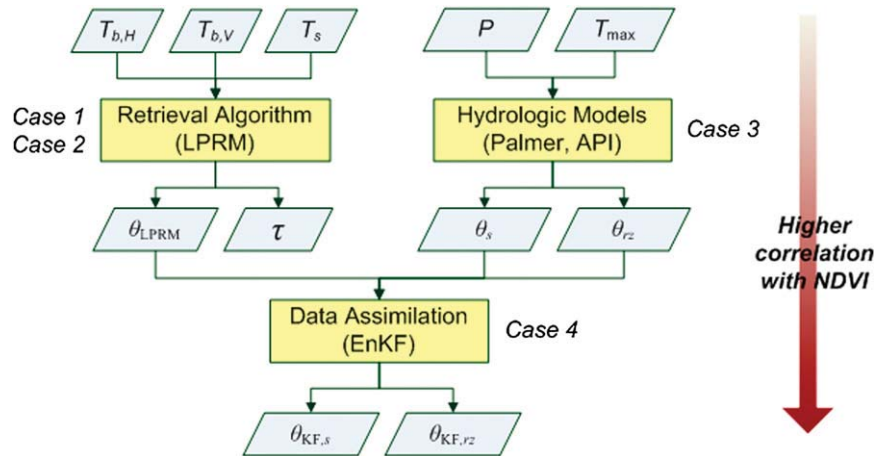
FIG. 1. Schematic of benchmarking framework (θ_{LPRM}: surface soil moisture observation, τ: vegetation optical depth, θ_s: model predicted surface soil moisture, θ_{rz}: model predicted root-zone soil moisture, $\theta_{\mathrm{KF},s}$: updated surface soil moisture, $\theta_{\mathrm{KF,rz}}$: updated root-zone soil moisture).

MODIS products flagged as "fully reliable" are extracted, and pixels in which more than 50% of the area consists of barren, tundra, forest cover, and open water surface are masked out to focus our study on areas with nonnegligible agricultural and grazing land uses.

3. Experiment setup

This experiment aims to benchmark individual components of a soil moisture data assimilation system for quasi-global (60°S–60°N) agricultural drought monitoring. Analysis is based on monthly data from July 2002 to June 2011. All daily observed or modeled data are averaged to monthly values before the analysis. However, months that have less than five daily data points are masked.

Unlike meteorological inputs for hydrologic models (precipitation and air temperature), remotely sensed brightness temperature and its processed products from LPRM have varying coverage with time because of several limiting factors such as dense vegetation, frozen surface condition, and active precipitation. For consistency, only pixels that contain retrievals for all five remote sensing products ($T_{b,V}$, $T_{b,H}$, T_s, θ_{LPRM}, and τ) are included in the analysis. Additionally, barren areas where no temporal NDVI variability is observed and (presumably energy limited) areas where the two-layer Palmer model demonstrated a nonsignificant rank correlation with NDVI anomalies (at 80% significance) are masked out in the present analysis (Crow et al. 2012b). Note that one potential reason for this loss in sensitivity is the saturation of NDVI in densely-vegetated regions. While the strict filtering of the data reduces the number of available data for the analysis, especially in the Northern Hemisphere, it strengthens the evaluation technique

(rank cross correlation with NDVI) by focusing on areas and seasons in which soil moisture and NDVI are viable drought indicators.

a. Evaluation metric

As an agricultural drought indicator, root-zone soil moisture estimates can be evaluated by measuring how much their anomalies are correlated with subsequent NDVI anomalies, an indicator of vegetation condition (Peled et al. 2010; Bolten and Crow 2012). Likewise, we use this criteria (rank correlation with NDVI) to determine the agricultural drought forecasting capability of a certain dataset (i.e., a soil moisture proxy predicted by a physical model or benchmarks). It is natural to expect that some rank correlation of a particular soil moisture dataset will be enhanced as the primary observations or forcing variables (e.g., precipitation) go through the nonlinear processes sequentially in the retrieval algorithm, model physics, and/or data assimilation step (Fig. 1). Magnitudes of the correlations allow us to compare the performance of each complex process relative to their simpler linear benchmarks.

The correlation analysis of this study is based on ranks because the rank time series are free from seasonality, only showing relative wetness of a particular month relative to the same month of the year in all other years. The analysis starts from ranking monthly anomalies of a certain dataset of soil moisture proxy grouped by month of year for the analysis period (July 2002 to June 2011). The resulting ranks—or Rank($\overline{\theta}_k$) of the dataset for month k—are normalized so that they are in the ranges between 0.0 and 1.0. Therefore, a month that has a rank of zero (one) has the driest (wettest) soil moisture condition compared to the same month of the year in other years. Monthly NDVI data are also ranked in the same way. Figure 2 shows an example
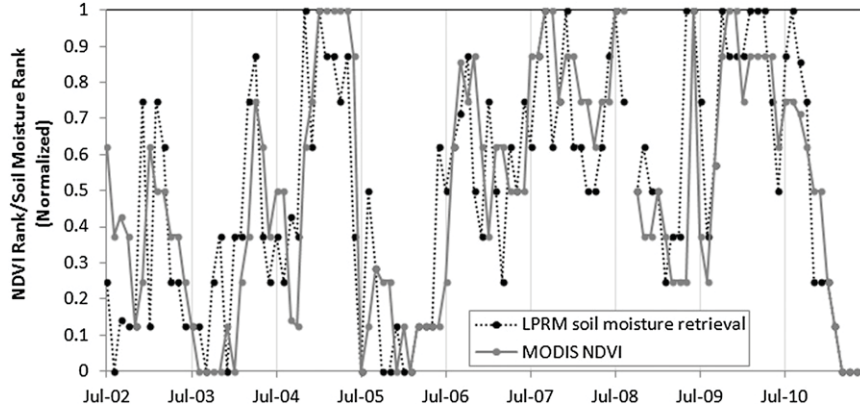
FIG. 2. Example time series of monthly ranks of LPRM soil moisture estimates and NDVI from a 0.25° grid centered at 30.125°N, 109.375°W.

time series of Rank($\overline{\theta}_k$) and Rank($\overline{\text{NDVI}}_k$) from a grid in North America.

The lag-L rank correlation coefficient $R(L)$ is calculated as the Pearson's correlation between Rank($\overline{\text{NDVI}}_k$) and Rank($\overline{\theta}_{k+L}$). In testing the skills of a soil moisture dataset as a leading indicator to NDVI, rank correlation at $L = -1$ (i.e., soil moisture precedes NDVI by 1 month) $R(-1)$ is our primary focus. It should be noted that optimal lags to reflect the best NDVI soil moisture correlation may be different for different land cover types (Musyimi 2011). However, a lag time of 1 month is consistent with the monthly operational cycle of many agricultural drought monitoring activities. In addition to the spatial masking described above, months having maximum air temperature below 5°C are excluded to minimize the impact of cold-season conditions. That is, snow dominated regions or seasons are not the focus of the present study. Finally, for each pixel, the correlation is calculated only if there are at least 30 pairs of Rank($\overline{\text{NDVI}}_k$) and Rank($\overline{\theta}_{k+L}$) to obtain statistically significant results.

To test significance of the sample correlation coefficient, the sample variance (σ^2) of a rank correlation $R(L)$ for a single pixel is estimated through a Fisher transformation (Von Storch and Zwiers 2002) as follows:

$$F[R(L)] = \frac{1}{2} \ln\{[1 + R(L)]/[1 - R(L)]\}. \quad (8)$$

This transformation converts the sample correlation into a normal distribution with variance $\sigma_F^2 = 1/(N_t - 3)$ (Fieller et al. 1957). The number of the temporal degrees of freedom (N_t) in the time series is computed by considering lag-1 temporal autocorrelation of soil moisture estimates ($\rho_{t,\theta}$) and NDVI ($\rho_{t,\text{NDVI}}$) (Dawdy and Matalas 1964):

$$N_t = n_t \times [(1 - \rho_{t,\theta}\rho_{t,\text{NDVI}})/(1 + \rho_{t,\theta}\rho_{t,\text{NDVI}})], \quad (9)$$

where n_t is number of monthly data in the time series.

The sample variance of a rank correlation $R(L)$ in Fisher space (σ_F^2) can be converted into a regular space as follows:

$$\sigma^2 = \sigma_F^2(\text{sech}^2\{F[R(L)]\})^2. \quad (10)$$

Likewise, the sample variance of spatially averaged $R(L)$ in space can be estimated by dividing the spatial average of Eq. (10) for each pixel by the number of effective spatial degrees of freedom (N_s) that are averaged across

$$N_s = n_s \times [(1 - \rho_s)/(1 + \rho_s)]^2, \quad (11)$$

where n_s is the total number of pixels averaged over and ρ_s is the lag-1 spatial autocorrelation of the field. These sample variance values are used to construct error bars (± 1 standard deviation) in Figs. 3, 4, and 5. The statistical significance of an $R(-1)$ difference for a single pixel between soil moisture outputs from each data assimilation component and their benchmarks can be expressed via the Z score:

$$Z = \{F[R_\theta(-1)] - F[R_{\text{BM}}(-1)]\}/\sqrt{\sigma_{F,\theta}^2 + \sigma_{F,\text{BM}}^2}. \quad (12)$$

b. Benchmarking models

Our benchmarking approach is based on a linear multivariate regression models. For each component of the data assimilation system (i.e., observation, modeling and assimilation), an empirical relationship between inputs and outputs is established. A general expression of the benchmarking model with p independent variables is
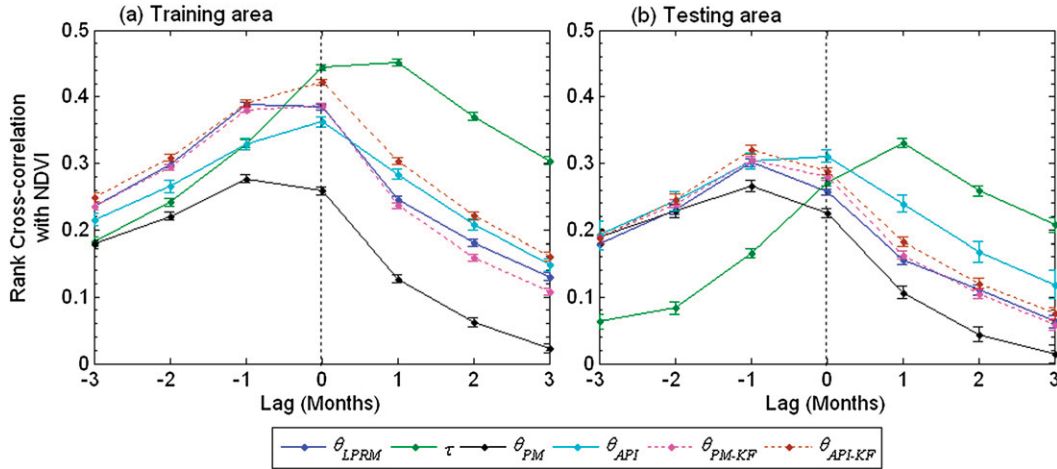
FIG. 3. Spatial average of $R(L)$ for the LPRM retrievals (θ_{LPRM} and τ), model predictions (θ_{PM} and θ_{API}), and updated model outputs ($\theta_{\text{PM-KF}}$ and $\theta_{\text{API-KF}}$) within the training (60°S–30°N) and testing (30°–60°N) areas.

$$\text{BM}(X_1, \ldots, X_p)_{i,j} = a_1 X'_{1,i,j} + a_2 X'_{2,i,j} + \cdots + a_p X'_{p,i,j} + \varepsilon_{i,j}, \quad (13)$$

where $X'_{p,i,j}$ is the anomaly of each input variable p for month i and pixel j normalized by its temporal standard deviation $[(X_{p,i,j} - \overline{X}_{p,j})/\sigma_{p,j}]$. The parameters $(a_1, \ldots, a_p)$ are constant in time and space and imply the relative contribution (weight) of each independent input variable in predicting the integrated soil moisture proxy $\text{BM}(X_1, \ldots, X_p)_{i,j}$. The time series of the resulting soil moisture proxies are ranked and used to compute $R(L)$ with NDVI.

In setting up the relationship, the objective function is the spatial average of $R(-1)$ or correlation between the statistical model predictions and NDVI at $L = -1$. That is, our goal in training the parameters is to achieve the highest spatial average of the lagged rank correlation with *next month's* NDVI. We reserve the extratropical Northern Hemisphere (ETNH, 30°–60°N) to test the empirical models and use the remaining area of the globe (60°S–30°N) to train the parameters. This spatial segregation allows us to have exclusively independent datasets for training and testing purposes. Note that our goal is only to obtain a *single* spatially and temporally constant set of parameters $(a_1, \ldots, a_p)$ for the entire globe. We take this very conservative approach to find a minimum reference level (benchmark) even though much better benchmark model performance is possible if tuned parameters in Eq. (13) are allowed to vary in time and space.

Unlike the nonlinear models such as LPRM or the Palmer model, this regression model combines available inputs $(X'_{1,i} \sim X'_{p,i})$ in a linear way and produces a soil moisture proxy for each month without any knowledge of the physical relationships between the input and output variables. In addition, this statistical model is a fully diagnostic model and therefore ignorant of information concerning the prior status of the variables (like LPRM, but unlike the Palmer model).

For the individual components of the data assimilation system, we define a series of benchmarking models as follows.

Case 1: The purpose of Case 1 is to test the efficiency of LPRM soil moisture retrievals versus a benchmark model consisting of a linear combination of inputs into LPRM (Fig. 1). For Case 1, our focus is on comparing the 1-month lagged rank correlation of θ_{LPRM} versus NDVI—denoted as $R(-1)$—versus the analogous rank correlation of a benchmarking model based solely on a simple linear combination of $T_{b,H}$, $T_{b,V}$, and T_s or

$$\text{BM}(T_s, T_{b,V}, T_{b,H})_i = a_{1,1} T'_{s,i} + a_{1,2} T'_{b,H,i} + a_{1,3} T'_{b,V,i} + \varepsilon_{1,i}. \quad (14)$$

Case 2: A defining feature of LPRM is that it produces τ as an intermediate product. Since τ reflects the presence of vegetation water content, it contains drought-relevant information. Based on this hypothesis, another way of testing the efficiency of the LPRM is to compare the information of both LPRM outputs (θ_{LPRM} and τ, after combining them into a single variable) against a benchmark model derived from LPRM inputs. Therefore, to represent complete skills from the two outputs, we use the same multiple linear regression model in Eq. (15) and compare it with the regression model from the inputs in Eq. (16):

$$\text{LR}(\theta_{\text{LPRM}}, \tau)_i = a_{2,1} \theta'_{\text{LPRM},i} + a_{2,2} \tau'_i + \varepsilon_{2,i} \quad (15)$$
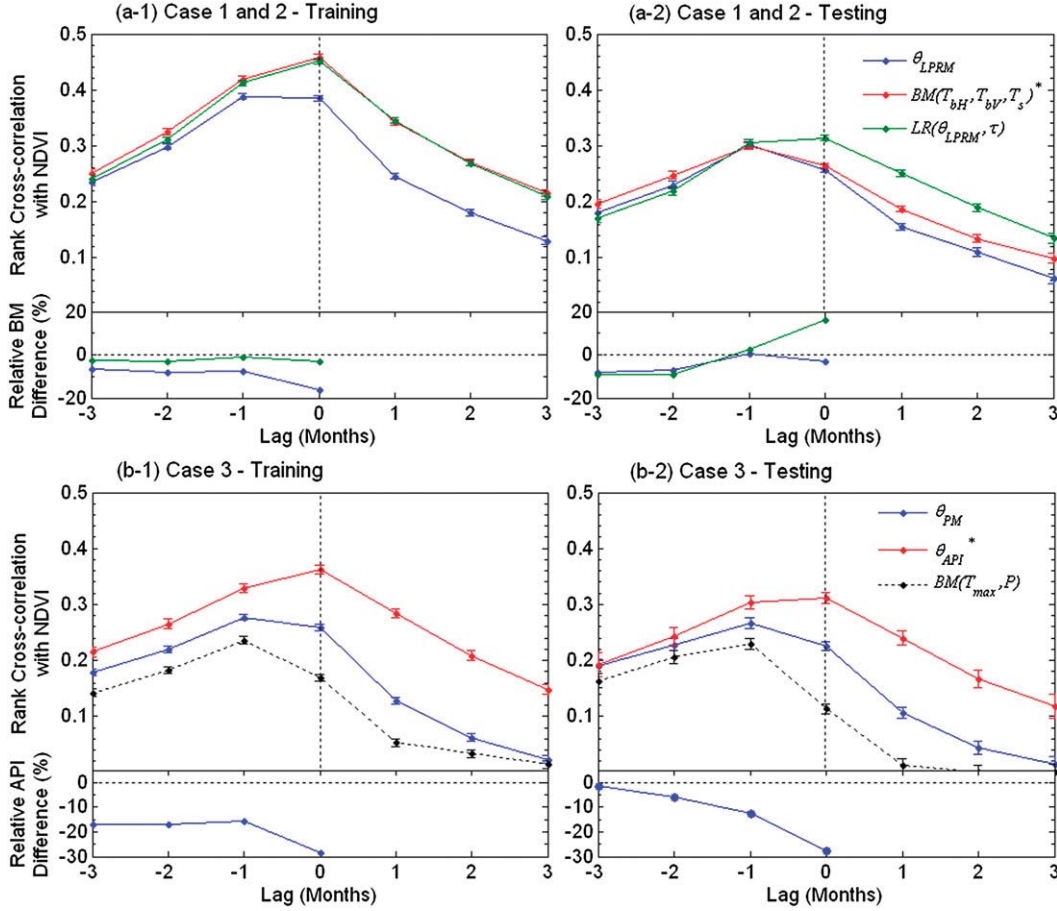
FIG. 4. Spatial average of $R(L)$ for the training (60°S–30°N) and testing (30°–60°N) areas for Cases 1–3. Subplots show relative $R(L)$ difference (%) of retrieval (or model) outputs to benchmarks indicated by the superscript (*) in the legend {e.g., $100*[R(L)\theta_{LPRM} - R(L)\theta_{BM}]/R(L)_{BM}$}. Note for $LR(\theta_{LPRM}, \tau)$ in (a) relative $R(L)$ differences are computed using the $BM(T_s, MPDI, T_{b,V})$ defined in Eq. (16).

$$BM(T_s, \text{MPDI}, T_{b,V})_i = a_{22,1} T'_{s,i} + a_{22,2} \text{MPDI}'_i$$
$$+ a_{22,3} T'_{b,V,i} + \varepsilon_{22,i}. \quad (16)$$

By comparing (15) and (16), we can get a sense of the value of θ_{LPRM} and τ outputs—combined via Eq. (15)—versus the simple combination of LPRM inputs in Eq. (16). In this case, the benchmark model in Eq. (16) uses MPDI as an input variable because τ is a function of the MPDI. The MPDI removes the effect of soil temperature from the microwave emission signal (Owe et al. 2001). In addition, $T_{b,V}$ is used rather than $T_{b,H}$ because a preliminary analysis (not shown) demonstrated that $T_{b,V}$ has a higher correlation with NDVI than $T_{b,H}$. It is also interesting to examine the marginal value of τ retrievals, above and beyond θ_{LPRM}, for agricultural drought monitoring. We can achieve this objective by comparing $R(-1)$ of the predictions

from Eq. (15) with the $R(-1)$ for only θ_{LPRM}.

Case 3: The performance of a nonlinear hydrologic model is evaluated by comparing it with a benchmark model based on the linear combination of its inputs [i.e., daily maximum air temperature (T_{max}) and daily accumulated precipitation (P)]:

$$BM(T_{max}, P)_i = a_{3,1} T'_{max,i} + a_{3,2} P'_i + \varepsilon_{3,i}. \quad (17)$$

Therefore, the key comparison in Case 3 is between $R(-1)$ for Palmer and/or API model predictions of root-zone soil moisture versus $R(-1)$ for the soil moisture proxy obtained via this benchmark model. Note that the prognostic structure of the API and Palmer model should provide a natural advantage over the purely diagnostic form of Eq. (17).

Case 4: The EnKF is designed to optimally combine model predictions (θ_{PM} or θ_{API}) and observations (θ_{LPRM}) using information about their relative

uncertainties. As an alternative to the standard EnKF, one can construct a benchmarking model that does not have any prior knowledge of those uncertainties and instead applies (globally constant) weights to merge model-predicted soil moisture (θ_{PM} or θ_{API}) with remotely sensed soil moisture retrievals (θ_{LPRM}):

$$\mathrm{BM}(\theta_{PM}, \theta_{LPRM})_i = a_{4,1}\theta'_{PM,i} + a_{4,2}\theta'_{LPRM,i} + \varepsilon_{4,i} \tag{18}$$

$$\mathrm{BM}(\theta_{API}, \theta_{LPRM})_i = a_{42,1}\theta'_{API,i} + a_{42,2}\theta'_{LPRM,i} + \varepsilon_{42,i}. \tag{19}$$

Note that the EnKF uses different uncertainties for each pixel (i.e., each pixel has its own model and observation error depending on land cover and distance from WMO rain gauges). In contrast, the benchmarking model applies spatially and temporally constant weights for the Palmer model predictions (θ_{PM}) and observations from LPRM (θ_{LPRM}) obtained from a training dataset. Comparing $R(-1)$ for soil moisture proxy obtained via Eqs. (18) or (19) with $R(-1)$ for a soil moisture analysis (acquired using a Palmer or API model-based EnKF) will evaluate how efficiently the EnKF is implemented.

4. Results

Figure 3 shows spatially averaged lagged $R(L)$ results for each component of an offline agricultural drought monitoring system (LPRM soil moisture and vegetation optical depth retrievals, open loop model predictions, and an EnKF analysis) for the training and testing dataset separately during the 9-yr period from July 2002 to June 2011. Although LPRM soil moisture retrievals (θ_{LPRM} in Fig. 3) reflect only surface soil moisture condition, it demonstrates nearly as large a $R(-1)$ as root-zone soil moisture estimates obtained from the EnKF analysis (θ_{PM-KF} or θ_{API-KF}). In contrast, the Palmer model–predicted root-zone soil moisture (θ_{PM}) has relatively low $R(-1)$ compared to the other soil moisture estimates. Most notably, it fails to match values obtained with the simpler API model. Error bars constructed using Eqs. (8)–(11) are added to each point in Fig. 3 to reflect 1σ sampling uncertainty. Relatively larger error bars for API results are due mainly to its higher temporal autocorrelation (and thus reduce temporal degrees of freedom) relative to other products.

For each case introduced above, Figs. 4 and 5 compare the $R(L)$ of each soil moisture product versus its appropriate benchmark in both the training and testing datasets. As shown in Crow et al. (2012b), semiarid areas have very strong correlation [$R(-1) > 0.5$] between soil moisture and NDVI due to water-limited plant growth. Therefore, the large fraction of semiarid areas found in the Southern Hemisphere training dataset ensures that the average $R(-1)$ is higher in the training set than in the Northern Hemisphere testing set. An in-depth explanation of results in Figs. 3, 4, and 5, focusing mainly on the testing set, is given below.

Spatially different coupling strengths between soil moisture and NDVI due to different climate conditions can be found on the maps of $R(-1)$ for the testing area (ETNH, 30°N – 60°N) in Fig. 6. In relation to climatic region, there are distinct differences in the magnitude of $R(-1)$ for different land cover types, as shown in Fig. 7. Land cover classification and their areal fractions are provided in Table 2 for both the training and testing areas. Error bars in Fig. 7 were computed using Eqs. (8)–(11). Note that two land cover types [evergreen broadleaf forests (EBF) and deciduous needleleaf forests (DNF)] are not displayed in Fig. 7 because of their small portions (Table 2) and large sample variance. Most forested areas show weak relationships between soil moisture and NDVI [$R(-1) < 0.2$], which makes sense because trees take water from deeper soil and are more resilient to surface soil moisture shortage than shrub or grass. Shrub and grasslands that correspond to the semiarid climate have strong correlations with NDVI [$R(-1) > 0.4$]. Croplands have relatively lower $R(-1)$ ($\sim$0.3) than the shrub or grasslands, likely because artificial human interference for agricultural practices such as irrigation and the installation of tile drainage systems may disturb the direct coupling between soil moisture and NDVI in managed agricultural landscapes.

Below, we define the efficiency of a given model monitor component (e.g., the soil moisture retrievals algorithm, the LSM, and the data assimilation system) as the difference between the $R(-1)$ of soil moisture output provided by each component and that is obtainable using its corresponding linear benchmark. This efficiency provides a means to evaluate the added value of nonlinear and/or complex processes embedded in each component relative to a baseline established by the simple benchmarks in Eqs. (14)–(19). Note that while each benchmark model is based on fitted parameters, these parameters are static in time and space and are fitted to a spatially distinct training dataset.

a. Case 1: Efficiency of the LPRM retrieval algorithm

For Case 1, comparisons between the benchmark model in Eq. (14), based on a linear combination of primary AMSR-E observations ($T_{b,V}$, $T_{b,H}$, and T_s), and
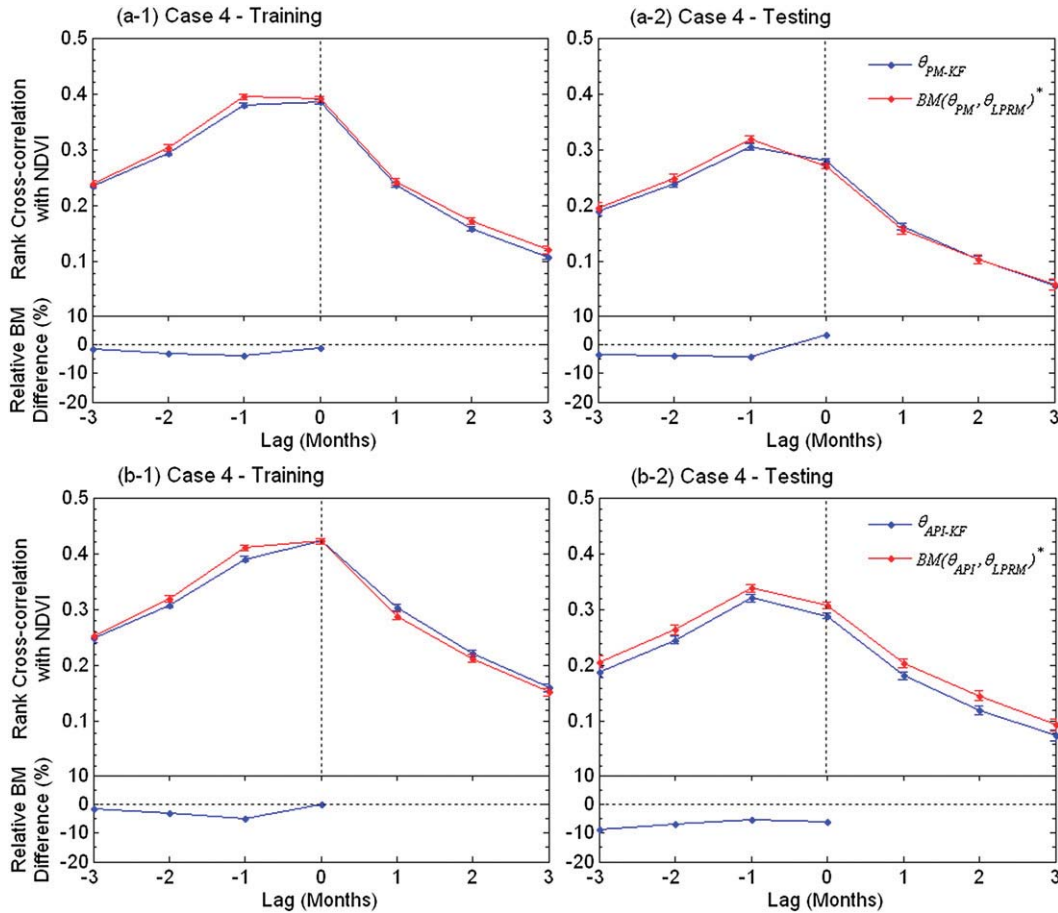
FIG. 5. Spatial average of $R(L)$ for the training (60°S–30°N) and testing (30°–60°N) areas for Case 4. Subplots show relative $R(L)$ difference (%) of updated outputs to benchmarks indicated by the superscript ("*") in the legend (e.g., $100*[R(L)\theta_{\text{PM-KF}} - R(L)\theta_{\text{BM}}]/R(L)\theta_{\text{BM}}$).

the full LPRM soil moisture product (θ_{LPRM}) are decidedly mixed. Averaged over the testing domain, the benchmark model produces roughly the same $R(-1)$ as the full LPRM soil moisture product (Table 3). For all lags other than $L = -1$, the benchmark model slightly outperforms the LPRM soil moisture product in predicting vegetation conditions (Fig. 4a-2). However, when focusing just on $R(-1)$ results for individual land cover types within the testing area, LPRM outperforms the benchmark model for most land cover types other than open shrublands (OS), closed bushland

or shrublands (CBS), deciduous forests (DBF) and bare soil in Fig. 7, and the Z-score map in Fig. 8a illustrates that LPRM is marginally superior to the benchmark model over broad areas of Europe and the eastern United States.

The evaluation approach in the present study adopted a benchmark model as a minimum reference level to evaluate the performance of the LPRM retrieval algorithm. Even though LPRM soil moisture has been shown to correlate well with in situ observations of soil moisture (Draper et al. 2009; Brocca et al. 2011), results
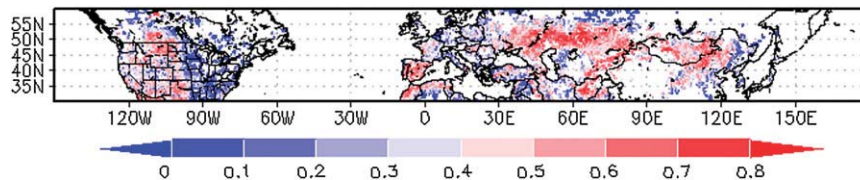

FIG. 6. Map of $R(-1)$ for LPRM soil moisture estimates within the testing area.
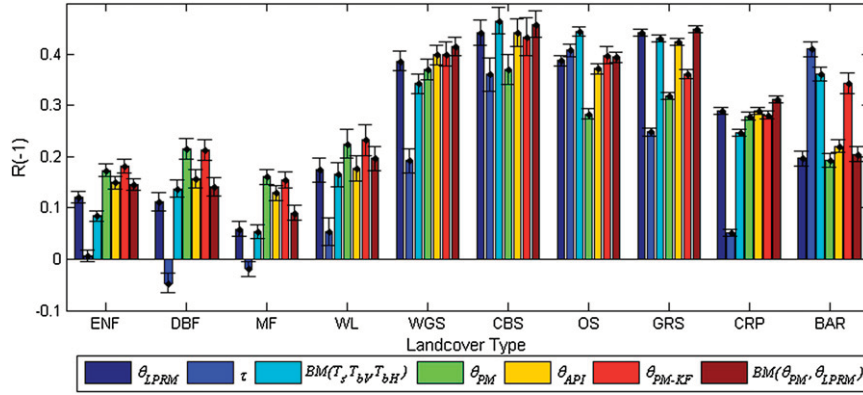
Fig. 7. Spatial average of $R(-1)$ of various soil moisture products within different land cover types in the ETNH (30°–60°N). See Table 2 for definitions of land cover acronyms.

here imply that LPRM benefits marginally from its nonlinear parameterization when judged against our NDVI validation metric. As a result, it may be possible to simplify (and/or linearize) the LPRM algorithm without adversely affecting its value for drought forecasting. Nevertheless, it should be stressed that in this study we benchmark the "quality" of soil moisture information only for a very specific application—forecasting the impact of agricultural drought on vegetation health—and these results might be application specific.

b. Case 2: Efficiency of the LPRM retrieval algorithm and added skill of τ

Case 2 is designed to investigate if LPRM τ estimates can add robust skill to agricultural drought monitoring. First, it should be noted the difference in measurements of NDVI and τ. In principle, NDVI observation is strongly influenced by the chlorophyll concentration in

vegetation and is thus related to green leaf biomass (Owe et al. 2001; Jones et al. 2011). However, τ is obtained based on the vegetation dielectric properties that are strongly related to vegetation water content in both foliage and woody biomass (Liu et al. 2011; Andela et al. 2013). Therefore, the NDVI/τ relationship is not known explicitly, even though both of them reflect some aspects of vegetation condition and NDVI was used to validate the τ in Owe et al. (2001).

From Fig. 3, the τ rank anomaly is significantly less successful (compared to surface soil moisture observations) as a leading indicator for future NDVI anomalies, especially in testing areas. Interestingly, however, the τ has strongest correlation with NDVI at $L = 1$ [i.e., Rank($\overline{\tau}_k$) temporally lags behind Rank($\overline{NDVI}_k$) by 1 month]. Therefore, τ information is more suitable for a retrospective analysis of NDVI rather than forecasting. Nonetheless, the τ shows relatively high $R(-1)$ with NDVI for several land cover types. For example,

TABLE 2. Land cover classification in the training and testing datasets. Pixels where all five remote sensing retrievals ($T_{b,V}$, $T_{b,H}$, T_s, θ_{LPRM}, and τ) are available and monthly maximum air temperature is above 5°C are included in the present analysis. Thus, many forested or frozen areas are excluded.

Land cover type	Training area (60°S–30°N) No. of pixels		Testing area (30°–60°N) No. of pixels	
Evergreen needleleaf forests (ENF)	6155	0.75%	70 485	10.88%
Evergreen broadleaf forests (EBF)	51 983	6.31%	112	0.02%
Deciduous needleleaf forests (DNF)	1	0.00%	2433	0.38%
Deciduous broadleaf forests (DBF)	17 695	2.15%	27 926	4.31%
Mixed forests (MF)	2640	0.32%	42 916	6.63%
Woodlands (WL)	154 142	18.72%	13 933	2.15%
Wooded grasslands/shrubs (WGS)	161 786	19.65%	14 795	2.28%
Closed bushlands or shrublands (CBS)	88 576	10.76%	8328	1.29%
Open shrublands (OS)	136 474	16.58%	76 824	11.86%
Grasses (GRS)	51 441	6.25%	192 866	29.77%
Croplands (CRP)	77 330	9.39%	160 329	24.75%
Bare (BAR)	74 993	9.11%	36 834	5.69%

TABLE 3. Optimized benchmark models and their performance. Key comparisons for each case are in bold.

Case	Model/algorithms/regression equation	Optimized parameters			$R(-1)$	
		$a_{*,1}$	$a_{*,2}$	$a_{*,3}$	Training	Testing
1	θ_{LPRM}	—	—	—	0.388	**0.301**
	$a_{1,1}T'_{s,i} + a_{1,2}T'_{b,H,i} + a_{1,3}T'_{b,V,i}$	0.5	0.8	−1.7	0.420	**0.299**
2	τ	—	—	—	0.330	0.166
	$a_{2,1}\theta'_{LPRM,i} + a_{2,2}\tau'_i$	0.25	0.12	—	0.412	**0.307**
	$a_{22,1}T'_{s,i} + a_{22,2}MPDI'_i + a_{22,3}T'_{b,V,i}$	0.5	−0.6	−1.3	0.417	**0.298**
3	θ_{PM}	—	—	—	0.277	**0.266**
	θ_{API}	—	—	—	0.329	**0.304**
	$a_{3,1}T'_{max,i} + a_{3,2}P'_i$	−0.19	0.52	—	0.241	0.223
4	θ_{PM-KF}	—	—	—	0.380	**0.305**
	θ_{API-KF}	—	—	—	0.390	0.320
	$a_{4,1}\theta'_{PM,i} + a_{4,2}\theta'_{LPRM,i}$	0.28	0.89	—	0.395	**0.319**
	$a_{42,1}\theta'_{API,i} + a_{42,2}\theta'_{LPRM,i}$	0.32	0.57	—	0.411	0.339

land cover types such as CBS, OS, and grasses (GRS) have $R(-1)$ greater than 0.25, but for forested areas and even crop land, τ shows little relationship with next month's NDVI (Fig. 7).

Because of low $R(L)$ for negative L, it is expected that the τ does not add significant value to the prediction of future vegetation conditions. Therefore, when θ_{LPRM} and τ are linearly combined into $LR(\theta_{LPRM}, \tau)$ through Eq. (15), it produces only marginally higher $R(-1)$ than the soil moisture retrievals only [0.307 for $LR(\theta_{LPRM}, \tau)$ and 0.301 for θ_{LPRM} only] for testing dataset as shown in Fig. 4a-2. However, because of the lagged nature of the τ response versus NDVI, the addition of τ does seem to increase $R(L)$ for $L \geq 0$.

In the context of evaluating the efficiency of nonlinear LPRM retrieval algorithm, the linear combination of the two LPRM outputs (θ_{LPRM} and τ) expressed in Eq. (15) is compared to the linear regression model of the three LPRM inputs (T_s, MPDI, and $T_{b,V}$) in Eq. (16). The combination of θ_{LPRM} and τ barely outperforms the benchmark model in Eq. (16) for the testing dataset (0.307 versus 0.298 in Table 3). Since τ is a function of MPDI, MPDI rather than $T_{b,H}$ is used in Case 2 for creating the benchmarking model. However, $R(L)$ of the MPDI does not correspond well to the $R(L)$ of the τ (not shown), and the use of MPDI as a predictor of the benchmarking model does not make much difference compared to the performance of the benchmark model
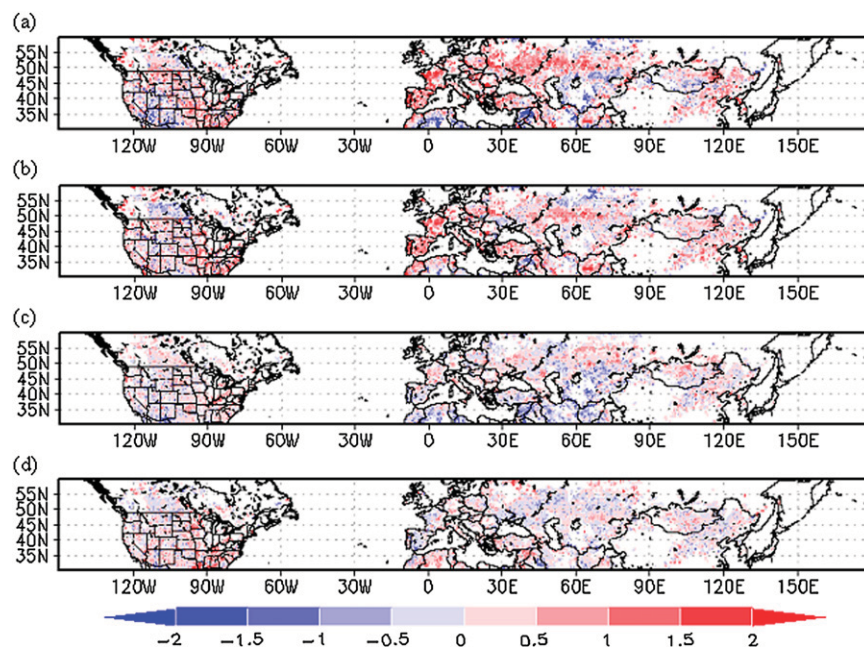

FIG. 8. Maps of Z score for (a) Case 1 $[R(-1)_{\theta_{LPRM}} - R(-1)_{\theta_{BM1}}]$, (b) Case 2 $[R(-1)_{LR(\theta_{LPRM}, \tau)} - R(-1)_{\theta_{BM2}}]$, (c) Case 3 $[R(-1)_{\theta_{PM}} - R(-1)_{\theta_{API}}]$, and (d) Case 4 $[R(-1)_{\theta_{PM-KF}} - R(-1)_{\theta_{BM4}}]$.

in Case 1 [Eq. (14)]. A plot of spatially averaged $R(L)$ for the benchmark in Eq. (16) is very similar to that of the Case 1 benchmark in Eq. (14). Therefore, $R(L)$ of the benchmark of Case 2 are not shown in Fig. 4a, but the relative difference of $LR(\theta_{LPRM}, \tau)$ was computed based on its benchmark in Eq. (16).

When spatially averaged, τ does not seem to add significant skill to NDVI prediction. However, for certain land cover types such as CBS and OS, the combination of θ_{LPRM} and τ contributes to an increase in $R(-1)$ from the soil moisture–only case because of the high $R(-1)$ of τ (Fig. 7). This result might be explained by the difference in what τ and NDVI actually measure; NDVI is sensitive to the chlorophyll concentration in the canopy while τ is sensitive to the water content both in foliage and woody biomass. Even though total above-ground biomass represented in τ decreases with a lack of precipitation, the NDVI may show lagged response because green canopy cover is maintained in grassland or shrubland land covers during drought (Liu et al. 2011). Therefore, the τ information may be more useful for open and closed shrubland vegetation types. This is also confirmed by the Z-score map in Fig. 8b. When θ_{LPRM} and τ are combined into $LR(\theta_{LPRM}, \tau)$ via Eq. (15), certain relatively arid regions (e.g., the western United States and northern Africa) are converted from negative to positive Z scores, indicating an improvement in performance relative to the benchmark (cf. Figs. 8a,b).

c. Case 3: Efficiency of hydrologic models

Case 3 examines the efficiency of a nonlinear hydrologic model by benchmarking Palmer model output against a linear regression model—given in Eq. (17)—with independent variables matching the meteorological inputs of the Palmer model (precipitation and daily maximum air temperature). If more complicated LSMs are evaluated, additional model inputs can be added to Eq. (17). Since soil moisture retains memory from the previous model time step, any kind of prognostic model should possess a natural advantage over a fully diagnostic model like the benchmark model in Eq. (17). Therefore, for more objective evaluation, performance of the Palmer model is also compared with a purely linear prognostic model (API model) as well as the benchmark in Eq. (17). The Palmer model produces soil moisture condition for each of the two layers. To mimic an integrated root-zone estimate, we use averages of the two soil moisture outputs weighted by the water holding capacity of each layer.

The Palmer model performs relatively poorly in predicting future vegetation dynamics as shown in Fig. 3. The values of $R(-1)$ of the Palmer model are 0.277 and 0.266 for training and testing dataset, respectively
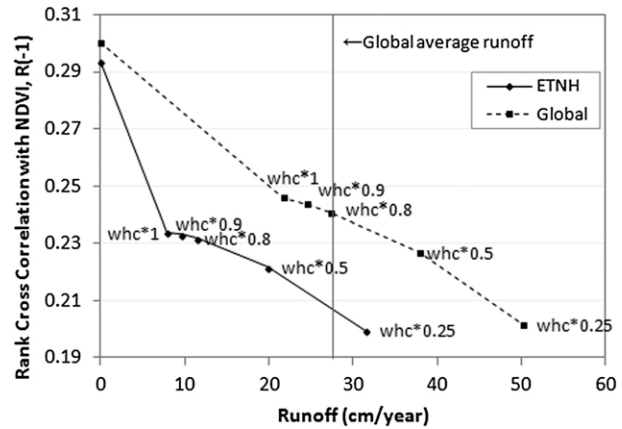

FIG. 9. Relationship between correlation with NDVI and runoff generation using the API model with different settings of water holding capacity. The original water holding capacity (whc) used for the Palmer model is indicated as whc*1. Global average runoff estimate is based on Trenberth et al. (2007).

(Table 3), and are significantly worse than the performance of the API model (0.329 and 0.304 for the training and testing dataset, respectively). However, the benchmark model in Eq. (17) is even worse than the Palmer model [$R(-1) < 0.25$]. As noted above, this is likely due to its diagnostic model and lack of memory regarding past soil moisture conditions. In Fig. 4b, the gap between the benchmark in Eq. (17) and the other prognostic modeling approaches reiterates the importance of representing month-to-month memory when deriving soil moisture proxies for agricultural drought monitoring.

The Z-score map in Fig. 8c also shows the relatively poor performance of the Palmer model compared to the API model serving as a benchmark. Looking at $R(-1)$ for each land cover type in Fig. 7, soil moisture variations in forested areas again demonstrate a weak correlation with NDVI, which is attributed to resilience of forest biomass to variations in soil moisture rather than weakness of a specific observations or model products. The root-zone soil moisture outputs from the Palmer model do not even catch up with the performance of LPRM surface soil moisture retrievals for most land cover types [wooded grasslands/shrubs (WGS), CBS, OS, GRS, and croplands (CRP)] in which agricultural drought monitoring is crucial (Fig. 7). This supports the earlier finding of Bolten and Crow (2012) that existing satellite-based soil moisture products provide at least as much global agricultural drought information as a simple, offline water balance model driven by available global precipitation datasets.

Based on comparisons with linear API results, the Palmer model does not appear to be fully utilizing the precipitation and air temperature information it requires
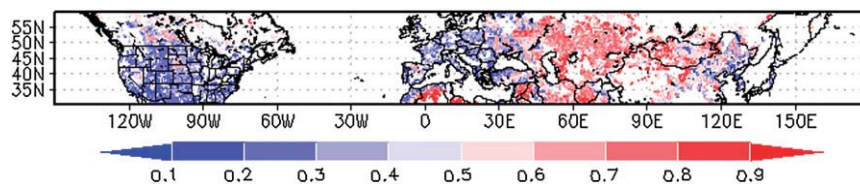
FIG. 10. An example spatial distribution of monthly averaged Kalman gain ($\mathbf{K}_k$) from the Palmer–EnKF system (July 2003).

as input. This implies that the nonlinear physics of the Palmer model are somehow squandering information present in its input. To pinpoint the nonlinearity responsible for this effective loss of information, we constrained the API model using different water holding capacity values as shown in Fig. 9. The original fully linear API model predicts no runoff but has a globally averaged $R(-1)$ of 0.294. Constraining the water holding capacity of the model to produce a realistic amount of global runoff [based on Trenberth et al. (2007)] leads to a sharp degradation in $R(-1)$ results. This suggests that the addition of a simple saturation threshold to the API model, required to generate surface runoff and any type of reasonable streamflow prediction, tends to decrease the utility of API predictions as a predictor of coarse-scale agricultural drought. As such, it provides a direct example of the so-called land surface model multi-objective parameterization problem (Yapo et al. 1998; Vrugt et al. 2003), whereby LSM parameterizations required to minimize error in one type of model output (e.g., runoff predictions) are often at odds with the optimal parameterization required for a second type of output (e.g., root-zone soil moisture monitoring for agricultural drought). Consequently, for coarse-scale agricultural drought monitoring, it appears there is no utility in enforcing a nonlinear saturation limit on soil moisture dynamics. Since such a threshold is required to generate surface runoff, these results also call into question the practice of monitoring agricultural drought using predictions from a land surface model calibrated using streamflow data.

d. Case 4: Efficiency of the EnKF

The two benchmarks in Eqs. (18) and (19) linearly combine anomalies of the model-predicted root-zone soil moisture (from a prognostic model) and observed surface soil moisture. Because of the relatively poor performance of the Palmer model in Case 3 and relatively strong correlation of LPRM soil moisture retrievals with NDVI, the optimized regression coefficient for LPRM retrievals is 3 times higher than the coefficient for Palmer model predictions (0.89 versus 0.28 in Table 3). The better performance of the API approach relative to the Palmer model in Fig. 3 causes relatively more

weight to be placed on the model, but LPRM soil moisture retrievals are still given almost 2 times more weight than API model predictions (Table 3).

The results of Case 4 in Table 3 and Figs. 5 and 8d show that the benchmarks outperform the outputs from the EnKF for both the Palmer and the API model. In other words, this benchmarking evaluation indicates an inefficient implementation of the EnKF systems for both the Palmer and API model. In addition, when the EnKF results of the Palmer model are compared with the performance of the LPRM soil moisture only (Fig. 3), we can see that most of the predictive skill of the EnKF analysis is already present in the assimilated observations; $R(-1) = 0.301$ for θ_{LPRM} while $R(-1) = 0.305$ for $\theta_{\text{PM-KF}}$ in Table 3. This suggests that the background model is contributing relatively little information to the analysis. For several land cover types such as WGS, CBS, GRS, and CRP, where soil moisture observations have relatively stronger correlations with NDVI, the benchmarks outperform the EnKF in Fig. 7.

In theory, the EnKF should assign optimal weights to background model predictions and observations and therefore outperform simple weighting based on spatially fixed parameters [Eqs. (18), (19)]. However, in practice, it is very difficult to specify those weights optimally for each grid because of the complicated nature of error sources in model and observed soil moisture products (Crow and Van den Berg 2010). Here, R and Q are assigned in a relatively simple way based on land cover and distance from WMO rain gauges, respectively, following Bolten et al. (2010) and Bolten and Crow (2012). However, comparisons between the Kalman gain map (based on specified R and Q) and optimized weights determined by the benchmark models suggests that R and Q have not been optimally specified. In particular, a spatially distributed Kalman gain map for the surface soil moisture in Fig. 10 suggests that the EnKF places excessive weight on the model background in the United States and Europe.

An additional assumption applied in this study is that of perfect vertical error correlation [i.e., $\rho = 1$ in Eq. (4)]. However, this assumption may not be always true for certain soil conditions. In that case, the Kalman gain in Eqs. (6) and (7) may link forecasted surface

observations with the root-zone soil moisture prediction inappropriately, which may also contribute to the apparently suboptimal performance of the Bolten and Crow (2012) EnKF system.

5. Discussion and summary

Over the past decade, soil moisture data assimilation has been actively investigated in the land surface modeling community as a tool for improving operational drought monitoring. However, the various soil moisture retrieval algorithms, land surface models, and data assimilation approaches comprising such a system have generally been evaluated separately. Therefore, their relative benefits within a comprehensive data assimilation system have never been fully assessed. In addition, traditional validation approaches based on direct comparison against ground-based observations have limited the evaluation of these systems to isolated, data-rich areas where sufficient ground-based observational resources are available.

Here, all three major components of a soil moisture data assimilation system (retrieval algorithm, hydrologic model, and an assimilation technique) for agricultural drought monitoring are evaluated against a series of benchmarks derived from linear statistical models. This benchmarking approach is designed to assess the value of complex and/or nonlinear processes directly by comparing them with the performance of simple linear models utilizing the same set of input variables. In addition, this study uses a novel evaluation metric, rank correlation with NDVI from Crow et al. (2012b), to quantify the predictive skill of various soil moisture proxies over broad continental-scale regions.

First, the efficiency of LPRM, a soil moisture retrieval algorithm, was evaluated in Case 1. The LPRM soil moisture product marginally outperformed a benchmark model constructed directly from AMSR-E radiance measurements. It should be stressed that the evaluation result of this study does not directly evaluate the intrinsic accuracy of the LPRM soil moisture products, but rather how much skill they posses for a single, specific application (the forecasting of future vegetation conditions). Nevertheless, in this specific context, the nonlinear LPRM retrieval algorithm does not appear to add much additional predictive information compared to a corresponding linear benchmark constructed using LPRM inputs.

In Case 2, vegetation optical depth (τ), the additional AMSR-E retrieval product of LPRM, shows a potential for the retrospective analysis (as opposed to forecasting) of vegetation condition since it demonstrates the highest correlation coefficient with NDVI at $L = 1$ [i.e.,

Rank($\overline{\tau}_k$) temporally lags behind Rank($\overline{NDVI}_k$) by 1 month]. The lagged response of τ to NDVI is consistent with the observations of Jones et al. (2011) made with regards to plant phenology. However, the relationship between τ and NDVI is not fixed across different land cover types (Liu et al. 2011). For $L = -1$, however, neither τ nor MPDI add any information over and above the benchmarks defined in Eqs. (14) or (16).

In Case 3, the modified two-layer Palmer model is evaluated by comparing its performance to the linear, prognostic API modeling approach illustrated in Eqs. (1) and (2) as well as a diagnostic benchmarking model in Eq. (17) based on the same meteorological forcing input utilized in the Palmer and API models. The Palmer model outperforms the diagnostic benchmark model in Eq. (17), but only because of its prognostic nature. More importantly, the poor performance of the Palmer model compared to the API implies the inefficiency of the nonlinear physical characteristics in the Palmer model. Crow et al. (2012b) also used a similar API model as a baseline to evaluate modern LSMs based on the same evaluation criteria for agricultural drought monitoring and showed that those complex LSMs do not produce significantly higher $R(L)$ than the API model. Based on the superior performance of the API model, we hypothesize that the finite saturation threshold of soil layers in the Palmer model may actually hinder its utilization of meteorological input information for agricultural drought monitoring. However, considering the relative simplicity of the models considered here, further research using more complex LSMs is required to better understand this issue.

Although Bolten and Crow (2012) showed the added benefit of assimilating surface soil moisture observations into the Palmer model, the evaluation results of the present study suggest that the benefit of soil moisture data assimilation comes mainly from the strength of the observations, not from the efficiency of a sophisticated assimilation technique such as the EnKF. That is, it is shown that the EnKF outputs from the Palmer model produce lower $R(-1)$ with NDVI than the benchmark model in Eq. (18) that simply combines model forecasts and observations using globally fixed weighting parameters. The relatively inefficient performance of the EnKF is likely linked to inappropriate model and observation error assumptions underlying the estimation of the Kalman gain. This suggests that additional work is still required to optimally parameterize large-scale land data assimilation systems for agricultural drought monitoring.

In spite of significant advances in developing soil moisture data assimilation systems, there are still limitations in evaluating them objectively for further

improvements. This study suggests a simple but effective evaluation approach using statistical benchmark models and successfully pinpoints weaknesses in each component of a soil moisture data assimilation system.

REFERENCES

Abramowitz, G., 2005: Towards a benchmark for land surface models. *Geophys. Res. Lett.,* **32,** L22702, doi:10.1029/2005GL024419.

——, 2012: Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geosci. Model Dev.,* **5,** 819–827, doi:10.5194/gmd-5-819-2012.

——, R. Leuning, M. Clark, and A. Pitman, 2008: Evaluating the performance of land surface models. *J. Climate,* **21,** 5468–5481, doi:10.1175/2008JCLI2378.1.

Adegoke, J. O., and A. M. Carleton, 2002: Relations between soil moisture and satellite vegetation indices in the U.S. Corn Belt. *J. Hydrometeor.,* **3,** 395–405, doi:10.1175/1525-7541(2002)003<0395:RBSMAS>2.0.CO;2.

Andela, N., Y. Liu, A. van Dijk, R. de Jeu, and T. McVicar, 2013: Global changes in dry land vegetation dynamics (1988–2008) assessed by satellite remote sensing: Combining a new passive microwave vegetation density record with reflective greenness data. *Biogeosci. Discuss.,* **10,** 8749–8797, doi:10.5194/bgd-10-8749-2013.

Bolten, J. D., and W. T. Crow, 2012: Improved prediction of quasiglobal vegetation conditions using remotely-sensed surface soil moisture. *Geophys. Res. Lett.,* **39,** L19406, doi:10.1029/2012GL053470.

——, ——, X. Zhan, T. J. Jackson, and C. A. Reynolds, 2010: Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. *IEEE J. Sel. Top. Appl. Earth Observ.,* **3,** 57–66, doi:10.1109/JSTARS.2009.2037163.

Brocca, L., and Coauthors, 2011: Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sens. Environ.,* **115,** 3390–3408, doi:10.1016/j.rse.2011.08.003.

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,** 137–149, doi:10.1016/S0309-1708(02)00088-X.

——, and M. 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.

——, and Coauthors, 2012a: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.,* **50,** RG2002, doi:10.1029/2011RG000372.

——, S. V. Kumar, and J. D. Bolten, 2012b: On the utility of land surface models for agricultural drought monitoring. *Hydrol. Earth Syst. Sci.,* **16,** 3451–3460, doi:10.5194/hess-16-3451-2012.

Dawdy, D. R., and N. C. Matalas, 1964: Statistical and probability analysis of hydrologic data, part III: Analysis of variance, covariance and time series. *Handbook of Applied Hydrology: A Compendium of Water-Resources Technology,* V. T. Chow, Ed., McGraw-Hill, 8.68–8.90.

De Jeu, R., and M. Owe, 2003: Further validation of a new methodology for surface moisture and vegetation optical depth retrieval. *Int. J. Remote Sens.,* **24,** 4559–4578, doi:10.1080/0143116031000095934.

Dorigo, W., K. Scipal, R. Parinussa, Y. Liu, W. Wagner, R. A. M. de Jeu, and V. Naeimi, 2010: Error characterisation of global active and passive microwave soil moisture datasets. *Hydrol. Earth Syst. Sci.,* **14,** 2605–2616, doi:10.5194/hess-14-2605-2010.

Draper, C. S., J. P. Walker, P. J. Steinle, R. A. M. de Jeu, and T. R. Holmes, 2009: An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sens. Environ.,* **113,** 703–710, doi:10.1016/j.rse.2008.11.011.

——, R. H. Reichle, R. de Jeu, V. Naeimi, R. M. Parinussa, and W. Wagner, 2013: Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens. Environ.,* **137,** 288–298, doi:10.1016/j.rse.2013.06.013.

Fieller, E., H. Hartley, and E. Pearson, 1957: Tests for rank correlation coefficients. I. *Biometrika,* **44,** 470–481, doi:10.2307/2332878.

Gruhier, C., and Coauthors, 2010: Soil moisture active and passive microwave products: Intercomparison and evaluation over a Sahelian site. *Hydrol. Earth Syst. Sci,* **14,** 141–156, doi:10.5194/hess-14-141-2010.

Hayes, M., M. D. Svoboda, B. D. Wardlow, M. C. Anderson, and F. Kogan, 2012: Drought monitoring: Historical and current perspectives. *Remote Sensing of Drought: Innovative Monitoring Approaches,* B. D. Wardlow, M. C. Anderson, and J. P. Verdin, Eds., CRC Press, 1–22.

Jackson, T. J., and Coauthors, 2010: Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Trans. Geosci. Remote Sens.,* **48,** 4256–4272, doi:10.1109/TGRS.2010.2051035.

Jones, M. O., L. A. Jones, J. S. Kimball, and K. C. McDonald, 2011: Satellite passive microwave remote sensing for monitoring global land surface phenology. *Remote Sens. Environ.,* **115,** 1102–1114, doi:10.1016/j.rse.2010.12.015.

Kumar, S. V., C. D. Peters-Lidard, J. Santanello, K. Harrison, Y. Liu, and M. Shaw, 2012: Land surface Verification Toolkit (LVT)—A generalized framework for land surface model evaluation. *Geosci. Model Dev. Discuss.,* **5,** 229–276, doi:10.5194/gmdd-5-229-2012.

Liu, Y. Y., R. A. de Jeu, M. F. McCabe, J. P. Evans, and A. I. van Dijk, 2011: Global long-term passive microwave satellite-based retrievals of vegetation optical depth. *Geophys. Res. Lett.,* **38,** L18402, doi:10.1029/2011GL048684.

Luo, Y., and Coauthors, 2012: A framework for benchmarking land models. *Biogeosciences,* **9,** 3857–3874, doi:10.5194/bg-9-3857-2012.

Meesters, A. G., R. A. De Jeu, and M. Owe, 2005: Analytical derivation of the vegetation optical depth from the microwave polarization difference index. *IEEE Geosci. Remote Sens. Lett.,* **2,** 121–123, doi:10.1109/LGRS.2005.843983.

Musyimi, Z., 2011: Temporal relationships between remotely-sensed soil moisture and NDVI over Africa: Potential for drought early warning. M.S. thesis, Dept. of Geo-Information Science and Technology, University of Twenty, Enschede, Netherlands, 41 pp. [Available online at www.itc.nl/library/papers_2011/msc/nrm/musyimi.pdf.]

Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, 2003: Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.,* **41,** 215–229, doi:10.1109/TGRS.2002.808243.

Owe, M., R. de Jeu, and J. Walker, 2001: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.,* **39,** 1643–1654, doi:10.1109/36.942542.

——, ——, and T. Holmes, 2008: Multisensor historical climatology of satellite-derived global land surface moisture. *J. Geophys. Res.,* **113,** F01002, doi:10.1029/2007JF000769.

Palmer, W. C., 1965: Meteorological drought. Research Paper 45, U.S. Department of Commerce, Washington, DC, 58 pp. [Available online at www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf.]

Parinussa, R. M., A. G. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. A. de Jeu, 2011: Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *IEEE Geosci. Remote Sens. Lett.,* **8,** 779–783, doi:10.1109/LGRS.2011.2114872.

Peled, E., E. Dutra, P. Viterbo, and A. Angert, 2010: Technical note: Comparing and ranking soil drought indices performance over Europe, through remote-sensing of vegetation. *Hydrol. Earth Syst. Sci. Discuss.,* **14,** 271–277, doi:10.5194/hess-14-271-2010.

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.

——, D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.,* **130,** 103–114, doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

Rüdiger, C., J.-C. Calvet, C. Gruhier, T. R. Holmes, R. A. de Jeu, and W. Wagner, 2009: An intercomparison of ERS-Scat and AMSR-E soil moisture observations with model simulations over France. *J. Hydrometeor.,* **10,** 431–447, doi:10.1175/2008JHM997.1.

Trenberth, K. E., L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeor.,* **8,** 758–769, doi:10.1175/JHM600.1.

Von Storch, H., and F. W. Zwiers, 2002: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian, 2003: Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res.,* **39,** 1214, doi:10.1029/2002WR001746.

Wagner, W., V. Naeimi, K. Scipal, R. de Jeu, and J. Martínez-Fernández, 2007: Soil moisture from operational meteorological satellites. *Hydrogeol. J.,* **15,** 121–131, doi:10.1007/s10040-006-0104-6.

Wang, J. R., and T. J. Schmugge, 1980: An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Trans. Geosci. Remote Sens.,* **GE-18,** 288–295, doi:10.1109/TGRS.1980.350304.

Wang, X., H. Xie, H. Guan, and X. Zhou, 2007: Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. *J. Hydrol.,* **340,** 12–24, doi:10.1016/j.jhydrol.2007.03.022.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1998: Multi-objective global optimization for hydrologic models. *J. Hydrol.,* **204,** 83–97, doi:10.1016/S0022-1694(97)00107-8.

Zhou, Y., D. McLaughlin, and D. Entekhabi, 2006: Assessing the performance of the ensemble Kalman filter for land surface data assimilation. *Mon. Wea. Rev.,* **134,** 2128–2142, doi:10.1175/MWR3153.1.